

**Microbial Source Module (MSM):
Documenting the Science and Software
for Discovery, Evaluation, and Integration
Updated – 4/17/17**

**Gene Whelan
Rajbir Parmar
Gerard F. Laniak**

4/17/17

U.S. Environmental Protection Agency
Office of Research and Development
National Exposure Research Laboratory
Ecosystems Research Division
Athens, GA 30605

TABLE OF CONTENTS

1. EXECUTIVE SUMMARY	6
2. INTRODUCTION	7
3. MICROBIAL SOURCE MODULE	11
4. APPLICATION OF AN ONTOLOGY TO THE MICROBIAL SOURCE MODULE.....	14
4.1 COMPONENT SUPERCLASS	14
4.1.1 Key Words	14
4.1.2 Component Description	14
4.2 RESOURCES LAYER.....	14
4.2.1 Developer Class	15
4.2.2 Organization Class	15
4.2.3 Project Class	15
4.2.4 Development Level Class	15
4.2.5 Data Class	16
4.3 COUPLING LAYER.....	16
4.3.1 Modeling Framework Class	16
4.3.2 Computational Resolution Class	18
4.3.3 Standards Interface Class	18
4.3.4 Architecture Class	19
4.4 TECHNICAL LAYER.....	19
4.4.1 Operating System Class	19
4.4.2 Programming Language Class	19
4.4.3 Memory Requirements Class	19
4.4.4 Number of Processors	19
4.5 SCIENTIFIC LAYER	20
4.5.1 Domain Class	20
4.5.2 Mathematical Classification Class	20
4.5.3 Symbol Class	20
4.5.3.1 Ontological Metadata Format	20
4.5.3.2 Indices	25
4.5.4 Equation Class	26
4.5.4.1 Summary of Assumptions and Constraints	27

<u>4.5.4.2 Domestic Animal Waste Available for Land Application and Wildlife Shedding Rates</u>	30
4.5.4.2.1 Domestic Animal Waste Available for Runoff.....	30
4.5.4.2.2 Wildlife Shedding Rates	30
<u>4.5.4.3 Accumulated Microbial Loading Rates on Cropland</u>	33
4.5.4.3.1 Wildlife	33
4.5.4.3.2 Dairy Cow	33
4.5.4.3.3 Beef Cattle.....	34
4.5.4.3.4 Poultry	34
4.5.4.3.5 Swine	34
<u>4.5.4.4 Accumulated Microbial Loading Rates on Pasture</u>	35
4.5.4.4.1 Shedding to Land Surface	35
<u>4.5.4.4.1.1 Wildlife</u>	35
<u>4.5.4.4.1.2 Beef Cattle</u>	35
<u>4.5.4.4.1.3 Horses</u>	35
<u>4.5.4.4.1.4 Sheep</u>	36
<u>4.5.4.4.1.5 Other Agricultural Animals</u>	36
4.5.4.4.2 Manure Application to Land Surface	36
<u>4.5.4.4.2.1 Dairy Cow</u>	36
<u>4.5.4.4.2.2 Beef Cattle</u>	37
<u>4.5.4.4.2.3 Horses</u>	37
<u>4.5.4.5 Accumulated Microbial Loading Rates on Forest</u>	37
<u>4.5.4.6 Accumulated Microbial Loading Rates on Urbanized Areas</u>	37
<u>4.5.4.7 Accumulated Overland Microbial Loading Rates to the Land Surface, Adjusted for Die-off</u>	38
4.5.4.7.1 Die-off Adjustment	38
4.5.4.7.2 Accumulated Overland Microbial Loading Rates and Maximum Microbial Storage with Die-off.....	39
<u>4.5.4.7.2.1 Cropland</u>	39
<u>4.5.4.7.2.2 Pasture</u>	39
<u>4.5.4.7.2.3 Forest</u>	40
<u>4.5.4.7.2.4 Urbanized</u>	40
4.5.4.7.3 Maximum Microbial Storage with Die-off	40

4.5.4.8 Microbial Point Source Loading Rates	40
4.5.4.8.1 Cattle in Streams	40
4.5.4.8.2 Septics	41
5. CONTEXT OF THE MICROBIAL SOURCE MODULE WITHIN A MULTI-COMPONENT WORKFLOW	42
5.1 MODELING WORKFLOW	42
5.2 MSM INPUT REQUIREMENTS.....	44
5.2.1 Input Data Variables	44
5.2.2 Microbial Input Comma Separated Values Files	44
5.2.2.1 Domestic Animals	49
5.2.2.2 Septic Systems and Point Sources	50
5.2.2.3 Wildlife.....	51
5.2.2.4 Microbial Die-off	51
5.2.2.5 Urban Sources	51
5.2.2.6 Supporting Literature Information	52
5.3 MSM OUTPUT REQUIREMENTS	53
5.4 TRANSFORMATION OF LATITUDE-LONGITUDE COORDINATES TO SUBWATERSHED DESIGNATIONS	53
5.5 WORKFLOW SET-UP AND EXECUTION.....	56
5.5.1 Automating Watershed Delineation and Microbial Source Mapping	56
5.5.2 Workflow Instantiation.....	58
5.6 ONTOLOGICAL RELATIONSHIPS BETWEEN VARIABLES, EQUATIONS, AND COMPONENTS	59
6. MAPPING THE MICROBIAL SOURCE MODULE ONTOLOGICAL METADATA DICTIONARIES TO AN EXTENSIBLE MARKUP LANGUAGE DOCUMENT.....	63
REFERENCES.....	68
APPENDIX A	76
EXAMPLE MICROBIAL SOURCE MODULE XML DOCUMENT FOR INPUT PARAMETERS/VARIABLES.....	76

1. EXECUTIVE SUMMARY

The Microbial Source Module (MSM) estimates microbial loading rates to land surfaces from non-point sources, and to streams from point sources for each subwatershed within a watershed. A subwatershed, the smallest modeling unit, represents the common basis for information consumed and produced by the MSM which is based on the HSPF Bacterial Indicator Tool. Non-point sources include numbers, locations, and shedding rates of domestic agricultural animals (beef cattle, dairy cow, swine, poultry, etc.) and wildlife (deer, duck, raccoon, etc.). Monthly maximum microbial storage and accumulation rates on the land surface, adjusted for die-off, are computed over an entire season for four land-use types (cropland, pasture, forest, and urbanized/mixed-use) for each subwatershed. Point-source microbial loadings to instream locations are determined for instream shedding by cattle, septic systems, and built structures. Instream shedding varies monthly, although average-annual loadings, regardless of year, are assumed for septic-system releases and built structures. Built structures, such as Publicly Owned Treatment Works/Wastewater Treatment Plants (POTWs/WWTPs), are captured as constant inputs through a complementing software module, inputs which that will eventually be replaced with actual time series through an externally watershed model.

The MSM functions within a larger modeling system that characterizes human-health risk resulting from ingestion of water contaminated with pathogens. The loading estimates produced by the MSM can be consumed by watershed models (e.g., HSPF model) that simulate flow and microbial fate/transport. Microbial cells within recreational waters can then be supplied as input to a risk module (e.g., MRA-IT model) to estimate human exposure and risk.

A new approach has been taken in the design and implementation of MSM documentation and software with the goal of enhancing the MSM's potential for reuse and interoperability with modeling systems. Satisfying this goal requires the MSM to be easy to discover, understand, evaluate, access, and integrate: therefore, the strategy is to 1) facilitate discovery, understanding, and evaluation by documenting the module with an ontological framework, and 2) facilitate access and integration by implementing the software as a web service.

The ontological framework is based on the Water Resources Component (WRC) ontology. The WRC is a structured way to describe the ontology of an environmental system represented by a science software component such as the MSM. The MSM ontology is documented in Protégé, an editor that implements the Web Ontology Language (OWL). The ontological framework also documents key aspects of the MSM including key words; module purpose, assumptions, and constraints; inputs; outputs; and internal variables. Finally, this document represents a traditional Theory Manual that accompanies the science; it has been structured to mirror the ontology, thus facilitating development in Protégé.

To facilitate access and integration, MSM software has been designed with object-oriented principles and is "published" as a Representational State Transfer web service. The web service, which consumes XML input and produces XML output, can be accessed directly via browser add-ons such as Postman for Chrome. The most common way to consume the web service is through a custom desktop or web client program. The web service is platform and programming language agnostic.

2. INTRODUCTION

Development of integrated watershed models increasingly requires coupling of multidisciplinary, independent models and collaboration between scientific communities since component-based modeling enables integration of models from different disciplines (Elag and Goodall, 2013). Many modeling frameworks have adopted an approach to compartmentalize science through individual models, linking them to create larger modeling workflows. Integrated Environmental Modeling (IEM) systems focus on transferring information between components by capturing a conceptual site model (CSM), establishing local metadata standards for input/output of models/databases, managing data/information flow between models and throughout the system, facilitating quality control of data/information exchanges (e.g., units checking, units conversion, inter-language transfers), handling warnings/errors, and coordinating sensitivity/uncertainty analyses (Whelan et al., 2014a). Although many computational software systems are designed to facilitate communication between, and execution of, components (Whelan et al., 2014a; Laniak et al., 2013), there are no common approaches, protocols, or standards for turn-key linkages between software systems and models, especially if the intent is not to modify components.

While there has been a notable increase in component-based modeling frameworks in recent years (Laniak et al., 2013; Whelan et al., 2014a), there has been less work on creating standard vocabularies, metadata, semantics, and ontologies (see Table 1) to ensure proper technical and conceptual assemblage, although work on ontologies is gaining traction. For example, Elag and Goodall (2012, 2013) and Morsey et al. (2014) designed an ontology for the water resources community, using a skeletal methodology described by Uschold and Gruninger (1996). Titled the Water Resources Component (WRC) ontology, it was developed to advance application of component-based modeling frameworks across water-related disciplines. Although their WRC ontology was designed for water resources, its design can be extended to include other domains, such as microbial source-term modeling, to document individual modeling components for eventual inclusion in larger, disparate systems. It advances the conceptual integration of components from different, but related, disciplines by handling semantic and syntactic heterogeneities to describe them, so they can be more easily reused, extended, and maintained by a larger group of model developers and end users. The WRC has four ontological layers (Elag and Goodall, 2012, 2013):

- Resources: defines digital resources related to the component.
- Coupling: defines coupling standards used by the component, the framework in which the component can be coupled, and its computational resolution.
- Scientific: describes the equations, symbols, mathematical classification, and component purpose.
- Technical: defines required computer architecture to employ and edit the component.

An overview of the WRC Ontology's four layers and their classes (Elag and Goodall, 2012, 2013) is presented in Figure 1. Details of the layers are presented in Figure 2. The strength of this ontology, like others, is its structure for capturing and documenting key information that define a component's vocabulary, metadata, semantics, and ontology to promote interoperability of components across disciplines and modeling frameworks.

The purposes of this effort are to 1) enable construction of scientifically consistent, coherent environmental software systems for multi-disciplinary data integration, decision and policy support, and

modeling; and 2) discover, access, and integrate components developed and published by different scientists (Laniak, 2012). The objectives are to 1) describe a model, using a standard ontology (see Table 1), so that the module can be discovered, understood, evaluated, accessed, and implemented on the cloud, and 2) place the model within the context of a workflow. A model called the Microbial Source Module (MSM) is described using this ontology. A glossary of terms related to interoperability is provided in Table 1. The ontology documents metadata, syntactics, and semantics of the model's Input/Output (I/O) through expanded dictionaries (Whelan et al., 2014a), mathematical formulations that define and/or use each I/O parameter/variable, constraints (i.e., assumptions) associated with each I/O parameter/variable, and an Extensible Markup Language (XML) file that encodes the I/O dictionaries for access and execution on the cloud. An in-depth discussion of applications of the WRC ontology relative to the MSM is presented, followed by a description of the MSM within a more complex modeling environment, where ontological relationships are captured within a more inclusive, multi-component paradigm.

Table 1. Definition of Terms Related to Interoperability

TERM	DEFINITION
Data	Information that is consumed and produced
Vocabulary	Terminological dictionary, which contains designations (e.g., names) and definitions from one or more specific subject fields (JCGM, 2008)
Taxonomy	Science of classification according to a pre-determined system with the resulting catalog used to provide a conceptual framework for discussion, analysis, or information retrieval (i.e., identifies, names, and classifies data, ¹ so it can be standardized, shared, and re-used in multiple systems ²)
Metadata	Information about the data used to capture content (Kashyap and Sheth, 2000)
Syntactics	Data structure [i.e., how elements are sequenced to form valid conditions (e.g., keywords, object names, operators, delimiters, and so on are in the correct places)]
Semantics	Data and their relationship to other data ³ by relating content and representation of information resources to entities and concepts in the real world (Meersman and Mark, 1997) and including not only the metadata about data but also the intended use (i.e., application) of data (Sheth, 2001)
Ontology	Explicit specification of conceptualization, describing knowledge about the domain ⁴ and relationships between domain concepts ⁵

¹ <http://en.wikipedia.org/wiki/Taxonomy>

² <http://it.toolbox.com/blogs/irm-blog/the-benefits-of-a-data-taxonomy-4916>

³ http://en.wikipedia.org/wiki/Semantic_data_model

⁴ <http://www.obitko.com/tutorials/ontologies-semantic-web/ontologies.html>

⁵ [http://en.wikipedia.org/wiki/Ontology_\(information_science\)](http://en.wikipedia.org/wiki/Ontology_(information_science))

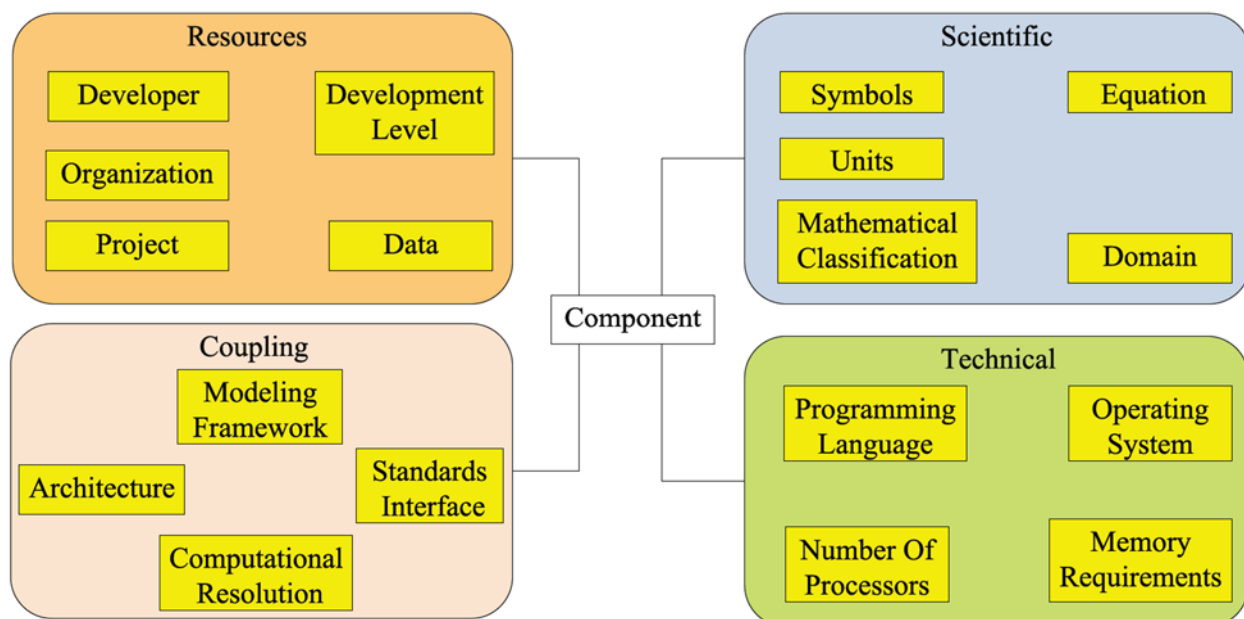


Figure 1. Overview of the Water Resources Component Ontology, describing the Four Basic Layers and their Classes (Elag and Goodall, 2012, 2013)

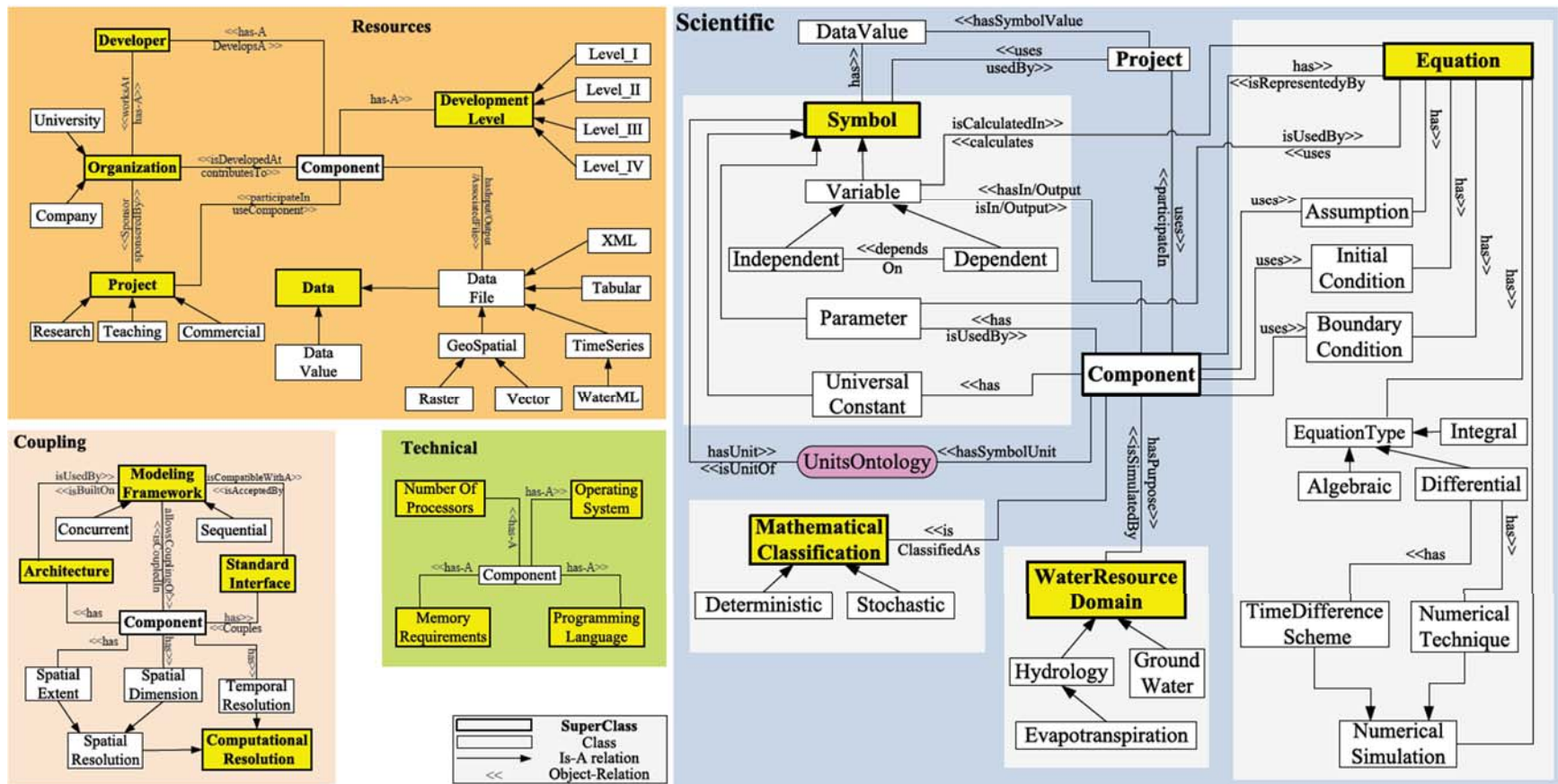


Figure 2. Details Comprising the Resource, Coupling, Scientific, and Technical Layers of the Water Resources Component Ontology (Elag and Goodall, 2012, 2013)

3. MICROBIAL SOURCE MODULE

A coupled software system is being developed that will connect IEM legacy technologies to support a watershed-scale Quantitative Microbial Risk Assessment (QMRA) source-to-receptor assessment, focusing on animal-impacted catchments, although point sources are also considered. A Quantitative Microbial Risk Assessment (QMRA) is a modeling approach that integrates disparate data (including fate/transport, exposure, and human health effect relationships) to characterize potential health impacts/risks from exposure to pathogenic microorganisms (Soller et al., 2010; Whelan et al., 2014b; Haas et al., 1999; Hunter et al., 2003). As Whelan et al. (2014b) note, a QMRA's conceptual design fits well within an integrated, multi-disciplinary modeling perspective (illustrated in Figure 3) which describes the problem statement, data access retrieval and processing [e.g., D4EM (EPA, 2013a; Whelan et al., 2009; Wolfe et al., 2007)]; software frameworks for integrating models and databases [e.g., FRAMES (Johnston et al., 2011)]; infrastructures for performing sensitivity, variability, and uncertainty analyses [e.g., SuperMUSE (Babendreier and Castleton, 2005)]; and risk quantification. Coupling modeling results with epidemiology studies allows policy-related issues (e.g., EPA, 2010; EPA and USDA, 2012) to be explored (Figure 3).

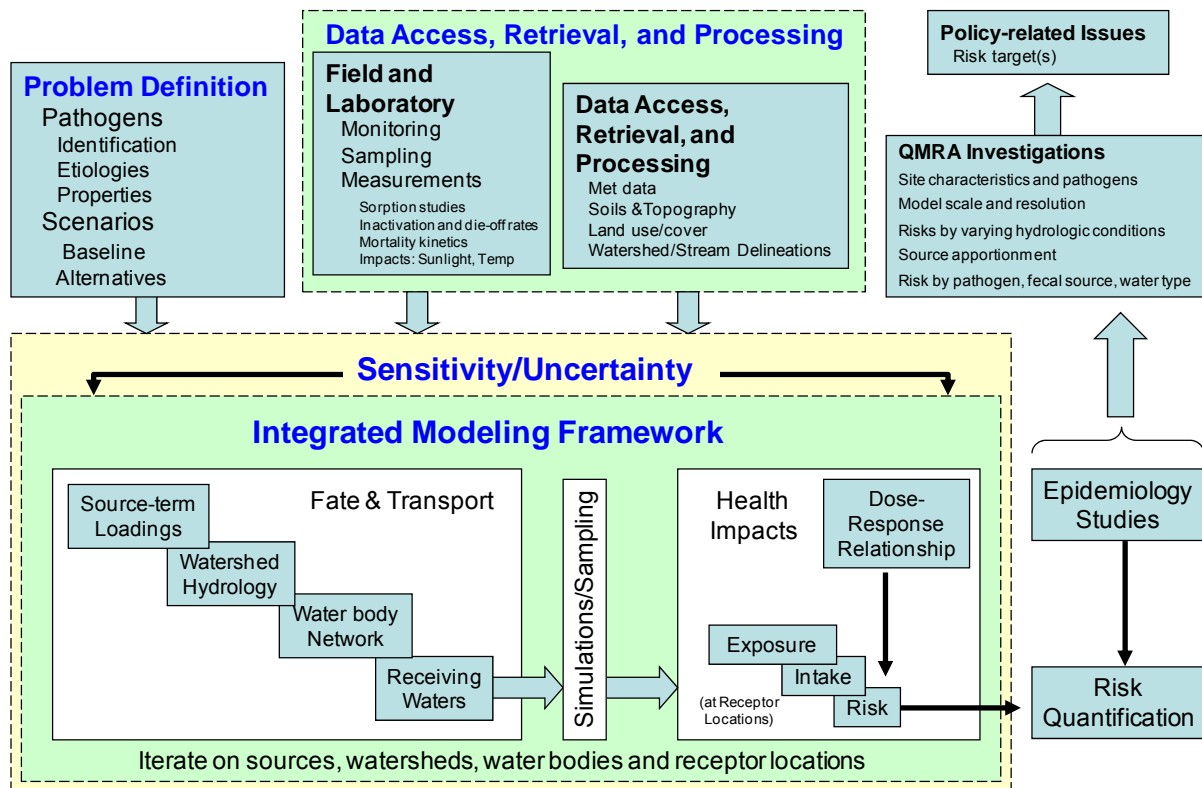


Figure 3. One possible rendition of QMRA from an integrated, multi-disciplinary multimedia modeling framework perspective that links problem definition; data access, retrieval, and processing; integrated modeling framework with source-to-receptor environmental models, housed within a sensitivity/uncertainty software structure; risk quantification linked to epidemiology studies and policy-related uses (After Whelan et al., 2014b)

An important piece of the IEM microbial workflow is the Microbial Source Module (MSM) that organizes, analyzes, and supplies data necessary to determine microbial loading rates within a watershed to support watershed modeling. The MSM makes this determination from sources correlated to four land-use types (cropland, pasture, forest, and urbanized/mixed-use) for each subwatershed, the smallest modeling unit within a watershed. Microbial sources include numbers and locations of domestic agricultural animals (beef cattle, dairy cow, swine, poultry, etc.) and wildlife (deer, duck, raccoon, etc.) with estimated shedding rates due to grazing; manure application rates where the manure is directly incorporated into a pasture's soil; and loading rates due to urbanized/mixed-use activities (commercial, transportation, etc.). Manure contains microbes, and the monthly maximum microbial storage and accumulation rates on the land surface, adjusted for die-off, are computed over an entire season, capturing seasonal trends (e.g., winter to summer), to represent the source for subsequent overland fate and transport to instream locations. Monthly point source microbial loadings to instream locations are also determined for instream shedding by cattle and septic systems, monthly values for septic systems are based on average-annual loadings. Built structures [POTWs/WWTPs (Publicly Owned Treatment Works/Wastewater Treatment Plants)] with average-annual loadings are captured as constants in a complementing software module that represents a place holder that will eventually be replaced with actual time series through the watershed model. The MSM module is based on the HSPF (Bicknell et al., 1997) Bacterial Indicator Tool (EPA, 2013b, 2013c). The subwatershed is the basis for spatial data consumed and produced by the MSM. Although microbial loadings maybe determined by land use type (e.g., pasture, cropland, urbanized, and residential), they are combined and assigned to the entire subwatershed. Attributes of the MSM, captured within the ontological description, include the following:

- The MSM considers only one microbe at a time and must be individually executed, if multiple microbes are being assessed; the MSM, therefore, does not need to consume any information that specifically identifies the microbe by name.
- Overland microbial loading rates, accounting for die-off, are computed for each subwatershed by land use type on a monthly basis.
- The MSM considers microbial loadings from sources correlated to four land-use types for each subwatershed, where a subwatershed is the smallest area associated with watershed modeling. Correlated sources and land use types are pictorially illustrated in Figure 4 and summarized as follows:
 - Cropland:
 - Land application of domestic animal waste (Beef Cattle, Dairy Cow, Swine, and/or Poultry)
 - Wildlife shedding
 - Pasture:
 - Shedding due to grazing (Beef Cattle, Horse, Sheep, and/or Other domestic animal)
 - Land application of domestic animal waste (Beef Cattle, Dairy Cow, and/or Horse)
 - Wildlife shedding
 - Forest: Wildlife shedding
 - Built: Urban-related releases:
 - Commercial and Services
 - Residential
 - Mixed Urban
 - Transportation, Communication, Utilities

- Direct Loading to Streams (point source releases):
 - Septic systems
 - Instream Beef Cattle shedding
- Instream loading rates are identified with each subwatershed.
- The MSM currently assumes the smallest time increment associated with nonpoint-source loadings is monthly, representing typical loadings for that month, regardless of the year.
- The MSM currently assumes monthly loadings from direct shedding to streams by Beef Cattle.

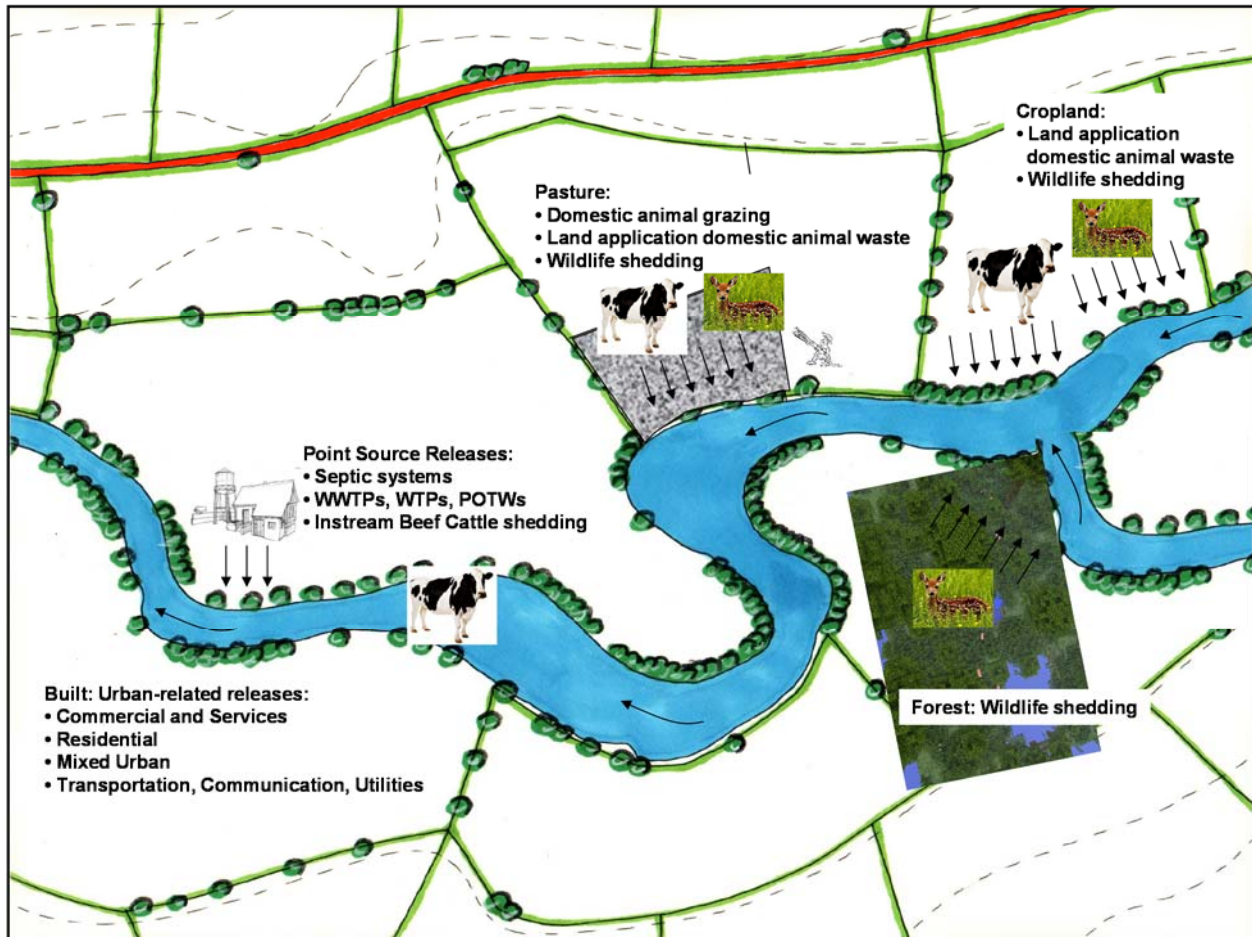


Figure 4. Schematic correlating microbial sources and land use types considered by the Microbial Source Module (After Whelan et al., 2014b)

4. APPLICATION OF AN ONTOLOGY TO THE MICROBIAL SOURCE MODULE

To demonstrate how ontologies like the WRC can help define a component's vocabulary, metadata, semantics, and ontology, the Microbial Source Module has been singled out, and an ontological analysis has been performed and documented. Using the WRC as a guide, this section provides an ontological description of the MSM using the Component superclass and four ontology “layers”: Resource, Coupling, Technical, and Scientific. Because the Component superclass represents the central hub in each layer, it is described first, followed by descriptions of the four layers.

4.1 COMPONENT SUPERCLASS

The Microbial Source Module (MSM) is the Component. Key words and a description of the MSM are provided for the purpose of discovery.

4.1.1 Key Words

Source-term model, microbial modeling, microorganisms, microbial loading rates, watershed, watershed modeling, microbial properties

4.1.2 Component Description

The Microbial Source Module (MSM) determines microbial loading rates within a watershed from sources correlated to four land-use types (cropland, pasture, forest, and urbanized/mixed-use) for each subwatershed, the smallest modeling unit within a watershed. Microbial sources include numbers and locations of domestic agricultural animals (beef cattle, dairy cow, swine, poultry, etc.) and wildlife (deer, duck, raccoon, etc.), with estimated shedding rates due to grazing; manure application rates where the manure is directly incorporated into a pasture's soil; and loading rates due to urbanized/mixed-use activities (commercial, transportation, etc.). The monthly maximum microbial storage and accumulation rates on the land surface, adjusted for die-off, are computed over an entire season, representing the source for subsequent overland fate and transport to instream locations. Instream shedding varies monthly, although average-annual loadings, regardless of year, are assumed for septic-system releases and built structures. The type of septic system (e.g., gravity, pressure distribution, sand filter, and mound) is not differentiated in the model. Although built structures, such as Publicly Owned Treatment Works/Wastewater Treatment Plants (POTWs/WWTPs), are initially captured as constant inputs and will eventually be replaced with actual time series through an externally supported user interface.

4.2 RESOURCES LAYER

The Resources layer has five super classes that collectively describe the component's “digital Resources” (as illustrated in [Figures 1 and 2](#), [Elag and Goodall, 2013](#)), identifying the developers, pertinent organization, projects supporting the component, its development level (Levels I through IV which represent basic model research up to a fully deployable, vetted model), and information on data used by the component.

4.2.1 Developer Class

The Developer class stores information about the component's development team (Elag and Goodall, 2013):

Rajbir Parmar (Software)
Gene Whelan (Science)
Gerard F. Laniak (Ontology)

4.2.2 Organization Class

The Organization class is related to the Developer class and identifies the agency or institute where the component is developed (Elag and Goodall, 2013):

U.S. Environmental Protection Agency
Office of Research and Development
National Exposure Research Laboratory
Ecosystems Research Division
960 College Station Road
Athens, GA 30605

4.2.3 Project Class

The Project class defines information about projects, where components are coupled to form a workflow. When a component is part of a modeling workflow, it is necessary to know where and how it is used within that project, including any specific project requirements (Elag and Goodall, 2013). Projects related to and supporting this effort include:

- Sustainable and Healthy Communities Research Program (SHCRP)
 - Task 1.1.2.2: Interoperability (2014)
 - The purpose is to develop Guidelines for Designing and Implementing Environmental Decision Support Software for Reuse and Interoperability
- Safe and Sustainable Water Research Program (SSWR)
 - Task 2.2.B.8: Integrated Public Health Evaluation of Pathogens (e.g., Occurrence, Exposure, Effects and Treatment) (2012-2015)
 - The purpose is to provide Quantitative Microbial Risk Assessment (QMRA) software infrastructure to perform predictive modeling and microbial risk assessments in mixed watersheds, using pathogen and indicator loadings and transport via models

Later, this document describes how the MSM component is coupled in a workflow containing multiple models.

4.2.4 Development Level Class

The Development Level class defines the component's development stage according to a four level scheme. Babendreier (2010) adopted guidance from the U.S. EPA National Exposure Research Laboratory's Modeling Workgroup to classify model development on four levels, as presented in Table 2.

The levels range from the most rigorous QA at Level I to Level IV. Level I directly and/or immediately supports specific Agency rule-making, enforcement, regulatory or policy decisions, and Level IV documents basic, exploratory, or conceptual model-based research to study basic phenomena or issues. The MSM is a QA'ed at Level IV.

4.2.5 Data Class

The Data class has two subclasses: Data File and Data Value (Elag and Goodall, 2013). The Data File has four subclasses: Geospatial, Tabular, Time Series, and Extensible Markup Language (XML) data and the Data Value class stores numerical or categorical values used by the MSM component. The relationship between the MSM Component and Data class can be input, output, or associated data. Examples of associated data include model parameters/variables or source code files (Elag and Goodall, 2013). Identifying existing data resources and describing the exact format of the data document could enable components to utilize remote data sources in an automated manner. MSM utilizes XML to describe input/output file content (xml schema) and exchange data with the user as input/output data values.

4.3 COUPLING LAYER

Elements of the Coupling Layer are presented in Figures 1 and 2. The Coupling Layer answers three questions about component coupling (Elag and Goodall, 2013): What coupling standards are used by the component? In which frameworks can components be coupled? What is the computational resolution of the component? The Coupling Layer addresses these questions through four classes: 1) Modeling Framework, 2) Standards Interface, 3) Architecture, and 4) Computational Resolution (Figure 2). Figure 5 presents workflow relationships and interactions between the MSM and other components from which it consumes and for which it produces data. Discussions of the MSM, other components, and the interactions between each other within the workflow are presented in Chapter 5.

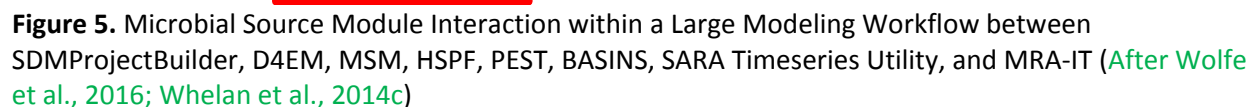
4.3.1 Modeling Framework Class

A Modeling Framework provides an environment where components can be coupled (Elag and Goodall, 2013). In component-based modeling, it couples components that adopt a specific Standards Interface and Architecture. A modeling Component can be used within a Modeling Framework if its design is consistent with the Framework's Standards Interface and Architecture. Elag and Goodall (2013) classify the Modeling Framework based on the level of interaction between components: 1) Concurrent, in which the framework allows components to communicate during the time horizon of the simulation (e.g., dynamic feedback during runtime, where data are shared at each time step) and 2) Sequential, in which the framework allows components to communicate after the simulation time horizon concludes (i.e., each model runs to completion in sequence).

The MSM operates as a stand-alone module, where necessary input data are available for consumption, or as a module integrated into a modeling framework. The MSM Component design accommodates the specific Standards Interface and Architecture associated with the Framework for Risk Assessment in Multimedia Environmental Systems (FRAMES) (Whelan et al., 2014a). Data transfer protocols are captured in ontological metadata dictionaries, extensions of the metadata described by Elag and Goodall (2013) and Whelan et al. (2014a). Either as a stand-alone or within a framework, the MSM operates sequentially, where the module communicates after the conclusion of the time horizon.

Table 2. Guidance for interpretation of QA level requirements for modeling projects (Babendreier, 2010)

						Example Model Evaluation Tasking							Recordkeeping		
Best Practice		Level	Description	Descriptors of Models as Technology and Science		Code Verification	Sensitivity Analysis	Parameter Est./Calibration	Corroboration			Peer Advice, Review	Uncertainty Analysis	Example Recordkeeping Requirements	
Best Practice Type		QA Level	Model-Based Category Description	Models as Technology (i.e., software)	Models as Science (i.e., archives of data/science)				Via Observations	Via Model Comparison	Via Other Means				
Problem Specification: Solving the Right Problem	Project Specific - All Types May Be Involved		Category IV	Basic, exploratory, or conceptual model-based research to study basic phenomena or issues.	Self-certification via lab notebook, publishing, or other means.	Self-certification via lab notebook, publishing, or other means.	Required (self)	As undertaken	As undertaken	As undertaken	As undertaken	As undertaken	Recommended	As undertaken	Lab notebook and peer-reviewed journal articles as appropriate for project needs. Source code, executables, data, and results.
	Model Evaluation	Development	Category III	Demonstration or proof of concept of model's technology basis.	Certification is geared to a reproducible demonstration of software behavior by the developer. The software does what we think it is supposed to do. Includes planned unit-level (i.e. module) testing. Includes an appropriate level of systems testing if the integrated system is being certified at this QA level (e.g., stress testing integrated components) .	Certification at this level does not necessarily require knowledge of the model's accuracy/precision; i.e., its ability to represent or predict the system of interest/focus. As appropriate and feasible for the project's needs, model evaluation studies are conducted and reported. A process of building overall confidence in model output data for specific uses.	Testable Document Required	As practical for project needs	As practical for project needs	As practical for project needs	As practical for project needs	Recommended	As practical for project needs	Source code, executables, data, and results. Documentation describing in some form software requirements, design (approach), specifications (I/O), test plan(s), and expected/actual test results. Model corroboration studies, sensitivity studies, and parameter estimation/calibration methodologies are increasingly best conducted with some level of software verification in place. Model evaluation studies are pursued as practical via peer-reviewed journal articles/presentations, and other independent peer review venues.	
			Category II	Model-based research of high programmatic relevance which, in conjunction with other ongoing or planned studies, is expected to provide complementary support of Agency rule-making, regulatory, or policy decisions.	Meets Category III requirements for code verification; systems level testing would typically be expected to be more thorough than for Category III. Code verification is conducted by person(s) other than immediate code developer (e.g., other teammate, etc).	Demonstration or proof of concept of the model's science basis is valued. Documentation of model and model evaluation tasking are sufficient to support the intended purpose/use of the model. Implies a minimal level of understanding by users of relative reliability and attendant uncertainties associated with model data. Typically includes peer-community support of the science basis of the model.	Required (non-developer)	Recommended	As needed	As needed	As needed	Required (as needed)	Required (as needed)	Source code, executables, data, and results. Supporting verification document(s) and model evaluation studies. Supporting and non-supporting peer-reviews, and responses to peer-reviews as appropriate. An overall statement of uncertainties involved relative to the needs of the specific use; may involve separating effects of natural variability on model output from the effects of sources of epistemic uncertainty.	
		Application	Category I	Model-based research which directly and/or immediately supports specific Agency rule-making, enforcement, regulatory, or policy decisions.	Meets Category II requirements for code verification; systems level testing would typically be expected to be more thorough than for Category III or II. Code compilation and verification is conducted by person(s) independent of the immediate software development team. As appropriate and feasible, engages best available or practically achievable methods.	Documentation objective: the model application is sufficient to support the intended purpose/use. Modelers focus their role on best describing sources of uncertainty in outputs, and the range and scale of associated outcomes possible. A practical level of understanding by users of the model is expected, with acceptable levels of community wide agreement on utility of use.	Required (independent)	Recommended	As appropriate; tends towards best available or practically achievable	As appropriate; tends towards best available or practically achievable	As appropriate; tends towards best available or practically achievable	As appropriate; tends towards best available or practically achievable	Required (as needed)	Required (as needed)	Source code, executables, data, and results. Supporting verification document(s) and model evaluation studies. Supporting and non-supporting peer-reviews, and responses to peer-reviews as appropriate. An overall statement of uncertainties involved relative to the needs of the specific use; may involve separating effects of natural variability on model output from the effects of sources of epistemic uncertainty.



The Computational Resolution class covers both temporal and spatial resolutions of the component model (Elag and Goodall, 2013). The Temporal Resolution class introduces the order of permissible operating time steps, and the Spatial Resolution class describes the space resolution. For numerically-based models, descriptive information such as grid or mesh size and dimensionality (1-D, 2-D, 3-D), as well as size of the time step to keep it numerically stable, are important to capture. For lumped-parameter or reduced-form models, data needs are less onerous. For the spatial resolution, the MSM is designed to work on polygon-shaped subwatershed elements, with no minimum or maximum size defined, although typical sizes range from HUC-8s to HUC-16s. For the temporal resolution, overland microbial loading rates and direct loading to the stream are on a monthly basis.

A Standards Interface is the way data, both input and output, are exchanged with the Component. From a developer's perspective, accessing the software functionality is achieved via MSM's Application Programming Interface (API) through a web service. An API is a set of routines, protocols, and tools for building software applications; it expresses a software component in terms of operations, inputs, outputs, and underlying types. An API defines functionalities that are independent of implementation which allows definitions and implementations to vary without compromising each other. A good API makes it easier to develop a program by providing all the building blocks for a programmer (API, 2015).

4.3.4 Architecture Class

Software architecture is the fundamental organization of a system embodied in its components, their relationships to each other and the environment, and the principles guiding its design and evolution (IBM, 2006; IEEE, 2000). It represents the high-level structure of a software system which facilitates communication (Wikipedia, 2014). The MSM is designed as a web service and reflects a Service-Oriented Architecture (SOA) which is a design pattern based on distinct pieces of software providing application functionality as services to other applications via a service-orientation. It is independent of any vendor, product, or technology. A service is a self-contained unit of functionality such as retrieving an online bank statement. Services can be combined by other software applications to provide the complete functionality of a large software application. SOA makes it easy for computers connected over a network to cooperate (i.e., share data and information). Every computer can run an arbitrary number of services, and each service is built to ensure it can exchange information with any other service in the network without human interaction and without needing to change the underlying program itself (SOA, 2015).

4.4 TECHNICAL LAYER

The Technical Layer answers questions about the computer architecture required to 1) run a component simulation, 2) edit or update the component code, 3) determine computational resources required by the component, and 4) optimize simulation time, given available computational resources (Elag and Goodall, 2013). The four Technical component classes (i.e., Operation System, Programming Language, Memory Requirements, and Number of Processors) are described as follows.

4.4.1 Operating System Class

The Operating System (OS) class defines the different systems that are compatible with the component (Elag and Goodall, 2013). The MSM was developed under the Microsoft Windows OS.

4.4.2 Programming Language Class

The Programming Language class determines the language used in writing the component (Elag and Goodall, 2013). The MSM software is written with a combination of C# and ASP.NET.

4.4.3 Memory Requirements Class

The Memory Requirements class describes required memory capacity to support a single component simulation (Elag and Goodall, 2013). Since the MSM software does not consume large volumes of data, it has no specific memory requirements.

4.4.4 Number of Processors

The Number of Processors class includes elements representing the number of processors the component can leverage (Elag and Goodall, 2013). The MSM software runs as a web service, so from a user's perspective (including software developers), the MSM is executed on a single processor.

4.5 SCIENTIFIC LAYER

The Scientific Layer describes the component's equations, Input and Output (I/O) variables, parameters, purpose, and mathematical classification (Elag and Goodall, 2013). Components of the Scientific Layer are shown in Figures 1 and 2. The four Scientific Component classes (Domain, Mathematical Classification, Symbol, and Equation) are described as follows.

4.5.1 Domain Class

The Domain describes the category with which the Microbial Source Module should be affiliated and is designated as a Source-term model.

4.5.2 Mathematical Classification Class

The Mathematical Classification class defines how variables are treated in space and time and if they are deterministic or stochastic. The MSM is classified as “deterministic,” as it uses algebraic equations in a deterministic mode.

4.5.3 Symbol Class

The Symbol class classifies symbols as Independent or Dependent Variables, Parameters, or Constants, where each must have a unique, unambiguous name, and where the names themselves can represent the symbols (Elag and Goodall, 2013). A variable is an entity that changes with respect to another, and a parameter is an entity that connects variables. A variable is a real world entity with a measurable quantity, while a parameter is an entity that may or may not be measurable; therefore, the same set of variables can be described by different parameters (e.g., indices) (Difference Between, 2012). For example, in the equation of a straight line ($y = mx + b$), x and y are independent and dependent variables, respectively, and m and b are parameters. When modeling this equation, x , m , and b are typically inputs, and y is typically an output. The output of one model, which produces dependent variables, could be classified as independent variables or parameters of a downstream model that consumes the information as input.

4.5.3.1 Ontological Metadata Format

Tables 3 and 4 extend the variable names and definitions associated with Tables 5 and 6 to succinctly capture the vocabulary, metadata, syntactics, semantics, and ontology associated with MSM input and output variables, respectively. Tables 3 and 4 are ontological dictionaries describing each variable's metadata, its relationship to other variables through indices, its use, mathematical expressions that define or use it, and relevant assumptions that impact its use and/or value. Table 7 summarizes the indices and provides their definitions. An ontological dictionary, as used here, groups like and related parameters and provides a single naming convention for variables and parameters shared by modeling components; specifically, each table provides the following information (Whelan et al., 2014a):

- Parameter/Variable Name
- Parameter/Variable Description (Definition of parameter/variable)

Table 3. Microbial Source Module Component – **Input:** Relevant Vocabulary, Taxonomy, Metadata, Syntactics, Semantics, and Ontology

Dictionary Name	Variable Name	Parameter Description	Cardinality	Data Type (Float, Integer, etc.)	Primary Key (i.e., used as a Universal parameter?)	Scaler [Not Self-indexed (i.e., not self-enumerated) = True]	Minimum	Maximum	Measure	Unit	Stochastic (Is it allowed to change in a Monte Carlo analysis?)	Index 1	Index 2	Index 3	Parameter Type [Independent, Dependent, Parameter (e.g., Index)]	Parameter Function (Input, Output, Internal)	Component	Document (Reference number with reference)	Equation in Document that Defines Variable (Reference numbers with relevant equations in parentheses)	Equations in Document that use Variable (Reference numbers with relevant equations in parentheses)	Equation Type	Relevant Assumption (Reference number with relevant assumptions in parentheses)
SDMPBOutput	Agricultural	Index on Domestic Animal (Name), self-indexed (i.e., self-enumerated)	1	STRING	TRUE	FALSE					FALSE				Parameter	Input	Microbial Source Module	1. Whelan et al. (2017)		1 (1,2,6-17, 32,37)		
SDMPBOutput	LandUse	Index on Land Use Type, self-indexed (i.e., self-enumerated)	1	STRING	TRUE	FALSE					FALSE				Parameter	Input	Microbial Source Module	1. Whelan et al. (2017)		1 (3-7,9-23, 27-31,35)		
SDMPBOutput	Subwatershed	Index on Subwatershed (Identification designation), self-indexed (i.e., self-enumerated)	1	Integer	TRUE	FALSE	0	1000			FALSE				Parameter	Input	Microbial Source Module	1. Whelan et al. (2017)		1 (6,7,9-23, 27-39)		1 (18)
SDMPBOutput	Urbanized	Index on type of Built up area, self-indexed (i.e., self-enumerated)	1	STRING	TRUE	FALSE					FALSE				Parameter	Input	Microbial Source Module	1. Whelan et al. (2017)		1 (23,39)		
SDMPBOutput	SubUrbanized	Index on Sub-urbanized Built up area, self-indexed (i.e., self-enumerated)	1	STRING	TRUE	FALSE					FALSE				Parameter	Input	Microbial Source Module	1. Whelan et al. (2017)		1 (18-21)		
SDMPBOutput	Wildlife	Index on Wildlife (Name), self-indexed (i.e., self-enumerated)	1	STRING	TRUE	FALSE					FALSE				Parameter	Input	Microbial Source Module	1. Whelan et al. (2017)		1 (3-5)		
SDMPBOutput	MonthID	Index on Month of the year (January, February, ..., December), self-indexed (i.e., self-enumerated)	1	STRING	TRUE	FALSE					FALSE				Parameter	Input	Microbial Source Module	1. Whelan et al. (2017)		1 (1,6-17,27-32,35,36)		1 (12,17)
SDMPBOutput	Area	Areas associated with each land use type (LandUse) per subwatershed (Subwatershed)	2	FLOAT	FALSE	TRUE	0.001	1.E+38	Area	Acre	TRUE	SDMPBOutput.Subwatershed	SDMPBOutput.LandUse		Independent	Input	Microbial Source Module	1. Whelan et al. (2017)		1 (6,7,9-17,23, 39)	Algebraic	1 (1,6,9-14,16)
SDMPBOutput	AreaFraction	Fraction of the Urbanized Area contributed by the four urbanized types (Urbanized) per subwatershed (Subwatershed), land use type (LandUse), and urbanized type (Urbanized) (i.e., ratio of the area associated with each urbanized type and total urbanized area). Fractions must total 1.0.	3	FLOAT	FALSE	TRUE	0.001	1	Ratio	Fraction	TRUE	SDMPBOutput.Subwatershed	SDMPBOutput.LandUse	SDMPBOutput.Urbanized	Independent	Input	Microbial Source Module	1. Whelan et al. (2017)	1 (39)	1 (23)	Algebraic	1 (1,6,9-14,16)
SDMPBOutput	NumberOfAnimals	Number of domestic animals (Agricultural) per subwatershed (Subwatershed)	2	FLOAT	FALSE	TRUE	0	1.E+38	Number	Animals	TRUE	SDMPBOutput.Subwatershed	SDMPBOutput.Agricultural		Independent	Input	Microbial Source Module	1. Whelan et al. (2017)	1 (37)	1 (6,7,9-17,32)	Algebraic	1 (9-14, 16)
SDMPBOutput	SepticNumber	Number of septic systems per subwatershed (Subwatershed)	1	Integer	FALSE	TRUE	0	10000	Number	Septics	TRUE	SDMPBOutput.Subwatershed			Independent	Input	Microbial Source Module	1. Whelan et al. (2017)	1 (38)	1 (33)	Algebraic	1 (18)
MSMInput	Application	Fraction of manure applied to soil each month (SDMPBOutput.MonthID) per domestic animal (Agricultural)	2	FLOAT	FALSE	TRUE	0	1	Ratio	Fraction	TRUE	SDMPBOutput.Agricultural	SDMPBOutput.MonthID		Independent	Input	Microbial Source Module	1. Whelan et al. (2017)		1 (1)	Algebraic	1 (9-14, 16)
MSMInput	SubUrbanizedBuiltUpRate	General microbial loading rates by sub-urbanized (SubUrbanized) category (Cells/Time/Area)	1	FLOAT	FALSE	TRUE	0	1.E+38	Cells/Area/Time	Cells/Acre/d	TRUE	SDMPBOutput.SubUrbanized			Independent	Input	Microbial Source Module	1. Whelan et al. (2017)		1 (18-21)	Algebraic	1 (1,2-10)
MSMInput	Density	Typical number of wildlife (Wildlife) per unit area by landuse (LandUse) pattern	2	FLOAT	FALSE	TRUE	0	1.E+38	Number/Area	Wildlife/Acre	TRUE	SDMPBOutput.Wildlife	SDMPBOutput.LandUse		Independent	Input	Microbial Source Module	1. Whelan et al. (2017)		1 (3)	Algebraic	1 (9-14, 16)
MSMInput	DieOff	First-order microbial inactivation/die-off rate on the land surface defined per month (SDMPBOutput.MonthID)	1	FLOAT	FALSE	TRUE	0	1.E+38	1/Time	1/d	TRUE	SDMPBOutput.MonthID			Independent	Input	Microbial Source Module	1. Whelan et al. (2017)		1 (31)	Algebraic	
MSMInput	GrazingDays	Number of grazing days per domestic animal (Agricultural) per month (SDMPBOutput.MonthID)	2	FLOAT	FALSE	TRUE	0	31	Number	d	TRUE	SDMPBOutput.Agricultural	SDMPBOutput.MonthID		Independent	Input	Microbial Source Module	1. Whelan et al. (2017)		1 (8,11-14,32)	Algebraic	1 (9-14, 16)
MSMInput	ManureIncorporatedIntoSoil	Fraction of amount of manure shed by domestic animal (Agricultural) incorporated into soil	1	FLOAT	FALSE	TRUE	0	1	Ratio	Fraction	TRUE	SDMPBOutput.Agricultural			Independent	Input	Microbial Source Module	1. Whelan et al. (2017)		1 (1,2)	Algebraic	1 (9-14, 16)
MSMInput	MicrobeAnimalProductionRates	Production or shedding rate of microbes from the domestic animal, which equals the multiple of the 1) Domestic animal shedding rate in mass of waste (ww) per time and 2) Microbial concentration based on mass of waste shed by domestic animal (Agricultural)	1	FLOAT	FALSE	TRUE	0	1.E+38	Cells/Time/Number	Cells//d/Animal	TRUE	SDMPBOutput.Agricultural			Independent	Input	Microbial Source Module	1. Whelan et al. (2017)		1 (6,7,9-17,32)	Algebraic	1 (8-14, 16)
MSMInput	MicrobeWildlifeProductionRates	Typical microbial production or shedding rate per wildlife (Wildlife)	1	FLOAT	FALSE	TRUE	0	1.E+38	Cells/Time/Number	Cells/d/Wildlife	TRUE	SDMPBOutput.Wildlife			Independent	Input	Microbial Source Module	1. Whelan et al. (2017)		1 (3)	Algebraic	1 (8,16)
MSMInput	SepticNumberPeople	Average number of people per septic system across the study area	0	FLOAT	FALSE	TRUE	0	1E+38	Number/septic	People/septic	TRUE				Independent	Input	Microbial Source Module	1. Whelan et al. (2017)		1 (33)	Algebraic	1 (18)
MSMInput	SepticConc	Typical microbial concentration in septic system waste across the study area	0	FLOAT	FALSE	TRUE	0	1.E+38	Cells/Volume	Cells/L	TRUE				Independent	Input	Microbial Source Module	1. Whelan et al. (2017)		1 (34)	Algebraic	1 (8)
MSMInput	SepticFailureRate	Typical fraction of septic systems that failure across the study area	0	FLOAT	FALSE	TRUE	0	1	Ratio	Fraction	TRUE				Independent	Input	Microbial Source Module	1. Whelan et al. (2017)		1 (33)	Algebraic	1 (8)
MSMInput	SepticOvercharge	Typical septic overcharge flow rate per person (e.g., gal/d/person)	0	FLOAT	FALSE	TRUE	0	1E+38	Volume/Time/Number	gal/d/Person	TRUE				Independent	Input	Microbial Source Module	1. Whelan et al. (2017)		1 (33)	Algebraic	1 (8)
MSMInput	TimeSpentInStreams	Fraction of the number of grazing days that a domestic animal (Agricultural) spends time in a stream per month (SDMPBOutput.MonthID)	2	FLOAT	FALSE	TRUE	0	1	Ratio	Fraction	TRUE	SDMPBOutput.Agricultural	SDMPBOutput.MonthID		Independent	Input	Microbial Source Module	1. Whelan et al. (2017)		1 (11,32)	Algebraic	1 (9-14,16)

¹Whelan, G. R. Parmar, G.F. Laniak. 2017. Microbial Source Module (MSM): Documenting the Science and Software for Discovery, Evaluation, and Integration. U.S. Environmental Protection Agency, Office of Research and Development, Athens, GA.

Table 4. Microbial Source Module Component – **Output:** Relevant Vocabulary, Taxonomy, Metadata, Syntactics, Semantics, and Ontology

Dictionary Name	Variable Name	Parameter Description	Cardinality	Data Type (Float, Integer, etc.)	Primary Key (i.e., used as a Universal parameter?)	Scaler (Not Self-Indexed (i.e., not self-enumerated) = True]	Minimum	Maximum	Measure	Unit	Stochastic (Is it allowed to change in a Monte Carlo analysis?)	Index 1	Index 2	Index 3	Parameter Type (Independent, Dependent, Parameter (e.g., Index])	Parameter Function (Input, Output, Internal)	Component	Document (Reference number with reference)	Equation in Document that Defines Variable (Reference numbers with relevant equations in parentheses)	Equations in Document that use Variable (Reference numbers with relevant equations in parentheses)	Equation Type	Relevant Assumption (Reference number with relevant assumptions in parentheses)
SDMPBOutput	Agricultural	Index on Domestic Animal (Name), self-indexed (i.e., self-enumerated)	1	STRING	TRUE	FALSE					FALSE				Parameter	Output	SDMProjectBuilder	1. Whelan et al. (2017)		1 (1,2,6-17, 32,37)		
SDMPBOutput	LandUse	Index on Land Use Type, self-indexed (i.e., self-enumerated)	1	STRING	TRUE	FALSE					FALSE				Parameter	Output	SDMProjectBuilder	1. Whelan et al. (2017)		1 (3-7,9-23, 27-31,35)		
SDMPBOutput	Subwatershed	Index on Subwatershed (Identification designation), self-indexed (i.e., self-enumerated)	1	Integer	TRUE	FALSE	0	1000			FALSE				Parameter	Output	SDMProjectBuilder	1. Whelan et al. (2017)		1 (6,7,9-23, 27-39)		1 (18)
SDMPBOutput	Urbanized	Index on type of Built up area, self-indexed (i.e., self-enumerated)	1	STRING	TRUE	FALSE					FALSE				Parameter	Output	SDMProjectBuilder	1. Whelan et al. (2017)		1 (23,39)		
SDMPBOutput	SubUrbanized	Index on Sub-urbanized Built up area, self-indexed (i.e., self-enumerated)	1	STRING	TRUE	FALSE					FALSE				Parameter	Output	SDMProjectBuilder	1. Whelan et al. (2017)		1 (18-21)		
SDMPBOutput	Wildlife	Index on Wildlife (Name), self-indexed (i.e., self-enumerated)	1	STRING	TRUE	FALSE					FALSE				Parameter	Output	SDMProjectBuilder	1. Whelan et al. (2017)		1 (3-5)		
SDMPBOutput	MonthID	Index on Month of the year (January, February, ..., December), self-indexed (i.e., self-enumerated)	1	STRING	TRUE	FALSE					FALSE				Parameter	Output	SDMProjectBuilder	1. Whelan et al. (2017)		1 (1,6-17,27-32,35,36)		1 (12,17)
SDMPBOutput	Area	Areas associated with each land use type (LandUse) per subwatershed (Subwatershed)	2	FLOAT	FALSE	TRUE	0.001	1.E+38	Area	Acre	FALSE	Subwatershed	LandUse		Dependent	Output	SDMProjectBuilder	1. Whelan et al. (2017)		1 (6,7,9-17,23, 39)	Algebraic	1 (1,6,9-14,16)
SDMPBOutput	AreaFraction	Fraction of the Urbanized Area contributed by the four urbanized types (Urbanized) per subwatershed (Subwatershed), land use type (LandUse), and urbanized type (Urbanized) (i.e., ratio of the area associated with each urbanized type and total urbanized area). Fractions must total 1.0.	3	FLOAT	FALSE	TRUE	0.001	1.E+00	Ratio	Fraction	FALSE	Subwatershed	LandUse	Urbanized	Dependent	Output	SDMProjectBuilder	1. Whelan et al. (2017)	1 (39)	1 (23)	Algebraic	1 (1,6,9-14,16)
SDMPBOutput	NumberOfAnimals	Number of domestic animals (Agricultural) per subwatershed (Subwatershed)	2	FLOAT	FALSE	TRUE	0	1.E+38	Number	Animals	FALSE	Subwatershed	Agricultural		Dependent	Output	SDMProjectBuilder	1. Whelan et al. (2017)	1 (37)	1 (6,7,9-17,32)	Algebraic	1 (9-14, 16)
SDMPBOutput	SepticNumber	Number of septic systems per subwatershed (Subwatershed)	1	Integer	FALSE	TRUE	0	10000	Number	Septics	FALSE	Subwatershed			Dependent	Output	SDMProjectBuilder	1. Whelan et al. (2017)	1 (38)	1 (33)	Algebraic	1 (18)
SDMPBOutput	PointFlow	Annual average point source discharge per subwatershed (Subwatershed)	1	FLOAT	FALSE	TRUE	0	1.E+38	Volume/Time	gal/d	FALSE	Subwatershed			Dependent	Output	SDMProjectBuilder	1. Whelan et al. (2017)	1 (38)	1 (36)	Algebraic	1 (13,14)
SDMPBOutput	PointMicrobeRate	Annual average microbial loadings to the stream associated with the annual average point source discharge per subwatershed (Subwatershed)	1	FLOAT	FALSE	TRUE	0	1.E+38	Cells/Time	Cells/yr	FALSE	Subwatershed			Dependent	Output	SDMProjectBuilder	1. Whelan et al. (2017)	1 (39)	1 (35)	Algebraic	1 (2,13,14)
SDMPBOutput	PointChemRate	Annual average chemical loadings to the stream associated with the annual average point source discharge per subwatershed (Subwatershed)	1	FLOAT	FALSE	TRUE	0	1.E+38	Mass/Time	Lbs/yr	FALSE	Subwatershed			Dependent	Output	SDMProjectBuilder	1. Whelan et al. (2017)				

¹Whelan, G. R. Parmar, G.F. Laniak. 2017. Microbial Source Module (MSM): Documenting the Science and Software for Discovery, Evaluation, and Integration. U.S. Environmental Protection Agency, Office of Research and Development, Athens, GA.

Table 5. Glossary of Microbial Source Module Input Variables [Descriptors in parentheses refer to indices outlined in [Table 7](#).]

Index	Definition
Application	Fraction of annual manure applied each month (MonthID) by domestic animal (Agricultural) [equivalent to the ratio of cells applied each month to cells applied per year] (Ratio)
Area	Areas associated with each land use type (LandUse) per subwatershed (Subwatershed)
AreaFraction	Fraction of the Urbanized Area attributed to each sub-urbanized area by subwatershed (Subwatershed) (i.e., ratio of the urbanized subcategory area and total urbanized area). Fractions must total 1.0.
Density	Typical number of wildlife (Wildlife) per area by land-use type (LandUse)
DieOff	First-order microbial inactivation/die-off rate on the land surface defined per month (MonthID)
GrazingDays	Number of grazing days per month (MonthID) by domestic animal (Agricultural)
ManureIncorporatedIntoSoil	Fraction of applied manure incorporated into the soil by domestic animal (Agricultural)
MicrobeAnimalProductionRates	Daily microbial production rate shed per domestic animal (Agricultural) [equals the multiple of domestic animal shedding rate of waste in mass of wet weight (ww) per time, and microbial density (concentration) based on mass of waste shed by domestic animal]
MicrobeWildlifeProductionRates	Typical Microbial shedding rate per wildlife (Wildlife)
NumberOfAnimals	Number of domestic animals (Agricultural) associated with Subwatershed indexed by (Subwatershed)
SepticNumberPeople	Average number of people per septic system across the study area
SepticConc	Typical microbial concentration in septic system waste across the study area
SepticFailureRate	Typical fraction of septic systems that fail across the study area
SepticNumber	Number of septic systems per subwatershed (Subwatershed)
SepticOvercharge	Typical septic overcharge flow rate per person (e.g., gal/d/person)
SubUrbanizedBuiltUpRate	General microbial loading rates by sub-urbanized (SubUrbanized) category
TimeSpentInStreams	Fraction of grazing days that a domestic animal (Agricultural = BeefCattle) spends time in a stream each month (MonthID)

- Cardinality [Number of elements in a set or grouping, as a property of that parameter/variable (dimensions). For example, if the variable “Area” (see [Tables 3 and 5](#)) is a function of its location (subwatershed) and land-use type (LandUse) (see [Table 3](#)), it has a cardinality of 2 and Subwatershed and LandUse (see [Tables 3 and 7](#)) will be classified as parameters (versus variables).
- Data Type (String, Float, Integer, Logical)

- Primary Key [Parameters/Variables that can be identified and defined only once in a workflow ontology, so that the universal parameter/variable is equally recognized by all components within a workflow, so that the universal parameter/variable is equally recognized by all components within a workflow (e.g., when all components use the same time reference)]
- Scaler [If TRUE, the variable is not part of a list. If FALSE, it is part of a list and is considered self-indexed (a function of itself) or self-enumerated (specified one after another). For example, a time series is typically self-enumerated, so the first time is indexed to 1, the second time to 2, etc. Self-indexing (i.e., being non-scaler) increases the parameter/variable cardinality by one.]
- Parameter/Variable Range (Minimum and Maximum)
- Measure (Categorizes a collection of units that inherit the same measuring properties; for example, meter, foot, and yard are units for the Measure “length.”)
- Parameter/Variable Units (Scaling properties within the same measure.)
- Stochastic (Identifies parameters/variables available for statistical manipulation, such as Monte Carlo simulation)
- Indices (Elements in a set or grouping, as a property of that parameter/variable; see [Table 7](#))
- Parameter/Variable Type (Independent, Dependent, Parameter, or Constant)
- Parameter/Variable Function (Input, Output, Internal: whether the parameter/variable represents input, output, or is associated with linking input to output)
- Component (Identifies the component that defines the parameter/variable)
- Document (Identifies the document related to the parameter’s/variable’s descriptions, equations, and assumptions)
- Equation in Document that Defines Parameter/Variable
- Equations in Document that use Parameter/Variable
- Equation Type (Algebraic, Differential, or Integral)
- Relevant Assumption (Assumptions that impact the parameter’s/variable’s use and/or value)

Table 6. Glossary of Microbial Source Module Output Variables [Descriptors in parentheses refer to indices outlined in [Table 7](#).]

Index	Definition
AccumulationRateMonth	Rate of microbial accumulation per area without die-off on the land surface by land-use type (LandUse) by month (MonthID) by subwatershed (Subwatershed) across all domestic animals (Agricultural) and wildlife (Wildlife) (a.k.a. ACQOP-Month in HSPF)
BeefCattleStreamMicrobeRate	Microbial loading rate of beef cattle (Agricultural=BeefCattle) shedding into a stream by subwatershed (Subwatershed) by month (MonthID)
SepticStreamFlowRate	Average septic flow rate to the stream by subwatershed (Subwatershed)
SepticStreamLoadingRate	Microbial loading rate to the stream from leaking septic systems by subwatershed (Subwatershed)
StorageLimitMonth	Maximum microbial storage per area by subwatershed (Subwatershed) by month (MonthID) by land-use type (LandUse), summed across all domestic animals (Agricultural) and wildlife (Wildlife), adjusted for die-off (removal) (a.k.a. SQOLIM-Month in HSPF)

Table 7. List of indices associated with Parameters

Index	Definition
Agricultural	Domestic Animal Name. There are seven domestic animal name designations: <ul style="list-style-type: none"> • DairyCow: Dairy Cow • BeefCattle: Beef Cattle • Swine • Poultry • Horse • Sheep • OtherAgAnimal: Other Agricultural Animal
LandUse	Land-use type. There are four land-use type designations: <ul style="list-style-type: none"> • Forest • Cropland • Pasture • Urbanized (a.k.a. Builtup)
MonthID	Name of the Month: January, February, March, April, May, June, July, August, September, October, November, December
SubUrbanized	Sub-urbanized Built up area. There are five name designations: <ul style="list-style-type: none"> • Commercial • SingleFamilyLowDensity: Single Family Low Density • SingleFamilyHighDensity: Single Family High Density • MultiFamilyResidential: Multi-family Residential • Road
Subwatershed	Subwatershed Identification designation
Urbanized	Urbanized or Builtup areas. There are four Urbanized designations: <ul style="list-style-type: none"> • CommercialAndServices: Commercial and Services • Residential • MixedUrban: Mixed Urban • TransportationCommunicationUtilities: Transportation, Communication, Utilities
Wildlife	Wildlife Name: There are six wildlife name designations: <ul style="list-style-type: none"> • Duck • Goose • Deer • Beaver • Raccoon • OtherWildlife: Other Wildlife

[4.5.3.2 Indices](#)

The first seven parameters listed in [Table 3](#) correspond to the seven indices outlined in [Table 7](#) (i.e., Agricultural, LandUse, Subwatershed, Urbanized, Wildlife, and MonthID) upon which other parameters and variables are dependent. If a parameter/variable has an index, as illustrated in [Tables 3 and 4](#), that parameter/variable is a function of that index (i.e., another parameter). For example, microbial die-off (DieOff in [Tables 3 and 5](#)) is a function of the month of the year (January, February, ..., December; as

captured with MonthID in [Tables 3 and 7](#)); hence, DieOff has 12 associated values, one for each month [i.e., DieOff(MonthID)]:

DieOff(January)
DieOff(February)
•
•
•
DieOff(December)

MonthID is a parameter but also an index. Each index may, therefore, be described by one or more elements: MonthID has 12, LandUse has four (Forest, Cropland, Pasture, and Urbanized), etc. Indices and their assigned elements are reported in [Table 7](#).

Indices organize the dimensionality of a system by providing hierarchical relationships (i.e., context) between variables and parameters, supporting the concept of semantics (see [Table 1](#)). [Whelan et al. \(2014a\)](#) note that semantics refers to the meaning of data and their relationship to other data, including indices, by relating content and representation of information resources to entities and concepts in the real world ([Meersman and Mark, 1997](#); [Wang et al., 2009](#)).

Some parameters/variables may be a function of multiple indices, such as the variable “Area” (see [Tables 3 and 5](#)), which is function of its location (i.e., Subwatershed) and land-use type (i.e., LandUse) (see [Tables 3 and 7](#)). When a parameter/variable is a function of multiple indices, a hierarchical relationship exists between multiple indices (i.e., one index is essentially contained within another). For example, the variable “TimeSpentInStreams” (see [Table 3](#)) is defined with the indices of Agricultural and MonthID; thus, there will be a value for “TimeSpentInStreams” for each combination of Agricultural and MonthID; a relationship that can be expressed as:

TimeSpentInStreams (Agricultural,MonthID)
or
TimeSpentInStreams (MonthID,Agricultural)

In this case, the list of values remains the same, and the order in which they are referenced, using indices, is simply reversed. Both expressions are valid, although it is desirable to establish a consistent ordering of indices to facilitate software and documentation development. The following logic was used to prioritize the order of indices for MSM parameters and variables: Subwatershed, Agricultural, Wildlife, LandUse, Urbanized, SubUrbanized, and MonthID. All ontological metadata contained in tables, such as [Tables 3 and 4](#), prioritize their indices (Index 1 to Index 3) in this order.

A glossary of indices defining associations between variables and parameters by identifying their correlations, which help to define metadata associated with input and output variables, are provided in [Table 7](#), and [Tables 5 and 6](#) provide glossaries of the MSM input and output parameters/variables, respectively. The glossaries are intended to be easy look-up tables.

4.5.4 Equation Class

The Equation Class describes all equations used by the MSM component, translating information from input to output. The purpose is to cross-correlate input, output, and internal variables; equations using

or defining the variables; and associated assumptions. Internal variables refer to those used within the mathematical formulations, not consumed as input or produced as output. This section is subdivided as follows:

- Summary of Assumptions and Constraints impacting the variables and their use within the MSM
- Domestic Animal Waste Available for Land Application and Wildlife Shedding Rates, Calculations associated with Domestic Animal Waste available for Land Application and Wildlife Shedding Rates
- Accumulated Microbial Loading Rates on Cropland
- Accumulated Microbial Loading Rates on Pasture
- Accumulated Microbial Loading Rates on Forest
- Accumulated Microbial Loading Rates on Urbanized Areas
- Accumulated Overland Microbial Loading Rates to the Land Surface, Adjusted for Die-off
- Microbial Point Source Loading Rates

4.5.4.1 Summary of Assumptions and Constraints

1. The MSM considers only one microbe at a time and must be individually executed.
2. Overland microbial loading rates, accounting for die-off, are computed for each subwatershed by land-use type on a monthly basis.
3. The MSM considers microbial loadings from sources correlated to four land-use types for each subwatershed, where the subwatershed is the smallest area associated with watershed modeling: 1) Cropland: Land application of some domestic animal waste (Beef Cattle, Dairy Cow, Swine, and/or Poultry) and Wildlife shedding; 2) Pasture: Some domestic animal grazing with shedding (Beef Cattle, Horse, Sheep, and/or Other), Land application of some domestic animal waste (Beef Cattle, Dairy Cow, and/or Horse), and Wildlife shedding; 3) Forest: Wildlife shedding; and 4) Built: Urban-related releases: Commercial and Services, Residential, Mixed Urban, Transportation, and Communication, Utilities.
4. The MSM considers instream beef cattle shedding, where loading rates are identified with each subwatershed.
5. The MSM currently assumes that manure loadings from land application and shedding are computed monthly and represent a typical year.
6. The land-use types associated with the National Land Cover Database (NLCD) are consolidated into Cropland, Pastureland, Forest, and Urbanized, providing a more manageable modeling set when land use is the index, since supporting data for finer granularity are not available.
7. Urbanized land is subdivided into Commercial and Services; Mixed Urban or Built-Up; Residential; and Transportation, Communications, and Utilities. A single, weighted Urbanized loading rate is quantified for each subwatershed (constant value for the year) based on all individual Urbanized land uses present. Each Urbanized category considers a weighted combination of the following five attributes: Commercial, Single-family low density, Single-family high density, Multi-family Residential, and Road. The combinations per Urbanized categories are as follows:
 - Commercial and Services: Commercial
 - Mixed Urban or Built-up: Average microbial accumulation rates for Road, Commercial, Single-family low density, Single-family high density, and Multi-family residential
 - Residential: Average microbial accumulation rates for Single-family low density, Single-family high density, and Multi-family residential

- Transportation, Communications, and Utilities
- 8. Fecal shedding from animals is used for microbial loading estimates to all land-use types except Urbanized.
- 9. Manures from Swine and Poultry are assumed to be collected and applied to Cropland.
- 10. Beef Cattle/Dairy Cow manure is assumed to be applied only to Cropland and Pastureland by the same method.
- 11. Dairy Cows are only kept in feedlots; therefore, all of their waste is used for manure application, divided equally between Cropland and Pastureland.
- 12. Beef Cattle are kept in feedlots or allowed to graze. During grazing, a specified percentage of cattle also have direct access to streams; therefore, Beef Cattle waste is either applied as manure to Cropland and Pastureland, or contributes directly to Pasture (shedding) or Streams (shedding). Direct contribution of microbes from Beef Cattle to a stream through shedding is thus represented as a monthly point source. Dairy Cows are not allowed to graze and, therefore, do not have access to streams.
- 13. Horse manure not deposited in Pastureland during grazing is assumed to be collected and applied to Pastureland.
- 14. Manures from Beef Cattle, Horses, Sheep, and Other domestic animals are assumed to contribute to Pastureland in proportion to time spent grazing. Sheep and Other domestic animal manures not deposited to Pastureland during grazing are assumed to be collected and treated or transported out of the watershed and not contributing to any loading.
- 15. Domestic animal designations are designed as placeholders to differentiate grazing and non-grazing animals by land-use type and manure application (land-applied versus direct shedding). For example, if Dairy Cows graze and/or shed directly to the stream, then the user would have to reclassify them as Beef Cattle.
- 16. Wildlife densities are provided for all land uses except Built-up and assumed to be the same in all subwatersheds. The wildlife population is the only microbial contributor considered to Forest.
- 17. The fraction of annual domestic animal manure application available for runoff each month (EPA, 2013b, 2013c)
 - = [Fraction of manure applied] * {1 - [Fraction of manure incorporated] / 3} for poultry
 - = [Fraction of manure applied] * {1 - [Fraction of manure incorporated] / 2} for other domestic animals (dairy cow, beef cattle, swine, and horse)
- 18. One input time series for direct input to streams is allowed per subwatershed; multiple septic and instream shedding are each aggregated separately, then combined to provide monthly loadings.

Assumptions and constraints that correlate manure application with land-use type by domestic animal and wildlife are summarized in Table 8. An index glossary that correlates subscripts used within the mathematical formulas is provided in Table 9. Subscripts relate to the indices associated with the MSM parameters, as summarized in Table 7; included is an index on the microbe, which accounts for indicator bacteria, pathogen bacteria, protozoa, and viruses:

Microbial Name (Name)

- Indicator Bacteria: *E. coli*, *Enterococci*, *Clostridium perfringens*, Fecal Coliforms, *Bacteroides*
- Pathogen Bacteria: *Salmonella spp.*, *Campylobacter jejuni*, *E. coli O157:H7*, *Listeria*, *Mycobacterium avium paratuberculosis*
- Pathogen Protozoa: *Cryptosporidium parvum*, *Giardia lamblia*, *Toxoplasma gondii*
- Pathogen Viruses: Enterovirus, Rotavirus, Adenovirus, Norovirus

Table 8. Correlation of Manure Application with Land-use type by Domestic Animal and Wildlife

Manure Application Correlated to Land Use	Domestic Animals and Wildlife							
	BeefCattle	DairyCow	Swine	Poultry	Horse	Sheep	Other	Wildlife
Cropland Grazing/Shedding								x
Pasture Grazing/Shedding	x				x	x	x	x
Forest Shedding								x
In Stream Shedding	x							
Cropland Application	x	x	x	x				
Pasture Application	x	x			x			

Notes:

1. Any domestic animal "Application" has a complementing value for "ManureIncorporatedIntoSoil."
2. All domestic animals and wildlife have production rates associated with them (i.e., "MicrobeAnimalProductionRates" and "MicrobeWildlifeProductionRates," respectively).

Table 9. Index Glossary used in the Mathematical Formulations

Index	Description
i	Subwatershed ID
k	Microbe (1 = <i>E. coli</i> , 2 = <i>Enterococci</i> , 3 = <i>Clostridium perfringens</i> , 4 = Fecal Coliforms, 5 = <i>Bacteroides</i> , 6 = <i>Salmonella spp.</i> , 7 = <i>Campylobacter jejuni</i> , 8 = <i>E. coli O157:H7</i> , 9 = <i>Listeria</i> , 10 = <i>Mycobacterium avium paratuberculosis</i> , 11 = <i>Cryptosporidium parvum</i> , 12 = <i>Giardia lamblia</i> , 13 = <i>Toxoplasma gondii</i> , 14 = Enterovirus, 15 = Rotavirus, 16 = Adenovirus, 17 = Norovirus)
ℓ	Land-use type (1 = Cropland, 2 = Pasture, 3 = Forest, 4 = Urbanized)
m	Domestic Animal [1 = Dairy Cow (DairyCow), 2 = Beef Cattle (BeefCattle), 3 = Swine, 4 = Poultry, 5 = Horse, 6 = Sheep, 7 = Other Agricultural Animal (OtherAgAnimal)]
n	Wildlife (1 = Duck, 2 = Goose, 3 = Deer, 4 = Beaver, 5 = Raccoon, 6 = Other Wildlife)
q	Month of the year (January to December)
r	Urbanized category (1 = Commercial and Services; 2 = Mixed Urban or Built-Up; 3 = Residential; and 4 = Transportation, Communications and Utilities)
u	Sub-urbanized category (1 = Commercial, 2 = Single Family Low Density, 3 = Single Family High Density, 4 = Multi-family Residential, 5 = Road)

Although the microbial name is not needed by the MSM because it handles only one microbe at a time, it is presented in the mathematical formulations for completeness. Names are also included in the formulas because other modules within a workflow may contain parameters/variables that are a function of the microbial name. A glossary of internal variables used to link input and output variables is presented in [Table 10](#), and the corresponding ontological dictionary (similar to [Tables 3 and 4](#)) is presented in [Table 11](#).

4.5.4.2 Domestic Animal Waste Available for Land Application and Wildlife Shedding Rates

4.5.4.2.1 Domestic Animal Waste Available for Runoff

The fraction of annual manure application available for runoff each month by domestic animal, based on the monthly fraction applied and incorporated into the soil, is computed as follows:

$$\text{FractionManureAvailableRunoff}_{m,q} = (\text{Application}_{m,q}) [1 - (\text{ManureIncorporatedIntoSoil}_m) / 2] \quad (1)$$

in which

$$\text{AnimalFractionAvailable}_m = 1 - (\text{ManureIncorporatedIntoSoil}_m) / 2 \quad \text{for } m = 1, 2, 3, \text{ or } 5 \quad (2)$$

$$\text{AnimalFractionAvailable}_m = 1 - (\text{ManureIncorporatedIntoSoil}_m) / 3 \quad \text{for } m = 4$$

where

- $\text{FractionManureAvailableRunoff}_{m,q}$ = Fraction of annual manure application available for runoff by month (q) by domestic animal (m) [equivalent to the ratio of microbial cells available for runoff each month to cells available for runoff per year] (Ratio)
- $\text{Application}_{m,q}$ = Fraction of annual manure applied each month (q) by domestic animal (m) [equivalent to the ratio of microbial cells applied each month to cells applied per year] (Ratio)
- $\text{ManureIncorporatedIntoSoil}_m$ = Fraction of applied manure incorporated into the soil by domestic animal (m) (Ratio)
- $\text{AnimalFractionAvailable}_m$ = Fraction of domestic animal (m) manure available for runoff (Ratio)

4.5.4.2.2 Wildlife Shedding Rates

Wildlife shedding is the only manure contribution to land-use type Forest ($\ell = 3$), although Wildlife also contributes to land-use types Cropland ($\ell = 1$) and Pasture ($\ell = 2$). The microbial shedding rate from Wildlife by land-use-type is:

$$\text{WildLifeMicrobeRateShed}_{k,\ell,n} = (\text{Density}_{\ell,n}) (\text{MicrobeWildlifeProductionRates}_{k,n}) \quad \text{for } \ell = 1, 2, 3 \quad (3)$$

$$\text{WildLifeMicrobeRateShed}_{k,\ell,n} = 0 \quad \text{for } \ell = 4 \quad (4)$$

where

- $\text{WildLifeMicrobeRateShed}_{k,\ell,n}$ = Microbial shedding rate per area by Wildlife (n) by microbe (k) by land-use-type (ℓ) (Cells/Time/Area)

Table 10. Glossary of Internal Variables (not including constants) [Descriptors in parentheses refer to indices outlined in [Table 7.](#)]

Index	Definition
AccumBuiltUpRate	Accumulated microbial loading rate associated with the Urbanized land-use type (LandUse = Urbanized) per subwatershed (Subwatershed), weighted by the areas associated with four Urbanized categories for all months (i.e., applicable throughout the year)
AnimalFractionAvailable	Fraction of domestic animal (Agricultural) manure available for runoff
BeefCattleMicrobeRateApply	The microbial loading rate due to manure application associated with the domestic animal Beef Cattle (Agricultural = BeefCattle) for land-use types Cropland and Pasture (LandUse = Cropland, LandUse = Pasture) by subwatershed (Subwatershed) by month (MonthID)
BeefCattleMicrobeRateShed	The microbial loading rate to land-use type (LandUse = Pasture) due to grazing of Beef Cattle (Agricultural = BeefCattle) by month (MonthID) by Subwatershed (Subwatershed)
BeefCattleStreamMicrobeRate	Microbial loading rate for domestic animal Beef Cattle (Agricultural = BeefCattle) shedding into a stream by subwatershed (Subwatershed) by month (q)
BuiltUpRate	Accumulation rates in median microbial cells by microbe per Urbanized land-use type (LandUse = Urbanized) per area per time, indexed by the Urbanized subcategories
DairyCowMicrobeRateApply	Microbial loading rate per area to Cropland (LandUse = Cropland) and Pasture (LandUse = Pasture) from land application of domestic animal (Agricultural = DairyCow) manure by month (MonthID) by subwatershed (Subwatershed)
FractionManureAvailableRunoff	Fraction of annual manure application available for runoff by month (MonthID) by domestic animal (Agricultural) [equivalent to the ratio of microbial cells available for runoff each month to cells available for runoff per year]
HorsesMicrobeRateApply	Microbial loading rate per area to land-use type Pasture (LandUse = Pasture) from land application of domestic animal Horses (Agricultural = Horses) manure by month (MonthID) by subwatershed (Subwatershed)
HorsesMicrobeRateShed	The microbial loading rate to land-use type (LandUse = Pasture) due to grazing of Horses (Agricultural = Horses) by month (MonthID) by Subwatershed (Subwatershed)
OtherAgAnimalMicrobeRateShed	The microbial loading rate to land-use type Pasture (LandUse = Pasture) due to grazing of Other Agricultural Animals (Agricultural = OtherAgAnimal) by month (MonthID) by Subwatershed Subwatershed)
PoultryMicrobeRateApply	Microbial loading rate per area by microbe to Cropland (LandUse = Cropland) from land application of domestic animal (Agricultural = Poultry) manure by month (MonthID) by subwatershed (Subwatershed)
SepticStreamFlowRate	Average septic flow rate to the stream by subwatershed (Subwatershed)
SepticStreamLoadingRate	Microbial loading rate to the stream from leaking septic systems by subwatershed (Subwatershed)
SheepMicrobeRateShed	The microbial loading rate to land-use type Pasture (LandUse = Pasture) due to grazing of Sheep (Agricultural = Sheep) by month (MonthID) by Subwatershed Subwatershed)
SwineMicrobeRateApply	Microbial loading rate per area by microbe to Cropland (LandUse = Cropland) from land application of domestic animal (Agricultural = Swine) manure by month (MonthID) by subwatershed (Subwatershed)
TotalGrazeDays	Total number of grazing days per year by agricultural domestic animal (Agricultural)
WildLifeMicrobeRateShed	Microbial shedding rate per area by wildlife (Wildlife) by land-use-type (LandUse) area
WildLifeMicrobeRateShedSum	Total microbial shedding rate per area by land-use-type (LandUse), summed across all wildlife (Wildlife)

Table 11. Microbial Source Module Component – Internally Computed Variables: Relevant Vocabulary, Taxonomy, Metadata, Syntactics, Semantics, and Ontology

Dictionary Name	Variable Name	Parameter Description	Cardinality Data Type (Float, Integer, etc.) (Privilege: 0=Input, 1=BC)	Primary Key (i.e., used as a Universal parameter?)	Scaler (Not Self-Indexed (i.e., not self-enumerated) True)	Minimum	Maximum	Measure	Unit	Stochastic (Is it allowed to change in a Monte Carlo analysis?)	Index 1	Index 2	Index 3	Parameter Type [Independent, Dependent, Parameter (e.g., Index)]	Parameter Function (Input, Output, Internal)	Component	Document (Reference number with reference)	Equation in Document that Defines Variable (Reference numbers with relevant equations in parentheses)	Equations in Document that use Variable (Reference numbers with relevant equations in parentheses)	Equation Type	Relevant Assumption (Reference number with relevant assumptions in parentheses)	
MSMInternalVariables	AccumBuiltUpRate	Accumulated microbial loading rate associated with the Urbanized land use type (LandUse = Urbanized) per subwatershed (Subwatershed), weighted by the areas associated with four Urbanized categories (i.e., Commercial and Service; Residential; Mixed Urban; Transportation, Communication, and Utilities) for all months (i.e., applicable throughout the year)	2	FLOAT	FALSE	TRUE	0	1E+38	Cells/Area/Time	Cells/ac/d	FALSE	SDMPBOutput.Subwatershed	SDMPBOutput.LandUse		Dependent	Internal	Microbial Source Module	1. Whelan et al. (2017)	1 (22,23)	1 (30)	Algebraic	1 (1,2-10)
MSMInternalVariables	AnimalFractionAvailable	Fraction of manure shed by domestic animal (Agricultural) that is applied to the land and available for runoff	1	FLOAT	FALSE	TRUE	0	1	Ratio	Fraction	FALSE	SDMPBOutput.Agricultural		Dependent	Internal	Microbial Source Module	1. Whelan et al. (2017)	1 (2)		Algebraic	1 (8-14, 16)	
MSMInternalVariables	BeefCattleMicrobeRateApply	The microbial loading rate due to manure application associated with the domestic animal Beef Cattle (Agricultural = BeefCattle) for land use types Cropland and Pasture (LandUse = Cropland, LandUse = Pasture) by subwatershed (Subwatershed) by month (MonthID)	3	FLOAT	FALSE	TRUE	0	1E+38	Cells/Area/Time	Cells/ac/d	FALSE	SDMPBOutput.Subwatershed	SDMPBOutput.LandUse	SDMPBOutput.MonthID	Dependent	Internal	Microbial Source Module	1. Whelan et al. (2017)	1 (7,16)	1 (27,28)	Algebraic	1 (8-14, 16)
MSMInternalVariables	BeefCattleMicrobeRateShed	The microbial loading rate to land use type Pasture (LandUse = Pasture) due to grazing of domestic animal Beef Cattle (Agricultural = BeefCattle) by subwatershed (Subwatershed) by month (MonthID)	3	FLOAT	FALSE	TRUE	0	1E+38	Cells/Area/Time	Cells/ac/d	FALSE	SDMPBOutput.Subwatershed	SDMPBOutput.LandUse	SDMPBOutput.MonthID	Dependent	Internal	Microbial Source Module	1. Whelan et al. (2017)	1 (11)	1 (28)	Algebraic	1 (8-14, 16)
MSMInternalVariables	BeefCattleStreamMicrobeRate	Microbial loading rate of domestic animal BeefCattle (Agricultural = BeefCattle) shedding into a stream by subwatershed (Subwatershed) by month (MonthID)	2	FLOAT	FALSE	TRUE			Cells/Time	Cells/d	FALSE	SDMPBOutput.Subwatershed	SDMPBOutput.MonthID		Dependent	Internal	Microbial Source Module	1. Whelan et al. (2017)	1 (32)	1 (35)	Algebraic	1 (9-14,16)
MSMInternalVariables	BuiltUpRate	Accumulation rates in median Cells per area per time per land use type (LandUse) and urbanized type (Urbanized) (i.e., built up areas)	2	FLOAT	FALSE	TRUE	0	1E+38	Cells/Area/Time	Cells/ac/d	TRUE	SDMPBOutput.LandUse	SDMPBOutput.Urbanized		Dependent	Internal	Microbial Source Module	1. Whelan et al. (2017)	1 (18-21)	1 (27)	Algebraic	1 (3,6,7-14,16)
MSMInternalVariables	DairyCowMicrobeRateApply	The microbial loading rate due to manure application associated with the domestic animal Dairy Cow (Agricultural = DairyCow) for land use types Cropland and Pasture (LandUse = Cropland, LandUse = Pasture) by subwatershed (Subwatershed) by month (MonthID)	3	FLOAT	FALSE	TRUE	0	1E+38	Cells/Area/Time	Cells/ac/d	FALSE	SDMPBOutput.Subwatershed	SDMPBOutput.LandUse	SDMPBOutput.MonthID	Dependent	Internal	Microbial Source Module	1. Whelan et al. (2017)	1 (6,15)	1 (27,28)	Algebraic	1 (8-14, 16)
MSMInternalVariables	FractionManureAvailableRunoff	Fraction of annual manure from domestic animal (Agricultural) applied to the land surface that is available for runoff each month (MonthID)	2	FLOAT	FALSE	TRUE	0	1	Ratio	Fraction	FALSE	SDMPBOutput.Agricultural	SDMPBOutput.MonthID		Dependent	Internal	Microbial Source Module	1. Whelan et al. (2017)	1 (1)	1 (6,7,9,10, 15-17)	Algebraic	1 (8-14, 16)
MSMInternalVariables	HorsesMicrobeRateApply	The microbial loading rate due to manure application associated with the domestic animal Horses (Agricultural = Horses) for land use type Pasture (LandUse = Pasture) by subwatershed (Subwatershed) by month (MonthID)	3	FLOAT	FALSE	TRUE	0	1E+38	Cells/Area/Time	Cells/ac/d	FALSE	SDMPBOutput.Subwatershed	SDMPBOutput.LandUse	SDMPBOutput.MonthID	Dependent	Internal	Microbial Source Module	1. Whelan et al. (2017)	1 (17)	1 (28)	Algebraic	1 (8-14, 16)
MSMInternalVariables	HorsesMicrobeRateShed	The microbial loading rate to land use type Pasture (LandUse = Pasture) due to manure application of domestic animal Horses (Agricultural = Horses) by subwatershed (Subwatershed) by month (MonthID)	3	FLOAT	FALSE	TRUE	0	1E+38	Cells/Area/Time	Cells/ac/d	FALSE	SDMPBOutput.Subwatershed	SDMPBOutput.LandUse	SDMPBOutput.MonthID	Dependent	Internal	Microbial Source Module	1. Whelan et al. (2017)	1 (12)	1 (28)	Algebraic	1 (8-14, 16)
MSMInternalVariables	OtherAgAnimalMicrobeRateShed	The microbial loading rate to land use type Pasture (LandUse = Pasture) due to manure application of Other Agricultural Animals (Agricultural = OtherAgAnimal) by subwatershed (Subwatershed) by month (MonthID)	3	FLOAT	FALSE	TRUE	0	1E+38	Cells/Area/Time	Cells/ac/d	FALSE	SDMPBOutput.Subwatershed	SDMPBOutput.LandUse	SDMPBOutput.MonthID	Dependent	Internal	Microbial Source Module	1. Whelan et al. (2017)	1 (14)	1 (28)	Algebraic	1 (8-14, 16)
MSMInternalVariables	PoultryMicrobeRateApply	The microbial loading rate due to manure application associated with the domestic animal Poultry (Agricultural = Poultry) for land use type Cropland (LandUse = Cropland) by subwatershed (Subwatershed) by month (MonthID)	3	FLOAT	FALSE	TRUE	0	1E+38	Cells/Area/Time	Cells/ac/d	FALSE	SDMPBOutput.Subwatershed	SDMPBOutput.LandUse	SDMPBOutput.MonthID	Dependent	Internal	Microbial Source Module	1. Whelan et al. (2017)	1 (9)	1 (27)	Algebraic	1 (8-14, 16)
MSMInternalVariables	SheepMicrobeRateShed	The microbial loading rate to land use type Pasture (LandUse = Pasture) due to manure application of domestic animal Sheep (Agricultural = Sheep) by subwatershed (Subwatershed) by month (MonthID)	3	FLOAT	FALSE	TRUE	0	1E+38	Cells/Area/Time	Cells/ac/d	FALSE	SDMPBOutput.Subwatershed	SDMPBOutput.LandUse	SDMPBOutput.MonthID	Dependent	Internal	Microbial Source Module	1. Whelan et al. (2017)	1 (13)	1 (28)	Algebraic	1 (8-14, 16)
MSMInternalVariables	SwineMicrobeRateApply	The microbial loading rate to land use type Cropland (LandUse = Cropland) due to manure application of domestic animal Swine (Agricultural = Swine) by subwatershed (Subwatershed) by month	3	FLOAT	FALSE	TRUE	0	1E+38	Cells/Area/Time	Cells/ac/d	FALSE	SDMPBOutput.Subwatershed	SDMPBOutput.LandUse	SDMPBOutput.MonthID	Dependent	Internal	Microbial Source Module	1. Whelan et al. (2017)	1 (10)	1 (27)	Algebraic	1 (8-14, 16)
MSMInternalVariables	TotalGrazedDays	Total number of grazing days per year by agricultural domestic animal (Agricultural)	1	FLOAT	FALSE	TRUE	0	365	Number	d	FALSE	SDMPBOutput.Agricultural		Dependent	Internal	Microbial Source Module	1. Whelan et al. (2017)	1 (8)	1 (7,16,17)	Algebraic	1 (9-14, 16)	
MSMInternalVariables	WildLifeMicrobeRateShed	Microbial shedding rate by wildlife (Wildlife) by land use type (LandUse) area	2	FLOAT	FALSE	TRUE	0	1E+38	Cells/Area/Time	Cells/ac/d	FALSE	SDMPBOutput.Wildlife	SDMPBOutput.LandUse		Dependent	Internal	Microbial Source Module	1. Whelan et al. (2017)	1 (3,4)	1 (5)	Algebraic	1 (8,16)
MSMInternalVariables	WildLifeMicrobeRateShedSum	Total microbial shedding rate per land use type (LandUse) area, summed across all wildlife (Wildlife)	1	FLOAT	FALSE	TRUE	0	1E+38	Cells/Area/Time	Cells/ac/d	FALSE	SDMPBOutput.LandUse		Dependent	Internal	Microbial Source Module	1. Whelan et al. (2017)	1 (5)	1 (27-29)	Algebraic	1 (8,16)	

¹Whelan, G. R. Parmar, G.F. Laniak. 2017. Microbial Source Module (MSM): Documenting the Science and Software for Discovery, Evaluation, and Integration. U.S. Environmental Protection Agency, Office of Research and Development, Athens, GA.

- $Density_{\ell,n}$ = Typical number of wildlife (n) per unit area by land-use type (ℓ) (Number of wildlife/Area)
- $MicrobeWildlifeProductionRates_{k,n}$ = Microbial shedding rate per microbe (k) per wildlife (n) (Cells/Time/Wildlife)

The total microbial shedding rate per land-use-type area by microbe, summed across all wildlife is:

$$WildLifeMicrobeRateShedSum_{k,\ell} = \sum_n WildLifeMicrobeRateShed_{k,\ell,n} \quad (5)$$

where

- $WildLifeMicrobeRateShedSum_{k,\ell}$ = Total microbial shedding rate per area by land-use-type (ℓ) by microbe (k), summed across all wildlife (n) (Cells/Time/Area)

4.5.4.3 Accumulated Microbial Loading Rates on Cropland

This section describes calculations to determine the accumulated microbial loading rate on Cropland, by month, by subwatershed area, by domestic animal, due to manure application (i.e., non-grazing) and wildlife shedding to the land surface.

4.5.4.3.1 Wildlife

The microbial loading rate to Cropland ($\ell = 1$) due to shedding per microbe (k), associated with all Wildlife, is equal to $WildLifeMicrobeRateShedSum_{k,\ell=1}$, with units of Cells/Time/Area.

4.5.4.3.2 Dairy Cow

The microbial loading rate to land-use type Cropland ($\ell = 1$), which is the same loading rate to land-use type Pasture ($\ell = 2$), due to manure application associated with domestic animal Dairy Cow ($m = 1$) by month by subwatershed, is equal to:

$$DairyCowMicrobeRateApply_{k,i,\ell=1,q} = (NumberOfAnimals_{i,m=DairyCow}) (MicrobeAnimalProductionRates_{k,m=DairyCow}) (FractionManureAvailableRunoff_{m=DairyCow,q}) (365 / DayInMonth_q) / (Area_{i,\ell=1} + Area_{i,\ell=2}) \quad (6)$$

where

- $DairyCowMicrobeRateApply_{k,i,\ell=1,q}$ = Microbial loading rate of microbe (k) per area to land-use type (ℓ) from land application of domestic animal Dairy Cow ($m = 1$) manure by month (q) by subwatershed (i) (Cells/Time/Area)
- $NumberOfAnimals_{i,m=DairyCow}$ = Number of domestic animals (m) associated with Subwatershed indexed by (i) (Number of domestic animals)
- $MicrobeAnimalProductionRates_{k,m=DairyCow}$ = Daily microbial production rate of microbe (k) shed per domestic animal (m) (Cells/d/domestic animal) [equals the multiple of domestic animal shedding rate of waste in mass of wet weight (ww) per time (Mass/d/domestic animal), and microbial density (concentration) based on mass of waste shed by domestic animal (Cells/Mass)]
- 365 = Conversion constant for days in a year

- DayInMonth_q = Conversion constant by month for days per month with January = 31, February = 28, ..., December = 31, in which the months are indexed (q) as 1 = January, ..., 12 = December (d/mo)
- $\text{Area}_{i,\ell=1}$ = Area associated the Cropland ($\ell = 1$) land-use type for Subwatershed (i) (Area)
- $\text{Area}_{i,\ell=2}$ = Area associated the Pasture ($\ell = 2$) land-use type for Subwatershed (i) (Area)

4.5.4.3.3 Beef Cattle

The microbial loading rate to land-use type Cropland ($\ell = 1$), which is the same loading rate to land-use type Pasture ($\ell = 2$), due to manure application associated with domestic animal Beef Cattle ($m = 2$) by month by subwatershed is equal to:

$$\text{BeefCattleMicrobeRateApply}_{k,i,\ell=1,q} = (\text{NumberOfAnimals}_{i,m=\text{BeefCattle}}) (\text{MicrobeAnimalProductionRates}_{k,m=\text{BeefCattle}}) (\text{FractionManureAvailableRunoff}_{m=\text{BeefCattle},q}) [(365 - \text{TotalGrazingDays}_{m=\text{BeefCattle}}) / \text{DayInMonth}_q] / (\text{Area}_{i,\ell=1} + \text{Area}_{i,\ell=2}) \quad (7)$$

in which

$$\text{TotalGrazingDays}_m = \sum_q \text{GrazingDays}_{m,q} \quad (8)$$

where

- $\text{BeefCattleMicrobeRateApply}_{k,i,\ell=1,q}$ = Microbial loading rate of microbe (k) per area to land-use type (ℓ) from land application of domestic animal Beef Cattle ($m = 2$) manure by month (q) by subwatershed (i) (Cells/Time/Area)
- $\text{GrazingDays}_{m=\text{BeefCattle},q}$ = Number of grazing days per month (q) by domestic animal (m) (d/mo)
- $\text{TotalGrazingDays}_{m=\text{BeefCattle}}$ = Total number of grazing days per year for domestic animal (m) (d/yr)

4.5.4.3.4 Poultry

The microbial loading rate to land-use type Cropland ($\ell = 1$), due to manure application associated with domestic animal Poultry ($m = 4$) by month by subwatershed, is equal to:

$$\text{PoultryMicrobeRateApply}_{k,i,\ell=1,q} = (\text{NumberOfAnimals}_{i,m=\text{Poultry}}) (\text{MicrobeAnimalProductionRates}_{k,m=\text{Poultry}}) (\text{FractionManureAvailableRunoff}_{m=\text{Poultry},q}) (365 / \text{DayInMonth}_q) / (\text{Area}_{i,\ell=1}) \quad (9)$$

where

- $\text{PoultryMicrobeRateApply}_{k,i,\ell=1,q}$ = Microbial loading rate per area of microbe (k) to land-use type Cropland ($\ell = 1$) from land application of domestic animal Poultry ($m = 4$) manure by month (q) by subwatershed (i) (Cells/Time/Area)

4.5.4.3.5 Swine

The microbial loading rate to land-use type Cropland ($\ell = 1$), due to manure application associated with domestic animal Swine ($m = 3$) by month by subwatershed, is equal to:

$$\text{SwineMicrobeRateApply}_{k,i,\ell=1,q} = (\text{NumberOfAnimals}_{i,m=\text{Swine}}) (\text{MicrobeAnimalProductionRates}_{k,m=\text{Swine}}) (\text{FractionManureAvailableRunoff}_{m=\text{Swine},q}) (365 / \text{DayInMonth}_q) / (\text{Area}_{i,\ell=1}) \quad (10)$$

where

- $\text{SwineMicrobeRateApply}_{k,i,\ell=1,q}$ = Microbial loading rate per area of microbe (k) to land-use type Cropland ($\ell = 1$) from land application of domestic animal Swine ($m = 3$) manure by month (q) by subwatershed (i) (Cells/Time/Area)

[4.5.4.4 Accumulated Microbial Loading Rates on Pasture](#)

This section describes calculations to determine accumulated microbial loading rate by month by subwatershed area by animal or wildlife for Pasture due to animal shedding (i.e., grazing) and manure application to the land surface (i.e., non-grazing).

[4.5.4.4.1 Shedding to Land Surface](#)

[4.5.4.4.1.1 Wildlife](#)

The microbial loading rate of per microbe (k) to Pasture ($\ell = 2$) associated with all Wildlife (n) is equal to $\text{WildLifeMicrobeRateShedSum}_{k,\ell=2}$ (see Equation 5).

[4.5.4.4.1.2 Beef Cattle](#)

The microbial loading rate to land-use type Pasture ($\ell = 2$), due to grazing associated with domestic animal Beef Cattle ($m = 2$) by month by subwatershed, is equal to:

$$\text{BeefCattleMicrobeRateShed}_{k,i,\ell=2,q} = (\text{NumberOfAnimals}_{i,m=\text{BeefCattle}}) (\text{MicrobeAnimalProductionRates}_{k,m=\text{BeefCattle}}) (\text{GrazingDays}_{m=\text{BeefCattle},q}) (1 - \text{TimeSpentInStreams}_{m=\text{BeefCattle},q}) / [(\text{DayInMonth}_q) (\text{Area}_{i,\ell=2})] \quad (11)$$

where

- $\text{BeefCattleMicrobeRateShed}_{k,i,\ell=2,q}$ = The microbial loading rate of microbe (k) to land-use type Pasture ($\ell = 2$) due to grazing of domestic animal Beef Cattle ($m = 2$) by month (q) by Subwatershed (i) (Cells/Time/Area)
- $\text{TimeSpentInStreams}_{m=\text{BeefCattle},q}$ = Fraction of grazing days of domestic animal Beef Cattle ($m = 2$) spends in a stream each month (q) (Ratio)

[4.5.4.4.1.3 Horses](#)

The microbial loading rate to land-use type Pasture ($\ell = 2$), due to grazing associated with domestic animal Horses ($m = 5$) by month by subwatershed, is equal to:

$$\text{HorsesMicrobeRateShed}_{k,i,\ell=2,q} = (\text{NumberOfAnimals}_{i,m=\text{Horses}}) (\text{MicrobeAnimalProductionRates}_{k,m=\text{Horses}}) (\text{GrazingDays}_{m=\text{Horses},q}) / [(\text{DayInMonth}_q) (\text{Area}_{i,\ell=2})] \quad (12)$$

where

- $HorsesMicrobeRateShed_{k,i,\ell=2,q}$ = The microbial loading rate of microbe (k) to land-use type Pasture ($\ell = 2$) due to grazing of domestic animal Horses (m = 5) by month (q) by Subwatershed (i) (Cells/Time/Area)

4.5.4.4.1.4 Sheep

The microbial loading rate to land-use type Pasture ($\ell = 2$), due to grazing associated with domestic animal Sheep (m = 6) by month by subwatershed, is equal to:

$$SheepMicrobeRateShed_{k,i,\ell=2,q} = (NumberOfAnimals_{i,m=Sheep}) (MicrobeAnimalProductionRates_{k,m=Sheep}) (GrazingDays_{m=Sheep,q}) / [(DayInMonth_q) (Area_{i,\ell=2})] \quad (13)$$

where

- $SheepMicrobeRateShed_{k,i,\ell=2,q}$ = The microbial loading rate of microbe (k) to land-use type Pasture ($\ell = 2$) due to grazing of domestic animal Sheep (m = 6) by month (q) by Subwatershed (i) (Cells/Time/Area)

4.5.4.4.1.5 Other Agricultural Animals

The microbial loading rate to land-use type Pasture ($\ell = 2$), due to grazing associated with domestic animal Other Agricultural Animals (m = 7) by month by subwatershed, is equal to:

$$OtherAgAnimalMicrobeRateShed_{k,i,\ell=2,q} = (NumberOfAnimals_{i,m=OtherAgAnimal}) (MicrobeAnimalProductionRates_{k,m=OtherAgAnimal}) (GrazingDays_{m=OtherAgAnimal,q}) / [(DayInMonth_q) (Area_{i,\ell=2})] \quad (14)$$

where

- $OtherAgAnimalMicrobeRateShed_{k,i,\ell=2,q}$ = The microbial loading rate of microbe (k) to land-use type Pasture ($\ell = 2$) due to grazing of domestic animal Other Agricultural Animals (m = 7) by month (q) by Subwatershed (i) (Cells/Time/Area)

4.5.4.4.2 Manure Application to Land Surface

4.5.4.4.2.1 Dairy Cow

The microbial loading rate of microbes to land-use type Pasture ($\ell = 2$), which is the same loading rate to land-use type Cropland ($\ell = 1$), due to manure application associated with domestic animal Dairy Cow (m = 1) by month by subwatershed is equal to:

$$DairyCowMicrobeRateApply_{k,i,\ell=2,q} = (NumberOfAnimals_{i,m=DairyCow}) (MicrobeAnimalProductionRates_{k,m=DairyCow}) (FractionManureAvailableRunoff_{m=DairyCow,q}) (365 / DayInMonth_q) / (Area_{i,\ell=1} + Area_{i,\ell=2}) \quad (15)$$

4.5.4.4.2.2 Beef Cattle

The microbial loading rate due to manure application to land-use type Pasture ($\ell = 2$), which is the same loading rate to land-use type Cropland ($\ell = 1$), associated with domestic animal Beef Cattle ($m = 2$) by month by Subwatershed is equal to:

$$\text{BeefCattleMicrobeRateApply}_{k,i,\ell=2,q} = (\text{NumberOfAnimals}_{i,m=\text{BeefCattle}}) (\text{MicrobeAnimalProductionRates}_{k,m=\text{BeefCattle}}) (\text{FractionManureAvailableRunoff}_{m=\text{BeefCattle},q}) [(365 - \text{TotalGrazeDays}_{m=\text{BeefCattle}}) / \text{DayInMonth}_q] / (\text{Area}_{i,\ell=1} + \text{Area}_{i,\ell=2}) \quad (16)$$

4.5.4.4.2.3 Horses

The microbial loading rate due to manure application to land-use type Pasture ($\ell = 2$), which is the same loading rate to land-use type Cropland ($\ell = 1$), associated with Horses by month by Subwatershed is equal to:

$$\text{HorsesMicrobeRateApply}_{k,i,\ell=2,q} = (\text{NumberOfAnimals}_{i,m=\text{Horses}}) (\text{MicrobeAnimalProductionRates}_{k,m=\text{Horses}}) (\text{FractionManureAvailableRunoff}_{m=\text{Horses},q}) [(365 - \text{TotalGrazeDays}_{m=\text{Horses}}) / \text{DayInMonth}_q] / (\text{Area}_{i,\ell=2}) \quad (17)$$

where

- $\text{HorsesMicrobeRateApply}_{k,i,\ell=2,q}$ = The microbial loading rate of microbe (k) to land-use type Pasture ($\ell = 2$) associated with domestic animal Horses by month (q) by Subwatershed (i) (Cells/Time/Area)

4.5.4.5 Accumulated Microbial Loading Rates on Forest

The microbial loading of microbe (k) rate to Pasture ($\ell = 3$) due to shedding associated with all Wildlife, is equal to $\text{WildLifeMicrobeRateShedSum}_{k,\ell=3}$ (see Equation 5).

4.5.4.6 Accumulated Microbial Loading Rates on Urbanized Areas

The Urbanized Land-use type category is divided into four Urbanized categories ($r = 1$ for Commercial and Services; $r = 2$ for Mixed Urban or Built-Up; $r = 3$ for Residential; and $r = 4$ for Transportation, Communications and Utilities) which are further divided into Sub-urbanized categories ($u = 1$ for Commercial, $u = 2$ for SingleFamilyLowDensity, $u = 3$ for SingleFamilyHighDensity, $u = 4$ for MultiFamilyResidential, and $u = 5$ for Road). Accumulation rates in median microbial cells by microbe, per Urbanized land type area per time, indexed by the Urbanized subcategories are computed as follows:

$$\text{BuiltUpRate}_{k,\ell=4,r=1} = \text{SubUrbanizedBuiltUpRate}_{k,u=1} \quad (18)$$

$$\text{BuiltUpRate}_{k,\ell=4,r=2} = \{ \sum_{u=1,5} [\text{SubUrbanizedBuiltUpRate}_{k,u}] \} / 5 \quad (19)$$

$$\text{BuiltUpRate}_{k,\ell=4,r=3} = \{ \sum_{u=2,4} [\text{SubUrbanizedBuiltUpRate}_{k,u}] \} / 3 \quad (20)$$

$$\text{BuiltUpRate}_{k,\ell=4,r=4} = \text{SubUrbanizedBuiltUpRate}_{k,u=5} \quad (21)$$

where

- BuiltUpRate_{k,ℓ=4,r} = Accumulation rates in median microbial cells of microbe (k) per Urbanized land type (ℓ = 4) per area per time, indexed by the Urbanized categories (r) (Cells/Time/Area)
- SubUrbanizedBuiltUpRate_{k,u} = General microbial loading rates of microbe (k) by sub-urbanized (SubUrbanized) category (u) (Cells/Time/Area)

Accumulated microbial loading rate by microbe, associated with the Urbanized land-use type per subwatershed, weighted by the areas associated with four Urbanized categories (i.e., Commercial and Service; Residential; Mixed Urban; Transportation, Communication, and Utilities), for all months (i.e., applicable throughout the year) is computed as follows:

$$\text{AccumBuiltUpRate}_{k,i,\ell=1,2,3} = 0 \quad (22)$$

$$\text{AccumBuiltUpRate}_{k,i,\ell=4} = \sum_r [(\text{AreaFraction}_{i,\ell=4,r}) (\text{BuiltUpRate}_{k,\ell=4,r})] \quad (23)$$

where

- AccumBuiltUpRate_{k,i,ℓ=4} = Accumulated microbial loading rate of microbe (k) associated with Urbanized land-use type (ℓ = 4) per subwatershed (i), weighted by areas associated with four Urbanized categories (r) (i.e., Commercial and Service; Residential; Mixed Urban; Transportation, Communication, and Utilities) for all months (i.e., throughout the year) (Cells/Time/Area)
- AreaFraction_{i,ℓ,r} = Fraction of the Urbanized Area for Urbanized land type (ℓ = 4) attributed to each sub-urbanized (r) area (i.e., Commercial and Service; Residential; Mixed Urban; Transportation, Communication, and Utilities) by subwatershed (i) (i.e., ratio of the urbanized subcategory area and total urbanized area). Fractions must total 1.0.

4.5.4.7 Accumulated Overland Microbial Loading Rates to the Land Surface, Adjusted for Die-off

4.5.4.7.1 Die-off Adjustment

Microbial accumulation on the land surface and maximum microbial storage accumulation calculations are based on formulations associated with the HSPF watershed model. Die-off or decay on overland surfaces is simulated as a function of the input accumulation rate and maximum storage of microbes which represents accumulation without removal (e.g., die-off, runoff, etc.). The unit removal rate represents processes such as die-off and wind erosion (Bicknell et al., 2005). The unit removal rate of the stored microbes (number removed per day) is computed as the microbial accumulation rate (Cells/ac/d), divided by the maximum microbial storage accumulation (storage limit) (Cells/ac). For example, the removal rate = (accumulation rate) / (storage limit). The general factor used to compute maximum microbial storage accumulation on the land surface ($N_{k,q}$) for a given microbe (k) is computed as the sum of storages for each day of the month (q):

$$N_{k,q} = No_{k,q} \int_0^{\text{DayInMonth}_q} 10^{-\text{DieOff}_{k,q} \cdot t} dt = [No_{k,q} / (2.303 \text{ DieOff}_{k,q})] [1 - 10^{-(\text{DayInMonth}_q \cdot \text{DieOff}_{k,q})}] \quad \text{from } t = 0 \text{ to } t = \text{DayInMonth}_q \quad (24)$$

where

- $N_{k,q}$ = maximum microbial storage accumulation on the land surface for microbe (k) for month (q) (Cells/Area)
- $No_{k,q}$ = Initial uniform loading to the overland surface for microbe (k) for month (q) (Cells/Time/Area)
- $DieOff_{k,q}$ = First-order microbial inactivation/die-off rate for microbe (k) on the land surface by month (q), which accounts for warm and cold months (1/d)
- 2.303 = Conversion constant for $\ln(10)$

When $DieOff_{k,q}$ is less than $10^{-6} d^{-1}$, $[1/(2.303 k)] \times [1 - 10^{-(DayInMonth_q \cdot DieOff_{k,q})}]$ can be approximated as $DayInMonth_q$.

When $DieOff_{k,q}$ is greater than $10^{-1} d^{-1}$, $[1/(2.303 DieOff_{k,q})] \times [1 - 10^{-(DayInMonth_q \cdot DieOff_{k,q})}]$ can be approximated as $[1/(2.303 DieOff_{k,q})]$. For example, for a die-off rate of $DieOff_{k,q} = 0.36/d$,

$$N_{k,q} = No_{k,q} / 0.83 = 1.21 No_{k,q} \quad (25)$$

Likewise, for a die-off rate of $k = 0.51/d$,

$$N_{k,q} = No_{k,q} / 1.17 = 0.85 No_{k,q} \quad (26)$$

4.5.4.7.2 Accumulated Overland Microbial Loading Rates and Maximum Microbial Storage with Die-off

The accumulated overland microbial loading rates per area and maximum microbial storage per area with die-off are presented by microbe (k) by subwatershed (i) by land-use type (ℓ) by month (q).

4.5.4.7.2.1 Cropland

The summation of microbial loading rate per area by microbe (k) by subwatershed (i) by month (q) across all domestic animals (m) and wildlife (n) on land-use type Cropland ($\ell = 1$) is computed as follows:

$$\begin{aligned} AccumulationRateMonth_{k,i,\ell=1,q} = & (WildLifeMicrobeRateShedSum_{k,\ell=1}) + \\ & (DairyCowMicrobeRateApply_{k,i,\ell=1,q}) + (BeefCattleMicrobeRateApply_{k,i,\ell=1,q}) + \\ & (PoultryMicrobeRateApply_{k,i,\ell=1,q}) + (SwineMicrobeRateApply_{k,i,\ell=1,q}) \end{aligned} \quad (27)$$

where

- $AccumulationRateMonth_{k,i,\ell=1,q}$ = Rate of microbial accumulation per area without die-off of microbe (k) on the land surface by month (q) by land-use type (ℓ) by subwatershed (i) across all domestic animals (m) and wildlife (n) (Cells/Time/Area)

4.5.4.7.2.2 Pasture

The summation of microbial loading rate per area by microbe (k) by subwatershed (i) by month (q) across all domestic animals (m) and wildlife (n) on land-use type Pasture ($\ell = 2$) is computed as follows:

$$\begin{aligned} AccumulationRateMonth_{k,i,\ell=2,q} = & (WildLifeMicrobeRateShedSum_{k,\ell=2}) + \\ & (BeefCattleMicrobeRateShed_{k,i,\ell=2,q}) + (HorsesMicrobeRateShed_{k,i,\ell=2,q}) + (SheepMicrobeRateShed_{k,i,\ell=2,q}) + \end{aligned}$$

$$\begin{aligned}
& (\text{OtherAgAnimalMicrobeRateShed}_{k,i,\ell=2,q}) + (\text{DairyCowMicrobeRateApply}_{k,i,\ell=2,q}) + \\
& (\text{BeefCattleMicrobeRateApply}_{k,i,\ell=2,q}) + (\text{HorsesMicrobeRateApply}_{k,i,\ell=2,q})
\end{aligned} \tag{28}$$

4.5.4.7.2.3 Forest

The summation of microbial loading rate per area by microbe (k) by subwatershed (i) by month (q) across all domestic animals (m) and wildlife (n) on land-use type Forest ($\ell = 3$) is computed as follows:

$$\text{AccumulationRateMonth}_{k,i,\ell=3,q} = \text{WildLifeMicrobeRateShedSum}_{k,\ell=3} \tag{29}$$

4.5.4.7.2.4 Urbanized

The summation of microbial loading rate per area by microbe (k) by subwatershed (i) by month (q) across all domestic animals (m) and wildlife (n) on land-use type Urbanized ($\ell = 4$) is computed as follows:

$$\text{AccumulationRateMonth}_{k,i,\ell=4,q} = \text{AccumBuiltUpRate}_{i,k,\ell=4} \tag{30}$$

4.5.4.7.3 Maximum Microbial Storage with Die-off

The maximum microbial storage per area by microbe (k) by subwatershed (i) by month (q) by land-use type (ℓ) across all domestic animals (m) and wildlife (n), adjusted for die-off, is computed as follows:

$$\begin{aligned}
& \text{StorageLimitMonth}_{k,i,\ell,q} = \\
& = \left[\text{AccumulationRateMonth}_{k,i,\ell,q} / (2.303 \text{ DieOff}_{k,q}) \right] \quad \text{for } 10^{-6} < \text{Die-off} < 10^{-1} \\
& \quad \left[1 - 10^{-(\text{DayInMonth}_q \cdot \text{DieOff}_{k,q})} \right] \\
& \approx \left[\text{AccumulationRateMonth}_{k,i,\ell,q} / (2.303 \text{ DieOff}_{k,q}) \right] \quad \text{for } \text{Die-off} > 10^{-1} \\
& \approx (\text{AccumulationRateMonth}_{k,i,\ell,q}) (\text{DayInMonth}_q) \quad \text{for } \text{Die-off} < 10^{-6}
\end{aligned} \tag{31}$$

where

- $\text{StorageLimitMonth}_{k,i,\ell,q}$ = Maximum microbial storage per area of microbe (k) by subwatershed (i) by month (q) by land-use type (k), across all domestic animals and wildlife, adjusted for die-off (removal) (Cells/Area)

4.5.4.8 Microbial Point Source Loading Rates

4.5.4.8.1 Cattle in Streams

The microbial loading rate of Beef Cattle ($m = \text{BeefCattle}$) shedding into a stream by microbe (k), by subwatershed (i), by month (q) is as follows:

$$\text{BeefCattleStreamMicrobeRate}_{k,i,q} = (\text{NumberOfAnimals}_{i,m=\text{BeefCattle}}) (\text{MicrobeAnimalProductionRates}_{k,m=\text{BeefCattle}}) [(\text{GrazingDays}_{m=\text{BeefCattle},q}) / \text{DayInMonth}_q] (\text{TimeSpentInStreams}_{m=\text{BeefCattle},q}) \quad (32)$$

where

- $\text{BeefCattleStreamMicrobeRate}_{k,i,q}$ = Microbial loading rate of microbe (k) for domestic animal Beef Cattle (m = 2) shedding into a stream by subwatershed (i) by month (q) (Cells/Time)

4.5.4.8.2 Septics

The average septic flow rate to the stream by subwatershed is as follows:

$$\text{SepticStreamFlowRate}_i = (\text{SepticNumber}_i) (\text{SepticNumberPeople}) (\text{SepticOvercharge}) (\text{SepticFailureRate}) \quad (33)$$

where

- $\text{SepticStreamFlowRate}_i$ = Average septic flow rate to the stream subwatershed (i) (Volume/Time)
- SepticNumber_i = Number of septic systems associated with Subwatershed (i) (Number of septics)
- $\text{SepticNumberPeople}$ = Average number of people per septic system (Number of people/septic)
- SepticOvercharge = Typical septic overcharge flow rate (Volume/Time/Person)
- SepticFailureRate = Typical fraction of septic systems that failure (Ratio)

The microbial loading rate associated with septic systems by microbe, by subwatershed, is as follows:

$$\text{SepticStreamLoadingRate}_{k,i} = (\text{SepticStreamFlowRate}_i) (\text{SepticConc}_k) \quad (34)$$

where

- $\text{SepticStreamLoadingRate}_{k,i}$ = Microbial loading rate of microbe (k) to the stream from leaking septic systems by subwatershed (i) (Cells/Time)
- SepticConc_k = Typical microbial density (concentration) of microbe (k) in septic system waste (Cells/Volume)

5. CONTEXT OF THE MICROBIAL SOURCE MODULE WITHIN A MULTI-COMPONENT WORKFLOW

Although the focus of this document is to describe how to capture the ontology associated with a component (i.e., MSM) for discovery, access, and execution on the web, it also provides context for where the component fits into a larger modeling workflow. Incorporating MSM into a workflow allows other components to provide MSM with key input. MSM needs microbial properties (e.g., die-off rates) and number of subwatersheds associated with the watershed delineation. In addition, MSM needs to know where the sources are located, relative to the subwatershed delineation and strength of each source (e.g., microbial loading rate). These data are supplied by other modules and databases associated with the workflow, and their original form may not match input requirements of the MSM; therefore, some transformation may be necessary. This section discusses the linkage and operation of MSM within a workflow.

5.1 MODELING WORKFLOW

A software infrastructure is developed to automate the manual process of characterizing transport of pathogens and indicator microorganisms, from sources of release to points of exposure, by loosely configuring a set of modules and process-based models. The MSM interacts directly with other components and receives information from a suite of files and other databases; a design schematic of the workflow, which tracks data from sources to downstream locations within a watershed and visualizes simulation results, is presented in [Figure 5](#). Major components are as follows:

- **Microbial Input CSV Files:** The microbe properties source data which need to be manually researched and populated capture microbial related properties of microorganisms, as well as data related to the release of microorganisms associated with fecal material. Microbial properties data are housed within 12 local data files, and these files allow users to modify input parameters that directly impact microbial loadings to a watershed. They contain microbial source locations, animal numbers and types, correlated with land-use type within subwatersheds. Although there is no user-friendly interface to manage them, these files allow users to identify specific locations (e.g., farms) of domestic animals, sampling/monitoring, septic systems, and point sources [e.g., Publicly Owned Treatment Works (POTWs) and Wastewater Treatment Plants (WWTPs)]. The files are included in the software download and stored in a “LocalData” subfolder which is concurrently created during execution of the QMRA software within the project’s working folder.
- **Data for Environmental Modeling (D4EM):** D4EM manages, accesses, retrieves, analyzes, and caches web-based environmental data. D4EM ([Wolfe et al., 2007](#); [Aqua Terra, 2011](#); [EPA, 2013a](#)) (1) is an open source automated data access and processing library that accesses a variety of data types including water quality, land use, hydrology, soils, meteorological (MET), stream flow, groundwater levels, and crop data; (2) uses DotSpatial geo-processing libraries ([CodePlex, 2014](#)) to perform cartographic re-projections, intersection, clipping, overlaying, joining and merging of geographic features, and areas-of-interest delineation (e.g., subwatersheds); (3) performs statistical processing (extraction, interpolation and averaging) of time series data; (4) incorporates automatic data access functionality; and (5) consists of a collection of .Net dynamic link libraries that can be linked to a modeling utility such as a batch processor or script to access data for multiple watersheds, or used with a custom-built user interface such as the SDMPProjectBuilder.

- **Site Data Manager Project Builder (SDMProjectBuilder or SDMPB):** The SDMProjectBuilder leverages D4EM to manage access, retrieval, analysis, and caching of web-based environmental data (e.g., NHDPlus, NLCD, NCDC, STORET, NLDAS, STATSGO/SSURGO, etc.); provides geographical information system (GIS) capabilities using DotSpatial technology; converts DotSpatial-based project files to MapWindow-based project files ([MapWindow 2011, 2013](#); [Watry and Ames, 2010](#)); and automatically pre-populates input files of fate and transport models. It automates the watershed delineation process, allowing for 8-digit and 12-digit Hydrologic Unit Codes (HUC-8 and HUC-12, respectively) or pour point analyses; assigns map-layer features automatically [e.g., slope, soil, land use, microbial sources, National Climatic Data Center (NCDC) and North American Land Data Assimilation System (NLDAS) ([Kim et al., 2014](#)) MET stations]; and accounts automatically for snow accumulation/melt, microbial fate and transport, and user-defined simulation time increments (e.g., hourly, daily, monthly, or annually). The ability to define time increments is valuable, especially when hourly (i.e., event-based) storm events drive an assessment.
- **Hydrologic Simulation Program – FORTRAN (HSPF):** HSPF ([Bicknell et al., 1997](#)) is a comprehensive package for simulating watershed hydrology and water quality for conventional and nonconventional pollutants (such as sediment/nutrients and toxic organics, respectively) and microbes, with microbes as the major focus in this application; implements a basin-scale analysis that includes hydrology and water quality to allow integrated simulation of land and soil contaminant runoff processes with instream hydraulic and microbial interactions on user-defined time scales (hour, day, month, or year); and provides a time history of runoff flow rates and microbial concentrations at any point in a watershed ([EPA, 2013d](#)). HSPF simulates flow, microbial, and chemical fate/transport within a watershed and executes as a stand-alone or from within BASINS.
- **Parameter ESTimation and Uncertainty Analysis (PEST):** The model-independent PEST is a nonlinear parameter estimation package that can estimate parameter values for almost any existing computer model, whether a user has access to the model's source code or not ([SSPA, 2010](#); [C⁴Site Factory, 2011](#); [Doherty, 2005](#)). PEST is designed to interface with an existing model, modify designated input, run the model as often as needed, and adjust its parameters until differences between simulated and monitored output results are minimized in a weighted least squares sense. PEST communicates with a model through the model's own input and output files. PEST implements a variant of the Gauss-Marquardt-Levenberg method of nonlinear parameter estimation. It also allows fine-tuning of parameter estimation through adjustment of control variables.
- **Better Assessment Science Integrating point & Non-point Sources system (BASINS):** BASINS ([EPA, 2001a](#); [Lahlou et al., 1998](#)) is a multipurpose environmental analysis infrastructure that performs watershed- and water quality-based analyses by integrating environmental data, analysis tools, and watershed and water quality models. A MapWindow-based GIS organizes spatial information that displays maps, tables, or graphics; analyzes landscape information; integrates and displays relationships among data at a user-chosen scale; and provides graphical and tabular viewers of flow and concentration output ([EPA, 2013e](#)).
- **San Antonio River Authority (SARA) Timeseries Utility:** The SARA Timeseries Utility supports analysis and management of time-varying environmental data including listing, graphing, computing statistics, computing MET data, and saving in a WDM or text file ([Aqua Terra, 2011](#)). Supported file formats include: WDM, HSPF Binary (.hbn), USGS RDB, and text files, using the same import scripts as WDMUtil and BASINS. Users of WDMUtil will find most of its functionality is included in this new utility tool which allows users to publish HSPF data as a txt file for consumption by other components.

- **Microbial Risk Assessment—Interface Tool (MRA-IT):** MRA-IT is a MathCad-based, integrated software tool based on the pathogen of interest, exposure, intake, and dose (Whelan et al., 2014b; Soller et al., 2008, 2004; Soller and Eisenberg, 2008). MRA-IT consumes a series of pathogenic densities from a text (.txt) file like that produced by the SARA Timeseries Utility and characterizes human-health risk from ingestion of reclaimed water through recreational activities, consumption of crops irrigated with reclaimed water, or incidental/inadvertent ingestion of reclaimed water associated with golf course/landscape irrigation. Key components include pathogen specification, exposure scenario identification, and dose-response relationships.

The process has been automated from source to receptor within a loosely coupled workflow which allows many of the components to operate individually or within the workflow. Those components that are fully linked to each other and automated include SDMPProjectBuilder, D4EM, MSM, accessing and reading the microbial input CSV files, and constructing and populating the HSPF input files.

5.2 MSM INPUT REQUIREMENTS

Table 3 lists the MSM input parameters and variables; some are supplied by the upstream module Site Data Manager Project Builder (SDMPProjectBuilder or SDMPB), and some are user-defined and supplied by a suite of Comma Separated Values (CSV) files.

5.2.1 Input Data Variables

Data on watershed characteristics are supplied by SDMPB/D4EM and designed in Table 3 by “SDMPBOutput” within in the first column. These data are automatically supplied to the MSM input file.

5.2.2 Microbial Input Comma Separated Values Files

Twelve user-defined CSV files and their microbial source-term input data requirements using MSM are listed in Table 12. Column 1 identifies the CSV file name and corresponding model or module (i.e., SDMPB or MSM, respectively) that consumes data. Columns 2 and 3 define each parameter and its corresponding units, respectively. SDMPB uses some of these data in calculations to produce output (Column 4), which is the input to MSM. For example, input location points defined by latitude and longitude for SDMPB (Column 2) are spatially overlaid onto the watershed to identify corresponding subwatersheds, which are required input to MSM. CSV files and example input are presented in Tables 13 through 22. Table 12 should be used to help correlate parameter names and their units with the information contained within Tables 13 through 22. Locations of farms that contain domestic animals; point sources that discharge directly into the stream; and septic system, output, and boundary points are documented in Tables 13, 14, and 15, respectively. Output points are intermediate locations within the watershed where simulation results are produced. Boundary points are locations where upstream areas have been evaluated *a priori* and represent flow and concentration boundary conditions for downstream evaluation. Information associated with domestic animals is captured in Tables 13 and 16 through 18. Instream loadings from septic systems and point sources are shown in Tables 15 and 19 and Tables 14 and 20, respectively. Wildlife shedding rates and microbial densities by land-use type are captured in Tables 18 and 21, respectively; Table 16 also captures loading rates associated with four different urbanized (built-up) areas, with each sub-divided as appropriate, into Urbanized sub-categories. Monthly first-order microbial die-off rates in surface soils appear in Table 22.

Table 12. Files providing data consumed by SDMPB or MSM (after [Wolfe et al., 2016](#))

CSV File Name and Model Consuming Data	Data and Definition, as contained in the CSV File	Units in CSV File	Parameter Consumed as Input by MSM (unless noted)
Domestic Animals and Wildlife			Area and AreaFraction*
AnimalLL.csv SDMPB	Domestic animal location by latitude and longitude	Degree (by fraction)	Subwatershed*
	Domestic animal numbers by type (m) by latitude and longitude location	Number	NumberOfAnimals
FCProdRates.csv MSM	Production or shedding rate of microbes (k) from domestic animal (m)	Cells/d/animal	MicrobeAnimalProductionRates
	Microbial (k) production or shedding rate per wildlife (n) per area	Cells/d/wildlife	MicrobeWildlifeProductionRates
	Microbial (k) loading rate by sub-urbanized category (u)	Cells/d/ac	SubUrbanizedBuiltupRate
GrazingDays.csv MSM	Number of grazing days per domestic animal ($m = 2, 5, 6$, and 7) per month (q)	Number	GrazingDays
	Fraction of the number of grazing days that Beef Cattle ($m = 2$) spend in stream per month (q)	Fraction	TimeSpentInStreams
ManureApplication.csv MSM	Fraction of manure applied to soil each month (q) per domestic animal ($m = 1 \rightarrow 5$)	Fraction	Application
	Fraction of amount of manure shed by the domestic animal ($m = 1 \rightarrow 5$) incorporated into soil	Fraction	ManureIncorporatedIntoSoil
MonthlyFirstOrderDieOffRateConstants.csv MSM	First-order microbial (k) inactivation/die-off rate on the land surface per month (q)	1/d	Die-off
WildlifeDensities.csv MSM	Number of wildlife (n) per unit area by land-use type (ℓ)	Number/mi ²	Density
Point Sources			
PointSourceLL.csv SDMPB	Point source locations by point source ID (PtSrcId) and latitude and longitude	Degree (by fraction)	Subwatershed* (not used by MSM)

PointSourceData.csv SDMPB	Annual-average discharge (Load) for each point source ID (PtSrcId) and facility name (FacName).	ft ³ /s	PointFlow (not used by MSM)
	Annual-average microbe (<i>k</i>) loading rate (Load) for each point source ID (PtSrcId) and facility name (FacName).	Cells/yr	PointMicrobeRate (not used by MSM)
	Annual-average chemical loading rate (Load) for each point source ID (PtSrcId) and facility name (FacName).	Lbs/yr	PointChemRate (not used by MSM)
Septic Systems			
SepticsLL.csv SDMPB	Septic system locations by latitude and longitude	Degree (by fraction)	Subwatershed* SepticNumber
SepticsDataWatershed.csv MSM	Number of people per septic unit	Number/septic	SepticNumberPeople
	Average fraction of septic systems that fail	Fraction	SepticFailureRate
	Average septic overcharge rate per person	gal/d/person	SepticOvercharge
	Microbial (<i>k</i>) density of septic overcharge reaching the stream	Cells/L	SepticConc
Intermediate Points			
BoundaryPointsLL.csv SDMPB	Boundary point locations by latitude and longitude	Degree (by fraction)	Subwatershed* (not used by MSM)
OutputPointsLL.csv SDMPB	Output point locations by latitude and longitude	Degree (by fraction)	Subwatershed* (not used by MSM)

*Produced by SDMPB, based on NHDPlus data and user-supplied delineation guidelines (i.e., minimum stream length and minimum subwatershed size). The SDMPB overlays and maps latitude-longitude locations to subwatersheds and supplies the corresponding subwatershed location to MSM, when appropriate.

Table 13. AnimalLL.csv template and example data*

Latitude	Longitude	BeefCow	Swine	DairyCow	Poultry	Horse	Sheep	OtherAg
44.23752	-88.0046	30	0	30	0	0	0	0
44.19700	-88.0954	0	0	0	0	0	0	0
44.16789	-88.0410	75	0	75	0	0	0	0
• •								
• •								
44.05979	-87.8475	0	0	0	93	0	0	0
44.09854	-88.0243	0	0	0	66	0	0	0
44.08787	-87.9167	0	0	0	25	0	0	0

*BeefCow = BeefCattle

Table 14. PointSourceLL.csv template and example data for two point sources

Latitude	Longitude	PtSrcId
44.112	-88.256	PT001
44.06	-88.191	PT002

Table 15. Template and example data for SepticsLL.csv, OutputPoints.csv, or BoundaryPoints.csv

Latitude	Longitude
44.23752	-88.0046
44.19700	-88.0954
•	•
•	•
44.09854	-88.0243
44.08787	-87.9167

Table 16. FCProdRates.csv template and example shedding rates by domestic animal and wildlife, and production rates by urban (built-up) areas*

Source	Value	Units
DairyCow	2.50E+10	CountPerAnimalPerDay
BeefCow	3.30E+10	CountPerAnimalPerDay
Swine	1.10E+10	CountPerAnimalPerDay
Sheep	1.20E+10	CountPerAnimalPerDay
Horse	4.20E+08	CountPerAnimalPerDay
Poultry	1.31E+08	CountPerAnimalPerDay
Duck	2.40E+09	CountPerAnimalPerDay
Goose	8.00E+08	CountPerAnimalPerDay
Deer	3.50E+08	CountPerAnimalPerDay
Beaver	2.50E+08	CountPerAnimalPerDay
Raccoon	1.25E+08	CountPerAnimalPerDay
OtherAgAnimal	0.00E+00	CountPerAnimalPerDay
OtherWildlife	0.00E+00	CountPerAnimalPerDay
Road	2.00E+05	CountPerAcrePerDay
Commercial	6.21E+06	CountPerAcrePerDay
SingleFamilyLowDensity	1.03E+07	CountPerAcrePerDay
SingleFamilyHighDensity	1.66E+07	CountPerAcrePerDay
MultifamilyResidential	2.33E+07	CountPerAcrePerDay

*BeefCow = BeefCattle

Table 17. ManureApplication.csv template and example data

ManureType	JanFractionApplied	FebFractionApplied	MarFractionApplied	AprFractionApplied	MayFractionApplied	JunFractionApplied	JulFractionApplied	AugFractionApplied	SepFractionApplied	OctFractionApplied	NovFractionApplied	DecFractionApplied	FractionIncorporatedIntoSoil
SwineManure	0	0	0	0.15	0.3	0.2	0.15	0.1	0.05	0.05	0	0	0.8
CowManure	0	0	0	0.15	0.3	0.2	0.15	0.1	0.05	0.05	0	0	0.75
CattleManure	0	0	0	0.15	0.3	0.2	0.15	0.1	0.05	0.05	0	0	0.75
HorseManue	0	0	0	0.1	0.1	0	0	0	0	0.4	0.4	0	0.75
PoultryLitter	0	0	0	0.15	0.3	0.2	0.15	0.1	0.05	0.05	0	0	0.96

Table 18. GrazingDays.csv template and example data

Month	BeefCattleGrazingDays	HorseGrazingDays	SheepGrazingDays	OtherAgAnimalGrazingDays	FractionOfTimeBeefCattleInStreams
January	0	3.1	31	0	0
February	0	2.8	28	0	0
March	0	3	31	0	0
April	30	27	30	0	0
May	31	27.9	31	0	0
June	30	27	30	0	0.1
July	31	27.9	31	0	0.1
August	31	27.9	31	0	0.1
September	30	27	30	0	0
October	31	27.9	31	0	0
November	0	27	30	0	0
December	0	3.1	31	0	0

Table 19. SepticsDataWatershed.csv template and example data, correlated with SepticsLL.csv

NumberOfPeoplePerSepticUnit	SepticFailureRate_Fraction	SepticOverchargeFlowRate_gallonsPerDayPerPerson	FCCConcentrationReachingStreamFromSepticOvercharge_CountsPerLiter
3.75	0.12	70	1.0E+7

Table 20. PointSourceData.csv template and example data for two point sources with flow, microbes, and chemicals, correlated with PointSourceLL.csv

PtSrcId	FacName	Load	Parm
PT001	PointSource1	1	FLOW
PT001	PointSource1	1000	Microbes
PT001	PointSource1	5	Diazinon
PT002	PointSource2	2	FLOW
PT002	PointSource2	2000	Microbes
PT002	PointSource2	4	Diazinon

Table 21. WildlifeDensities.csv template and example data

Animal	DensityPerSqMile_Cropland	DensityPerSqMile_Pasture	DensityPerSqMile_Forest
Duck	14.13	14.13	14.13
Goose	2.22	2.22	2.22
Deer	20.51	20.51	20.51
Beaver	0	0	0
Raccoon	0	0	0
OtherWildlife	0	0	0

Table 22. Template and example data for MonthlyFirstOrderDieOffRateConstants.csv

Month	DieOffRateContant
January	0.027
February	0.035
March	0.042
April	0.050
May	0.058
June	0.065
July	0.073
August	0.065
September	0.058
October	0.050
November	0.042
December	0.035

5.2.2.1 Domestic Animals

The number and type of domestic animals associated with each farm location are documented in [Table 13](#). County-wide agricultural census data can be retrieved from USDA (e.g., [USDA, 2016](#)), although many states track the numbers and types of domestic animals by farm – especially if the numbers exceed a threshold, as with designated concentrated animal feeding operations (CAFOs).

A single shedding rate is captured in [Table 16](#) and associated with each domestic animal, as defined in [Table 8](#). If different age groups or types of domestic animals (e.g., calf, heifer, cow, bull, steer, etc.) are of concern for grazing/shedding, these may also be captured in MSM by using another category such as “Other” in [Table 8](#), “OtherAg” in [Table 13](#), and “OtherAgAnimal” in [Table 16](#). MSM assumes that wastes generated from and associated with a location are released within the assigned subwatershed.

[Table 17](#) captures the monthly land-application schedule of manure for each domestic animal, as the fraction of manure applied to soil; it also captures the fraction of manure shed by each domestic animal that is eventually incorporated into soil. Monthly schedules capture seasonal trends. [Table 18](#) captures

the number of days per month that a domestic animal and fraction of the number of grazing days per month that beef cattle spend in stream. The MSM assumes that the category “Dairy Cow” is confined and that “Beef Cattle” are allowed to graze and enter the stream; the terms only differentiate between non-grazing and grazing animals, respectively. For example, if dairy cows graze and/or shed directly to the stream, users can designate them as beef cattle. If beef cattle are restricted from entering the stream, then time spent in stream can be set to zero. Schedules in [Table 18](#) apply across the watershed.

[5.2.2.2 Septic Systems and Point Sources](#)

Instream loadings from septic systems and point sources are captured by [Tables 19 and 20](#), respectively; [Table 19](#) captures average number of people per septic unit, fraction of systems that fail, and overcharge flow rate as well as microbial density associated with the overcharge. Information supporting septic releases in [Table 19](#) represent the watershed as a whole – that is, the same average usage rate, failure and overcharge rates, and microbial densities are applied to each septic location. All septic systems within each subwatershed are combined to represent a single loading to the respective subwatershed stream segment.

The SDMPB combines the time series of microbial loadings to the stream associated with septic and instream loading by beef cattle by microbe (k), by subwatershed (i), by month to compute a combined monthly time series, noting that there is only one point source per subwatershed:

$$\text{PointMicrobeRateToStream}_{k,i,q} = \text{BeefCattleStreamMicrobeRate}_{k,i,q} + \text{SepticStreamLoadingRate}_{k,i} \quad (35)$$

where

- $\text{PointMicrobeRateToStream}_{k,i,q}$ = Microbial loading rate time series to the stream from Point Sources in the stream by microbe (k), by subwatershed (i), by month (q) (Cells/Time)

The flow rate that is produced by SDMPB is equal to the total septic flow rate to the stream subwatershed (i) applied each month (q):

$$\text{PointFlowToStream}_{i,q} = \text{SepticStreamFlowRate}_i \quad (36)$$

where

- $\text{PointFlowToStream}_{i,q}$ = Point source discharge to the stream by subwatershed (i) by month (q) (Volume/Time)

Equation (36) could be construed as redundant, but it explicitly assigns a point source discharge to the stream in this assessment.

Point source discharges also include direct input to streams from engineered sources such as WWTPs and POTWs, but these types of point source discharges are directly handled by the HSPF watershed model input data stream and not by the SDMPB or MSM. To establish a place holder within the HSPF UCI input file, the SDMPB consumes a single annual average discharge and microbial and chemical loading rates at specified point locations, as shown in [Table 20](#). If the user wants to assess only microbes, the file is modified by removing rows for chemicals. If there is only one point source, rows related to other

point sources are removed. Additional point sources can be identified and added. SDMPB consumes these data as a function of point source ID, name, and latitude and longitude (Table 14); overlays latitude and longitude locations onto the watershed, mapping to corresponding subwatershed locations; and produces the following data for watershed model consumption: annual-average discharge by subwatershed (*i*) (ft³/s), annual-average microbe loading rate by subwatershed (*i*) (Cells/yr), and annual-average chemical loading rate by subwatershed (*i*) (Lbs/yr). The annual average point-source data are essentially placeholders for when the user replaces these constant values with actual point source time series (e.g., daily values); this is accomplished by editing the HSPF WDM file through the BASINS system and activating the point-source location in the HSPF interface. Because MSM consumes only microbial data, chemical data are of no importance for a microbial assessment in this case and, thus, not discussed. Automatically including a place holder in the watershed input file allows the file to be more easily updated with the actual time series through watershed model user interfaces (HSPF and BASINS).

[5.2.1.3 Wildlife](#)

MSM considers six wildlife categories (Table 9). Wildlife shedding rates (Table 16) and microbial densities (Table 21) vary by land-use type. Wildlife is assumed to shed on forest, cropland, and pasture, but not in urban areas or streams, although certain wildlife (e.g., geese) may shed in large quantities in urban areas, golf courses, and streams. Numbers for selected wildlife (e.g., deer) are typically available for each state.

[5.2.2.4 Microbial Die-off](#)

Monthly first-order microbial die-off rates due to manure on surface soils, captured and used in Table 22, allow users to account for variations in die-off by season. Die-off rates apply to both domestic animals and wildlife loadings to land surfaces.

[5.2.2.5 Urban Sources](#)

Potential loading rates in urban areas according to the following categories and subcategories are accounted for in Table 16:

- Commercial and Services: Commercial
- Mixed Urban or Built-up: Road, Commercial, Single-family low density, Single-family high density, and Multi-family residential
- Residential: Single-family low density, Single-family high density, and Multi-family residential
- Transportation, Communications, and Utilities

Urbanized built-up areas include roads (Road), commercial property (Commercial), single-family-low-density residence (SingleFamilyLowDensity), single-family high density residence (SingleFamilyHighDensity), and multi-family residential (Multi-familyResidential). Single-family low density is a single-detached dwelling, single-family residence, or separate house which is a free-standing residential building (Wikipedia, 2015a). Single-family high density is a suite of smaller-scale single-family dwellings, representing a more compact single-family residential development (e.g., 13 – 40 units/ac) (Garnett, 2012). Multi-family residential is a unit with multiple separate housing units for residential inhabitants contained within one building, or several buildings within one complex such as an apartment or condominium (Wikipedia, 2015b). A single, weighted urbanized loading rate is quantified for each

subwatershed (all months) based on individual urbanized land uses. Although microbial loadings maybe determined by land-use type (pasture, cropland, urbanized, and residential), accumulated shedding and land application are computed by month, across all domestic animals and wildlife, and assigned to the entire subwatershed.

5.2.2.6 Supporting Literature Information

Because normal microbial composition of animal feces is different from human feces and can change dramatically over time and space (Boehm et al., 2002; Dorner et al., 2007; NRC, 2004), animal and human sources of pathogens and indicators can be treated differently, depending on characterization of fecal material and availability of technology that can accurately and reliably differentiate between sources (EPA, 2009). There still remains wide variability within and between sites (Fraser et al., 1998) and in relevant literature. For example, Kim et al. (2016) performed detailed monitoring of microbial release from manure and subsequent overland runoff on 36 identically prepared, side-by-side plots in the same field which resulted in 144 plot-scale, rainfall-runoff events. The range in microbial densities was more than eight orders of magnitude. Wolfe et al. (2016) demonstrated that microbial loading rates to a mixed-use watershed – based on numbers and types of domestic animals and wildlife; microbial densities; shedding and production rates by domestic animal, wildlife, and septic; and microbial die-off rates – represent only estimates and thus require calibration using observed densities downstream. A summary of densities and production rates is not tabularized herein, although suggested values are in the published literature.

Soller et al. (2015) and EPA (2010) provided studies related to occurrence and abundance (shedding densities in cells/g manure) of pathogens (e.g., *E. coli* O157:H7, *Campylobacter*, *Salmonella* spp., *Cryptosporidium* spp., *Giardia* spp.) in manures from domestic animals (beef cattle, dairy cows) and disinfected secondary effluent. EPA (2009) provided representative fecal indicator bacteria and zoonotic pathogen densities in human and animal feces and sewage. Soller et al. (2010) documented ranges used to characterize densities of indicators (*E. coli* and enterococci) and reference pathogens (*E. coli* O157:H7, *Cryptosporidium* spp., *Salmonella* spp., *Giardia* spp., Norovirus) in the fecal sources (primary sewage, secondary chlorinated effluent, gulls, cattle, pigs, chickens) (Schoen and Ashbolt, 2010). EPA (2010) and Butler et al. (2008) provided example shedding rates for cows (Whelan et al., 2014b). Geldreich (1978) and ASAE (2005) provided manure production rates and fecal coliform shedding rates associated with various domestic animals and wildlife (Zeckoski et al., 2005). Overcash et al. (1983) provided fecal coliform densities from domestic animal manures (Moore et al., 1989). Walker et al. (1990) used Geldreich's (1978) values for fecal coliform densities in manure; they also provided suggested values for the fraction of manure incorporated by month. EPA (2000) provided example values for fecal coliform shed from domestic animals and wildlife, fecal coliform production rates associated with urban areas (road; commercial; single-family low and high density; and multi-family residential), and supporting information on septic.

Users of the SELECT model (e.g., McFarland and Adams, 2014; Riebschleager et al., 2012; McKee et al., 2011; Teague et al., 2009) provided example shedding rates for *E. coli*, based on fecal coliform production rates for domestic animals and wildlife (EPA, 2001b), assuming a fecal coliform-to-*E. coli* conversion factor of 0.5 recommended by Doyle and Erikson (2006). Riebschleager et al. (2012) and Teague et al. (2009) provided production rates for septic (on-site wastewater treatment) systems, and Riebschleager et al. (2012) also considered urban development and built areas including low-, medium- and high-density land use consisting of single- and multi-family housing, commercial service, industrial and utilities/transportation.

Kim et al. (2016) provided a summary of *E. coli* die-off rates from published literature (Crane and Moore, 1986; Wang et al., 2004; Meals and Braun, 2006; Gu et al., 2012; Blaustein et al., 2013; Martinez et al., 2013; Olandeinde et al., 2014). Based on Moore et al. (1988), Walker et al. (1990) and Moore et al. (1989) tried to account for bacterial die-off in stored manure.

5.3 MSM OUTPUT REQUIREMENTS

Information produced by MSM are summarized in Table 23:

1. Microbial (k) loading rate by subwatershed (i) by month (q), summed across all domestic animals (m) and wildlife (n) for each land-use type (ℓ) without die-off (a.k.a. MON-ACCUM in HSPF)
2. Maximum microbial (k) storage per land-use type (ℓ) area per subwatershed (i) by month (q), summed across all domestic animals (m) and wildlife (n), adjusted for die-off (a.k.a. MON-SQOLIM in HSPF)
3. Microbial (k) loading rate of domestic animal beef cattle ($m = 2$), shedding to streams by subwatershed (i) by month (q)
4. Average septic flow rate to the stream by subwatershed (i)
5. Microbial (k) loading rate to the stream from leaking septic systems by subwatershed (i).

Table 23. Microbial Source Module output variables [Table 6 re-written using indices in equations.]

Index	Definition
AccumulationRateMonth	Rate of microbial (k) accumulation per area without die-off on the land surface by land-use type (ℓ) by month (q) by subwatershed (i) summed across all domestic animals (m) and wildlife (n) (a.k.a. ACQOP-Month in HSPF)
BeefCattleStreamMicrobeRate	Microbial (k) loading rate of beef cattle ($m=BeefCattle$) shedding into a stream by subwatershed (i) by month (q)
SepticStreamFlowRate	Average septic flow rate to the stream by subwatershed (i)
SepticStreamLoadingRate	Microbial (k) loading rate to the stream from leaking septic systems by subwatershed (i)
StorageLimitMonth	Maximum microbial (k) storage per area by subwatershed (i) by month (q) by land-use type (ℓ), summed across all domestic animals (m) and wildlife (n), adjusted for die-off (removal) (a.k.a. SQOLIM-Month in HSPF)

5.4 TRANSFORMATION OF LATITUDE-LONGITUDE COORDINATES TO SUBWATERSHED DESIGNATIONS

Microbial source characterization identifies types and locations of sources and information that capture microbial loadings and influence fate and transport in the watershed, including locations and types of microbial sources, shedding and production rates, and die-off. Microbial sources include domestic animals, wildlife, septic systems, point sources (WWTPs and POTWs), and urban loadings. Microbial source characterization data supplied by the user are captured in Tables 13 through 22. Example file

formats that document locations of farms containing domestic animals; point sources that discharge directly to the stream; and septic system, output, and boundary points as a function of latitude and longitude are provided in [Tables 13, 14, and 15](#), respectively. Latitude and longitude are used because the data-gathering process on microbial sources is typically determined prior to watershed delineation and can dictate how delineation proceeds. For example, output and boundary locations are of particular importance and, therefore, must be identified in advance. SDMPB automatically delineates a watershed into subwatersheds, accounting for the output and boundary locations ([Table 15](#)); overlays latitude-longitude locations defined in [Tables 13, 14, and 15](#); and assigns these locations to subwatersheds. If locations are outside the watershed boundary, these data will be ignored. SDMPB automates delineation and overlaying of source locations, so mapping microbial sources and output and boundary points align exactly to the correct subwatersheds without user intervention.

A transformation of information is required between components because data produced and consumed by two components (e.g., databases or modules) typically do not exactly align by name or indices. An example is the conversion of Latitude-Longitude locations to subwatershed locations: Latitude-Longitude locations of point sources, domestic animals, and septic systems must be correlated to specific subwatershed locations [i.e., indexed to the subwatershed (i)]. A glossary of these external variables, used by the SDMPProjectBuilder to develop spatially based input for the MSM, is presented in [Table 24](#). [Table 25](#) presents a mapping of names and Indices of relevant parameters/variables from the SDMPProjectBuilder output (SDMPBOutput, [Table 3](#)) to the MSM input (MSMInput, [Table 3](#)). Latitude-Longitude coordinates (i.e., AnimalLat-AnimalLong, PointLat-PointLong, SepticLat-Septic-Long) are overlaid on subwatershed polygons, and subsequently re-designated by Subwatershed (i). The number of domestic agricultural animals by latitude-longitude pairs (AnimalLat, AnimalLong) and domestic animal name (m) (each row in [Table 13](#)) within a subwatershed (i) is summed as follows:

$$\text{NumberOfAnimals}_{i,m} = \sum_{\text{within } i} (\text{AgriculturalAnimalNumber}_m) \quad (37)$$

where

- $\text{AgriculturalAnimalNumber}_m$ = Number of domestic agricultural animals (m) by latitude (AnimalLat) and longitude (AnimalLong) pair (Number of domestic animals)
- AnimalLat = Latitude associated with a domestic animal location by Domestic Animal (m) (Coordinates)
- AnimalLong = Longitude associated with a domestic animal location by Domestic Animal (m) (Coordinates)
- i = index on Subwatershed
- m = index on domestic animal name

The number of septic systems by subwatershed (i) is a summation of the septic systems (each row in [Table 15](#)) corresponding to latitude-longitude pairs (SepticLat, SepticLong) within a subwatershed (i):

$$\text{SepticNumber}_i = \sum_{\text{within } i} (\text{SepticByLocation}) \quad (38)$$

Where

- SepticNumber_i = Number of septic systems associated with Subwatershed (i) (Number of septics)

- SepticByLocation = Septic system within subwatershed (i), where each row in [Table 15](#) constitutes a septic system (Boolean Number)
- SepticLat = Latitude associated with a septic system location (Coordinates)
- SepticLong = Longitude associated with a septic system location (Coordinates)

The assessment within subwatersheds for urbanized areas is based on the fraction of the urbanized area attributed to each sub-urbanized area by subwatershed (i.e., ratio of the urbanized subcategory area and total urbanized area). Fractions must total 1.0:

$$\text{AreaFraction}_{i,\ell=4,r} = (\text{BuiltupArea}_{i,\ell=4,r}) / (\text{Area}_{i,\ell=4}) \quad (39)$$

where

- BuiltUpArea_{k,ℓ=4,r} = Area associated with each Urbanized land type (ℓ = 4) by Subwatershed (i), indexed by the Urbanized subcategories (r) (Area)

Table 24. Glossary of External Parameters/Variables that are used by the SDMPProjectBuilder to develop spatially based input for the Microbial Source Module [Alphabetic descriptors in parentheses refer to the glossary presented in [Table 7](#) from main body of report.]

Parameter/Variable Name	Parameter/Variable Description
AgriculturalAnimalNumber	Number of domestic agricultural animals associated with a designated latitude (AnimalLat) and longitude (AnimalLong)
AnimalLat	Latitude associated with a domestic animal location by Domestic Animal Name (Agricultural)
AnimalLong	Longitude associated with a domestic animal location by Domestic Animal Name (Agricultural)
BuiltUpArea	Area associated with each Urbanized land type (ℓ = 4), indexed by the Urbanized subcategories (r)
SepticLat	Latitude associated with a septic system location
SepticLong	Longitude associated with a septic system location

Table 25. Mapping of Names and Indices of relevant parameters/variables from the SDMPB output to the MSM input

Naming Convention for Names and Indices of SDMPProjectBuilder Input Ontological Metadata		Names and Indices of the SDMPProjectBuilder Output (SDMPBOutput) and Microbial Source Module Input (MSMInput) Ontological Metadata		
Name	Index 1	Name	Index 1	Index 2
AnimalLat	DomesticAnimalName	Subwatershed		
AnimalLong	DomesticAnimalName			
AgriculturalAnimalNumber	DomesticAnimalName	NumberOfAnimals	Subwatershed	Agricultural
SepticLat		Subwatershed		
SepticLong				
SepticNumber		SepticNumber	Subwatershed	

5.5 WORKFLOW SET-UP AND EXECUTION

Coupled with microbial properties data contained in the CSV data files and D4EM-retrieved data, SDMPB produces a delineated watershed containing a suite of GIS map layers that include farms with domestic-animal types, numbers, and locations; septic-system locations; NHDPlus flowlines; subwatersheds; and waterbody network; in addition to mapping elevation (e.g., slope), soil types, land-use types, meteorological (MET) stations, gaging stations, and multiple legal boundaries (Table 26).

5.5.1 Automating Watershed Delineation and Microbial Source Mapping

The workflow that contains MSM allows for automated watershed delineation and collation of microbial sources within each subwatershed. This allows users to easily change the number and size of subwatersheds, and microbial sources are automatically placed within the correct subwatershed and collated accordingly; users, therefore, do not have to manually assign sources (domestic animals, humans, engineered point sources or septics) to subwatersheds. The SDMPB/D4EM automates watershed delineation and microbial source mapping as it:

- links to a GIS system (MapWindow) to visualize map layers of data.
- accesses and retrieves web-based data from sources outlined in Table 26 to automatically create input files for MSM, including automatic delineation of watersheds into subwatersheds, areas for and land-use types in each subwatershed, etc.
- provides user control for watershed delineation as it relates to number of subwatersheds, minimum subwatershed size, and minimum stream length. The latter two prevent watershed modeling of areas and streams that are too small, although the smallest areas are those defined by the minimum NHD delineations.
- allows users to specify intermediate (e.g., gaging or monitoring) locations within the watershed to ensure that the automated delineation process has subwatershed boundaries going through those points.
- accesses and retrieves user-defined local data (Tables 13 through 22) which compute microbial loading rates distributed spatially and temporally by subwatersheds.
- allows manual manipulation of input data for more refined, boutique assessments.
- assigns North American Land Data Assimilation System (NLDAS) radar or National Climatic Data Center (NCDC) land-based MET data to individual subwatersheds (Kim et al., 2014).
- allows users to designate snow accumulation/melt, microbial fate and transport, and simulation time increments (e.g., hourly, daily, monthly, or annually).
- creates the MSM input file.
- facilitates linkage between microbial sources and loadings through MSM, with fate and transport modeling within a mixed-use watershed.
- creates map layers to visualize locations of subwatersheds, land-use types, farms, domestic animals, septics, engineered point sources, monitoring and gaging stations, and MET stations. An example watershed with delineated subwatersheds, water body network, gaging stations, and farms with domestic-animal and septic-system locations is presented in Figure 6.

Table 26. Databases automatically accessed and used by D4EM and SDMPB

NASA NLDAS (North America Land Data Assimilation System)
USGS NLCD (National Land Cover Data)
USGS NWIS (National Water Information System)
USGS NAWQA (National Water-Quality Assessment program)
USDA NASS (National Agricultural Statistics Service)
USDA Soils
SSURGO (Soil Survey Geographic database)
STATSGO (State Soil Geographic dataset)
NOAA NCDC (National Climatic Data Center)
NOAA NDBC (National Buoy Data Center)
EPA STORET (STOrage and RETrieval)
EPA Waters Web Services
EPA BASINS (Better Assessment Science Integrating Point and Nonpoint Sources)
Land use/land cover
Urbanized areas
Populated place locations
Reach File version 1 (RF1)
Elevation [DEM (Digital Elevation Model)]
National Elevation Dataset (NED)
Major roads
USGS HUC (Hydrologic Unit Code) boundaries
Accounting unit
Cataloging unit
Dam sites
EPA regional boundaries
State boundaries
County boundaries
Federal and Indian lands
Ecoregions
Legacy STORET
NHDPlus
NHD (National Hydrography Dataset)
NED (National Elevation Dataset)
WBD (Watershed Boundary Dataset)

NatureServe

EPA = U.S. Environmental Protection Agency
NASA = National Aeronautics and Space Administration
NOAA = National Oceanic and Atmospheric Administration
USDA = U.S. Department of Agriculture
USGS = U.S. Geological Survey

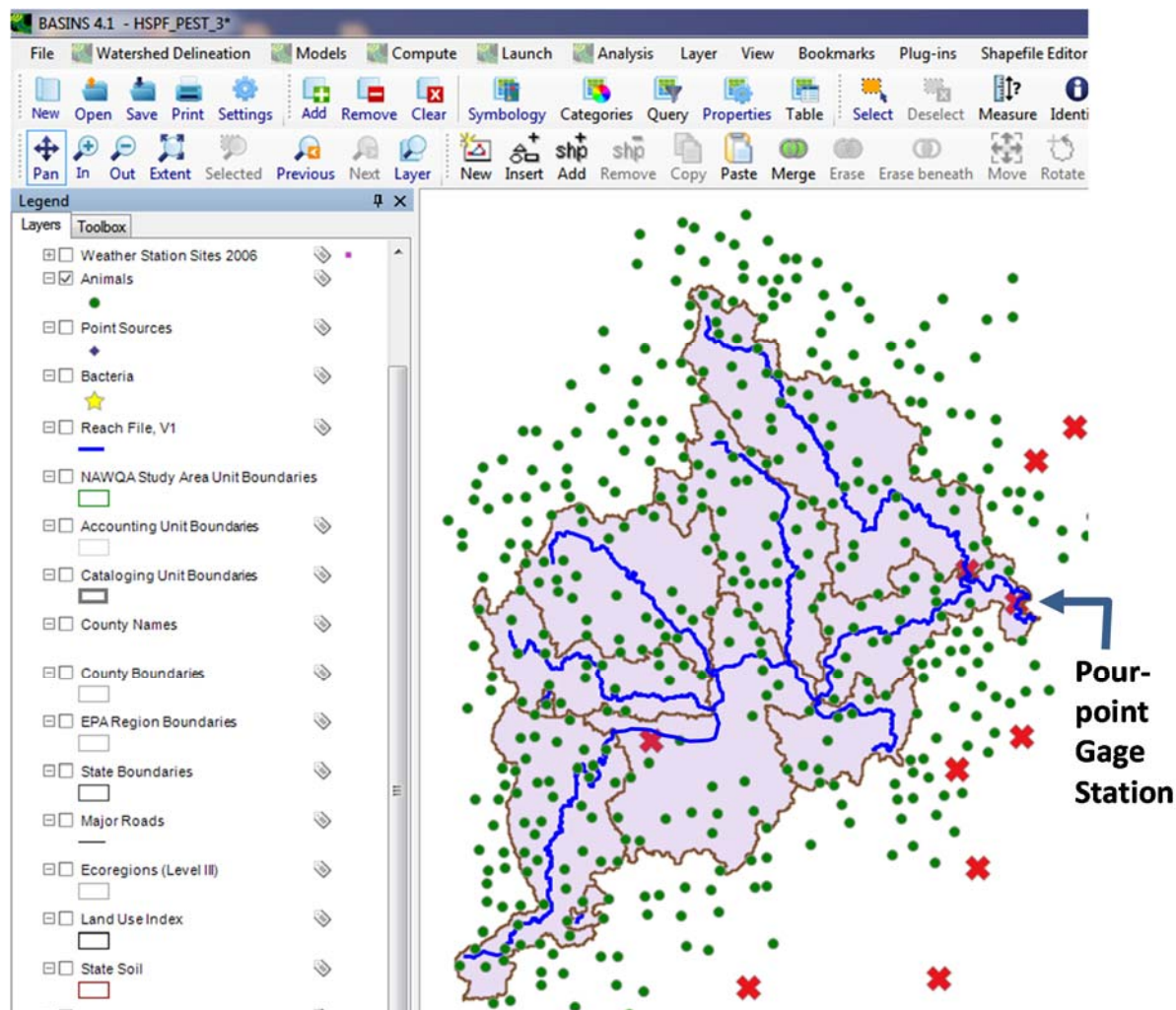


Figure 6. Example watershed with subwatersheds (brown outline), water body network (blue lines), gaging stations (X), and farms with domestic animal and septic system locations (●) (after Wolfe et al., 2016)

5.5.2 Workflow Instantiation

Instantiation of the QMRA workflow (Figure 5) begins with the user initiating SDMPB/D4EM; navigating the United States by state, county, and HUC-8; and selecting the pour point, HUC-12, or HUC-8 of interest. With user-defined simulation output intervals (hourly, daily, etc.), simulation start and end times, selected data sources and a pour-point selection, SDMPB automatically identifies the upstream basin boundary and registers the 12 user-defined CSV data files (Tables 13 through 22). Users can also control the number of subwatersheds with user-defined minimum subwatershed areas and stream lengths. The SDMPB manages data acquisition from standard national databases using D4EM and caches web-based data (Table 26). Coupled with boundary and output points (e.g., Table 15) and user-defined minimum stream lengths and subwatershed areas, the SDMPB produces a delineated watershed, as illustrated in Figure 6, which contains a suite of GIS map layers that include gaging stations, farms with domestic-animal and septic-system locations, waterbody network, elevation (e.g., slope), soil types,

land-use types, and MET stations. Number and type of domestic animals, as well as wildlife density, were collected *a priori* and thus known beforehand; although these data exist, they are not always routinely known due to privacy/security. The MSM develops microbial loadings (e.g., Cells/Area/Time), adjusted for die-off, to the overland subwatershed areas by land use and to instream (e.g., Cells/Time) locations within a watershed.

The SDMPB automatically pre-populates input needs of the fate and transport watershed model HSPF by automatically creating its input UCI file, a collection of geospatial data files, a DotSpatial-based project file, and a MapWindow-based project file (*.wmpri) currently used by BASINS. Using HSPF/BASINS Windows interfaces, non-spatially related data may be modified without requiring re-delineation. For example, if a point source exists within the watershed, its time series loadings can be registered within BASINS prior to HSPF execution, replacing the annually averaged default values provided in the CSV file (Table 20). HSPF is then executed, creating flows and microbial concentrations both spatially and temporally distributed throughout the watershed.

BASINS (e.g., EPA, 2013b, 2013c) provides a user interface and visualization tool for HSPF, and accesses gage data for subsequent inverse modeling. PEST uses HSPF flow and microbial density simulations with monitored flow and microbial density data at a point of interest in the watershed for an initial calibration that will require a final manual calibration. HSPF flow calibration has been discussed by Duda et al. (2012). Key calibration parameters produced by MSM and consumed by HSPF included loadings by microbe and by land-use type, maximum microbial storage accumulation on the land surface, and point source loading rates to the stream from septic systems and direct shedding (Table 23). Key HSPF microbial calibration parameters include rate of surface runoff (which will remove 90 percent of stored microbes per hour), microbial densities in interflow and active groundwater outflow, instream first-order die-off rate and temperature correction for first-order die-off.

The Sara Timeseries Utility (Aqua Terra, 2011) reads a time series of microbial densities in a BASINS WDM file and converts it to a txt file in a Unicode Transformation Format 8-bit (UTF-8) format, which is a character encoding capable of encoding all possible characters in Unicode using 8-bit code units. MRA-IT consumes txt files using UTF-8. Using a time series of pathogen densities, MRA-IT (Soller et al., 2008, 2004; Soller and Eisenberg, 2008) computes risk of infection during a user-identified exposure period.

5.6 ONTOLOGICAL RELATIONSHIPS BETWEEN VARIABLES, EQUATIONS, AND COMPONENTS

Using the ontological metadata dictionaries, a definitive relationship can be established between variables, components within the workflow, equations, metadata, and assumptions. Even when a variable is shared between multiple components (e.g., SDMPProjectBuilder and Microbial Source Module), relationships can be established using an approach similar to a Resource Description Framework (RDF) triple (Price, 2004). An RDF format is the standard for encoding metadata and other knowledge on the semantic web (GitHub, 2014), and an RDF triple consists of the 1) subject that identifies the object the triple is describing, 2) predicate that defines the piece of data in the object to which we are giving a value, and 3) object that is the actual value. In other words, a subject and an object are linked by a predicate. For our purposes, Elag and Goodall (2013) describe this as a “3-ary” because neither the equation, variable (symbol), nor component can be considered the primary subject.

Expanded examples of a “3-ary” are presented in [Figures 7 and 8](#); [Figure 7](#) illustrates the relationships between the variable “AreaFraction” and the components (SDMProjectBuilder and Microbial Source Module) and equations that define and use it (Equations 41 and 23, respectively). The ontological metadata associated with AreaFraction is provided in [Table 3](#). [Figure 7](#) illustrates that both equations and components link to the same variable, through which the metadata and assumptions are described. Interesting features that are captured include 1) linkage of two different modules (SDMProjectBuilder and Microbial Source Module); 2) definition of the variable in one module which registers it as output (SDMProjectBuilder), and consumption of same variable in another module as boundary-condition input (Microbial Source Module); and 3) demonstration of how one accounts for input as “module-specific” (MSMInput ontological dictionary) and as a boundary condition (SDMPBOutput ontological dictionary), where this is an example of the boundary condition case. Since multiple modules could be linked to this variable, this figure is not necessarily limited to two.

This type of “3-ary” can be applied to all registered variables. For example, [Figure 8](#) presents a “3-ary” for an internal variable which is defined and used wholly within the MSM (i.e., TotalGrazeDays). Its interesting features include the variable being defined and used in the same module, and its indices (SDMOutput ontological dictionary) being associated with a different ontological dictionary than where the variable is registered (MSMInternalVariables ontological dictionary). In summary, model developers are able to focus on the model code itself rather than linkages between components by expressing model and variable descriptions and assumptions within the ontology.

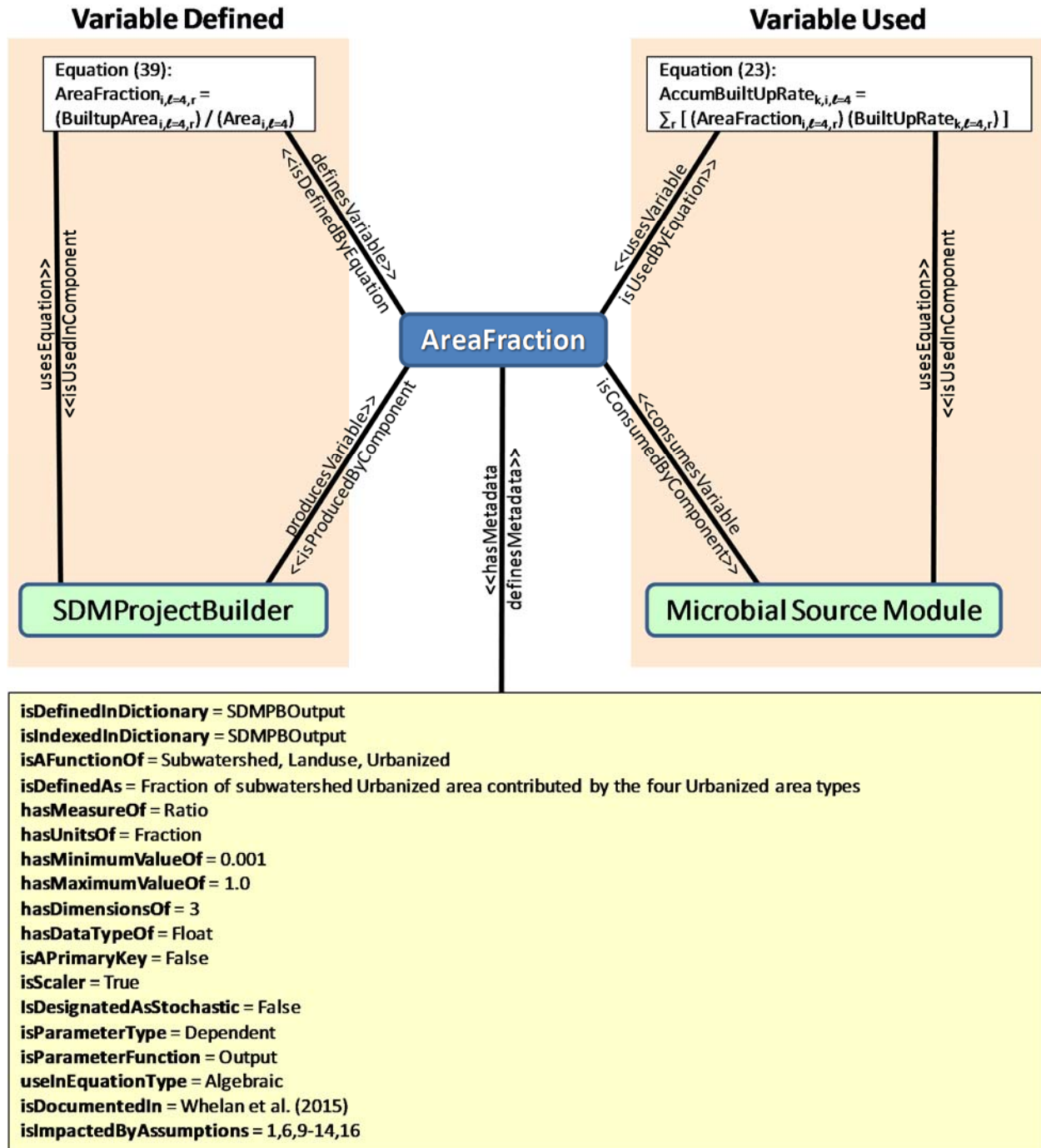


Figure 7. Relationships between the Variable "AreaFraction" and Components, Equations, Metadata, and Assumptions

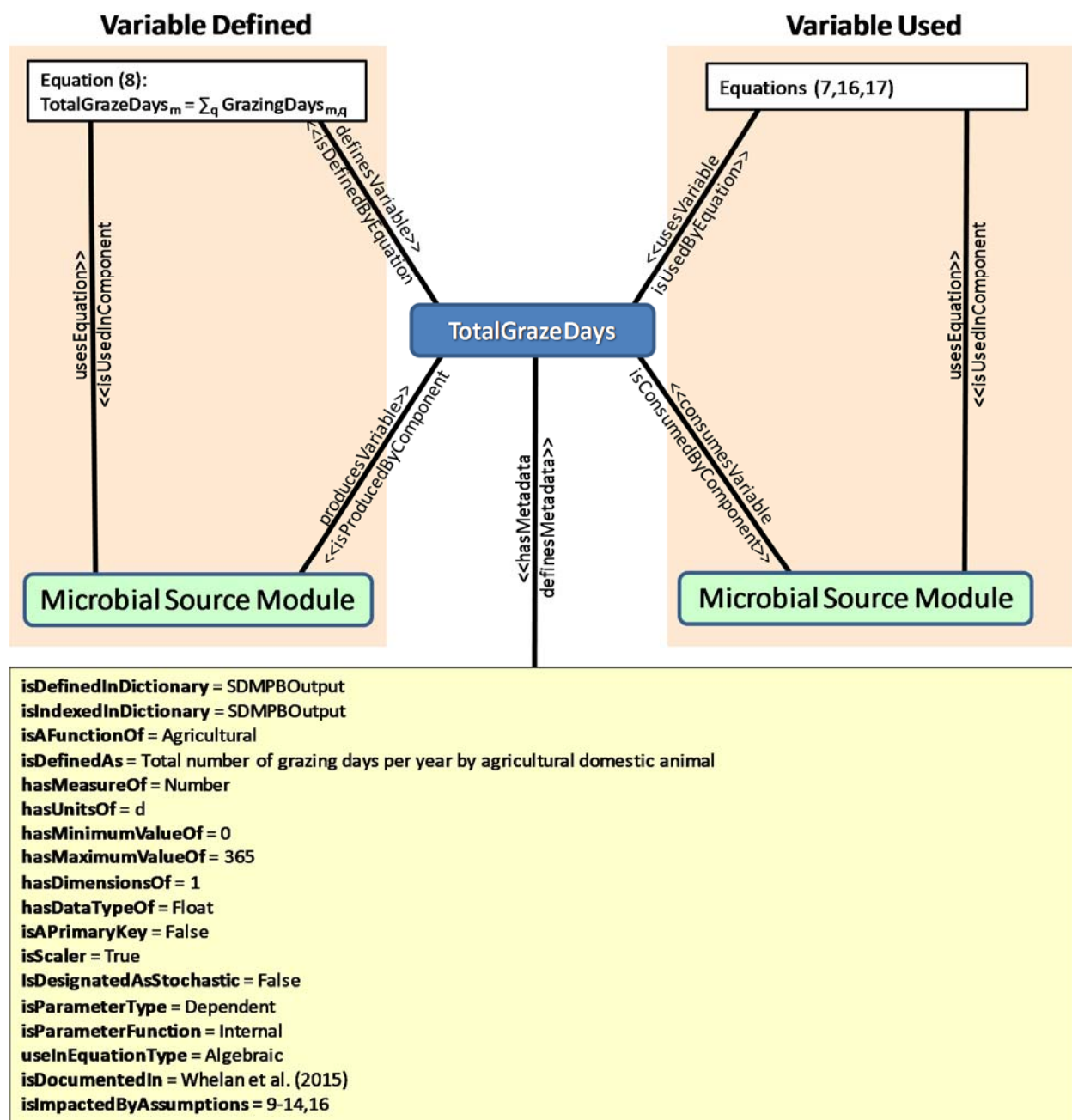


Figure 8. Relationships between the Variable “TotalGrazeDays” and Components, Equations, Metadata, and Assumptions

6. MAPPING THE MICROBIAL SOURCE MODULE ONTOLOGICAL METADATA DICTIONARIES TO AN EXTENSIBLE MARKUP LANGUAGE DOCUMENT

The concept of an ontological framework for documenting science software “products” (e.g., components, models, databases, assessments, etc.) lends itself to describing knowledge about the product and relationships between product concepts (see [Table 1](#), for example). A science-based software product communicates science theory and software usability where the traditional means were text-based, although technology is changing to digitized formats that facilitate not only product understanding but automated discovery, evaluation (for a purpose), and integration (with other products). To achieve digitized documentation for communication, discovery, evaluation, and integration, an ontological framework provides a structured and possibly standardized way to combine data, taxonomy, and relationships among concepts and data. Hence, the WRC ontology framework described by [Elag and Goodall \(2013, 2012\)](#) encompasses many elements (data, taxonomy, concepts, and relationships) in one format such as OWL (web ontology language).

Spreadsheets (illustrated by [Tables 3, 4, and 11](#)) combine essential variables with metadata and intra- and inter-parameter relationships between variables (see [Figures 7 and 8](#)). They have traditionally been used because they are intuitive and most software developers are comfortable with them, although they are not the only format that could capture the ontological metadata. By agreeing on a format to express data exchange, tools can be developed that facilitate the process to higher-level ontological frameworks (e.g., OWL). With standardization, for example, user-friendly, graphical user interfaces (GUIs) can be developed from a spreadsheet to capture ontological metadata, as illustrated in [Figure 9](#) with the FRAMES Dictionary Editor ([Whelan et al., 2014a](#)). Likewise, spreadsheet-based ontological metadata can be easily converted to GUIs, as illustrated by the Dictionary Registration Tool ([Pelton, 2009](#)), resulting in interchangeable forms of the same ontological metadata (i.e., spreadsheet to GUI and vice versa).

Moving toward controlled vocabularies to name data elements and associated metadata, standardized tools can facilitate linking controlled vocabularies (and definitions) to individual software product digitized formats. Coupled with taxonomy (classification of concepts) relative to science software, a more complete ontology can be documented and tools developed to compare, merge, and produce such files and formats, as illustrated by expression of the WRC ontology using Protégé ([Protégé, 2014](#)). [Figure 10a](#) illustrates interchangeable forms describing ontological metadata (or schema) between spreadsheets, GUIs, and ontology editors.

When coupled with input values, data transfer with metadata can ensure proper quality control within and between components; not only is the value known, but its metadata (description, units, range, relationships to other parameters/variables, etc.) accompanies it. Standardization facilitates multiple formats for expressing values with their ontological metadata ([Figure 10b](#)). For example, it allows standardized user interfaces to deliver input to or produce data from models (illustrated in [Figure 11](#)) which is the user interface for FRAMES’s Data Client Editor ([DCE, 2010](#)) that captures three input variables (TimePts, CumMass, and TotalFlux) with metadata and values associated the ChemAquiferTotalFlux dictionary. This information can be easily converted into a flat file (e.g., csv, txt) or expressed electronically, as illustrated by the MSM output captured in the Extensible Markup Language (XML) in [Figure 12](#) (see bottom third of figure).

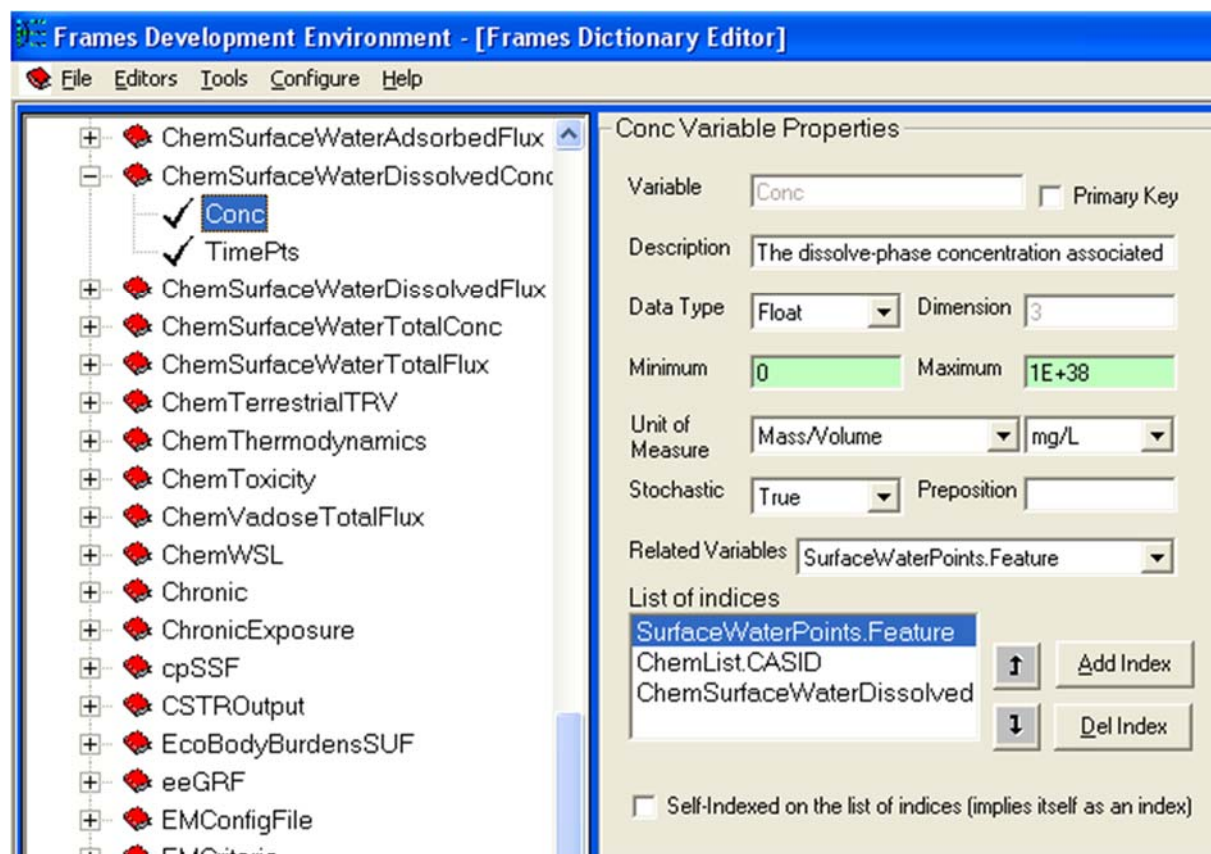


Figure 9. Example of Ontological Metadata captured by the FRAMES Dictionary (DIC) Editor (Whelan et al., 2014a)

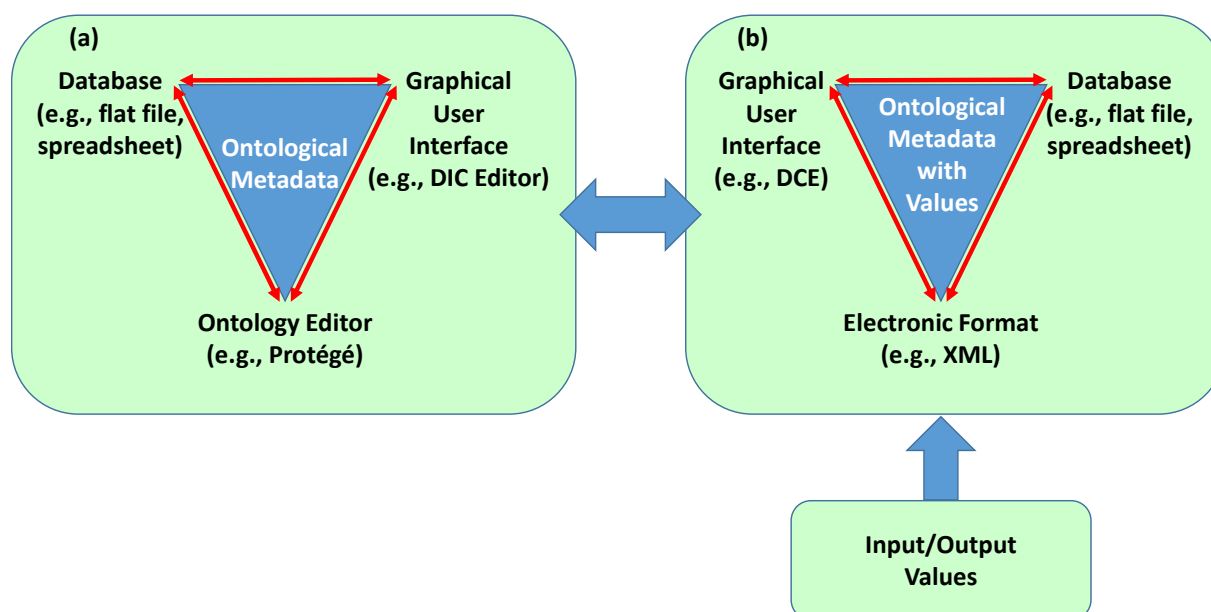


Figure 10. Interchangeable forms (a) describing ontological metadata (or schema) and (b) documenting instances of a dataset related to the schema (After Parmar et al., 2016)

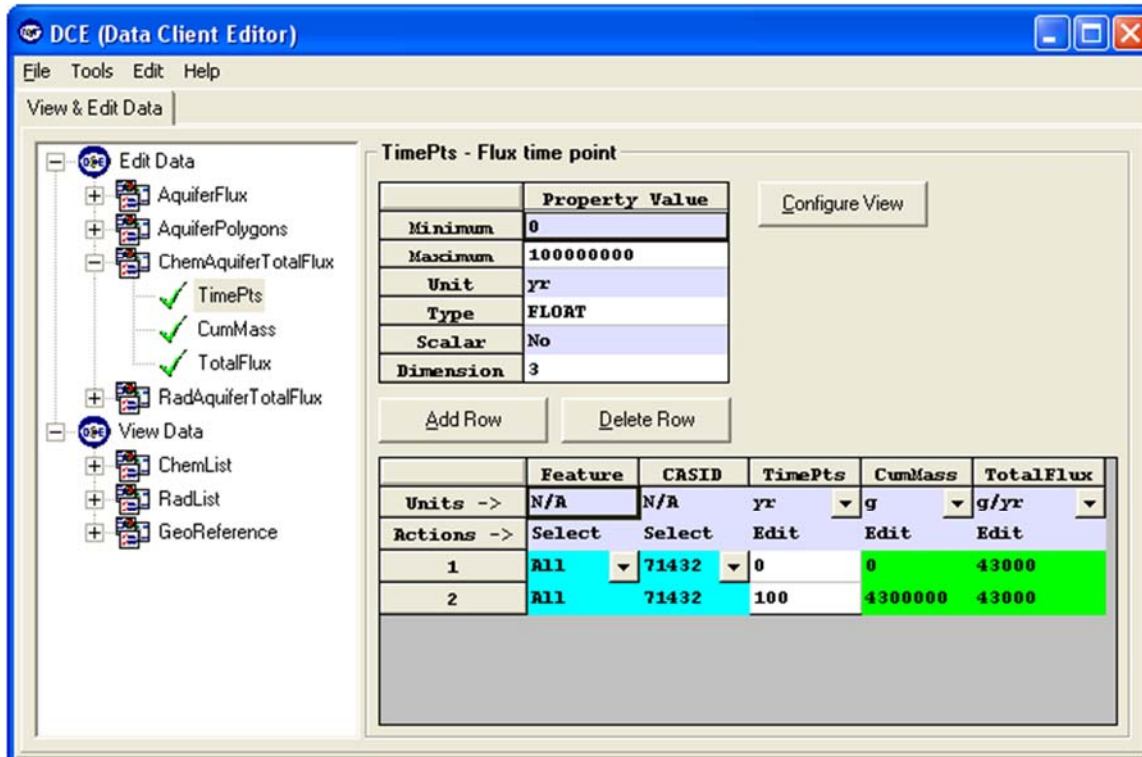


Figure 11. Standardized Graphical User Interface for FRAMES' Data Client Editor (DCE, 2010), capturing Ontological Metadata and Values for Three Variables (TimePts, CumMass, and TotalFlux) associated with the ChemAquiferTotalFlux Dictionary

XML is 1) a standard or set of rules that governs encoding of documents into an electronic format (Difference Between, 2014) that is human- and machine-readable; 2) a textual data format with strong support via Unicode for different human languages; and 3) widely used to represent arbitrary data structures such as those in web services (XML Wikipedia, 2014). An XML document captures rules in a readable form and is compared to the XML schema (XSD) developed to execute the web service. The purpose of comparison is to verify the syntax and validate the structure of an XML document. The purpose of this section is to illustrate how an MSM ontological metadata dictionary maps to its corresponding XML document, so the model can be executed as a web service.

There is a logical, natural mapping of MSM variables to an XML document. To facilitate access and integration, MSM software has been designed with object-oriented principles and is "published" as a Representational State Transfer (REST, 2015) web service. The web service consumes XML input and produces XML output which can be accessed directly via browser add-ons such as Postman for Chrome. The most common way to consume the web service is through a custom desktop or web client program. The web service is platform and programming language agnostic. For example, Appendix A presents an example and partial listing of MSM input variables and metadata directly with the MSM XML document. The only metadata directly captured by the XML document are names, units, and indices, including each parameter's hierarchical relationships between indices. The XML document also captures the value for each variable and lets the user include comments/explanations. Input

parameter/variable names are easily mapped from an ontological dictionary such as Table 3 to an XML document. The parameter/variable name represents the lowest level in the hierarchy, telescoping from the highest mapped index (Index 1) to the lowest index such as Index 3. Figure 13 illustrates how the metadata for input variables “Area” and “AreaFraction” are mapped to the XML document.

“Area” is a function of the Subwatershed, of which a watershed contains one or more subwatersheds, and LandUse, of which there are four types (Cropland, Pasture, Forest, and Urbanized). For each subwatershed, therefore, an area is assigned to each land-use type (see Figure 13.) In addition to being a function of Subwatershed and LandUse, “AreaFraction” is a function of the Urbanized land-use type, of which there are four (CommercialAndServices, Residential, MixedUrban, TransportationCommunicationUtilities). “AreaFraction,” therefore, is captured in the XML document under Subwatershed, LandUse, and Urbanized for each Urbanized land-use type, with Figure 13 illustrating the telescoping indices. A similar procedure can be followed when mapping the remaining MSM input variables listed in Table 3, as illustrated by Appendix A. Using this template, the MSM ontological metadata output in Table 4 can also be captured in an XML document.

Subwatershed Areas File: Areas.txt
 FC Production Rates File: FCProdRates.txt
 Ag Animal Count File: AnimalSub.txt
 Wildlife Densities File: WildlifeDensities.txt
 Manure Application File: ManureApplication.txt
 Grazing Days File: GrazingDays.txt
 Septics Data File: SepticsData.txt
 Dieoff Rates File: MonthlyFirstOrderDieOffRateConstants.txt
 Manure Application File: PointSources.txt

Outputs have been produced. You may view them using the drop down box on this form.

Get Output Display output for: Pasture

SubWatershedID	JanAccum	FebAccum	MarAccum	AprAccum	MayAccum	JunAccum	JulAccum	AugAccum	SepAccum	OctAccum
P1	3085712045.528...	3415905523.620...	3085712045.528...	4596099815.344...	4459130668.876...	3446759819.438...	3344032959.587...	3344032959.587...	3534588930.218...	25004502639.06...
P2	1326890074.330...	1468638341.223...	1326890074.330...	5924322062.432...	5865522781.351...	4428928414.925...	4384828954.114...	4384828954.114...	5468627634.052...	14685414943.55...
P3	50746112.65865...	55764669.37208...	50746112.65865...	1028749881.918...	1026668110.244...	768955732.4834...	767394403.7281...	767394403.7281...	1012616151.447...	1338933861.302...
P4	238930975.1139...	264112195.6618...	238930975.1139...	3909985764.015...	3899540220.676...	2920099215.098...	2912265057.595...	2912265057.595...	3829032803.142...	5466371721.436...
P5	519724939.8360...	574991228.0327...	519724939.8360...	7189477629.541...	7166552354.437...	5368924755.360...	5351730799.032...	5351730799.032...	7011806747.486...	10605343620.01...

OR Run Module Using XML Input File

XML Input File: MSMInputExample.xml

Run Module Using XML

```

<Output>
<Subwatershed>
<ID>P2</ID>
<MonthID>
<January>
<Forest>
<Accum units="Microbial Counts/d">257971875</Accum>
<SQOLIM units="Microbial Counts/Acre">92869875</SQOLIM>
</Forest>
<Cropland>
<Accum units="Microbial Counts/d">15465933829.531027</Accum>
  
```

Figure 12. Microbial Source Module Graphical User Interface: 1) input data files (top third), 2) tabular form of output results, values (middle third), and 3) XML-based output results, values and ontological metadata.

Acknowledgements

Thanks are extended to Ms. Fran Rauschenberg of EPA for editing the document. Mention of trade names or commercial products does not constitute endorsement or recommendation for use. This document has been subjected to Agency review and approved for publication.

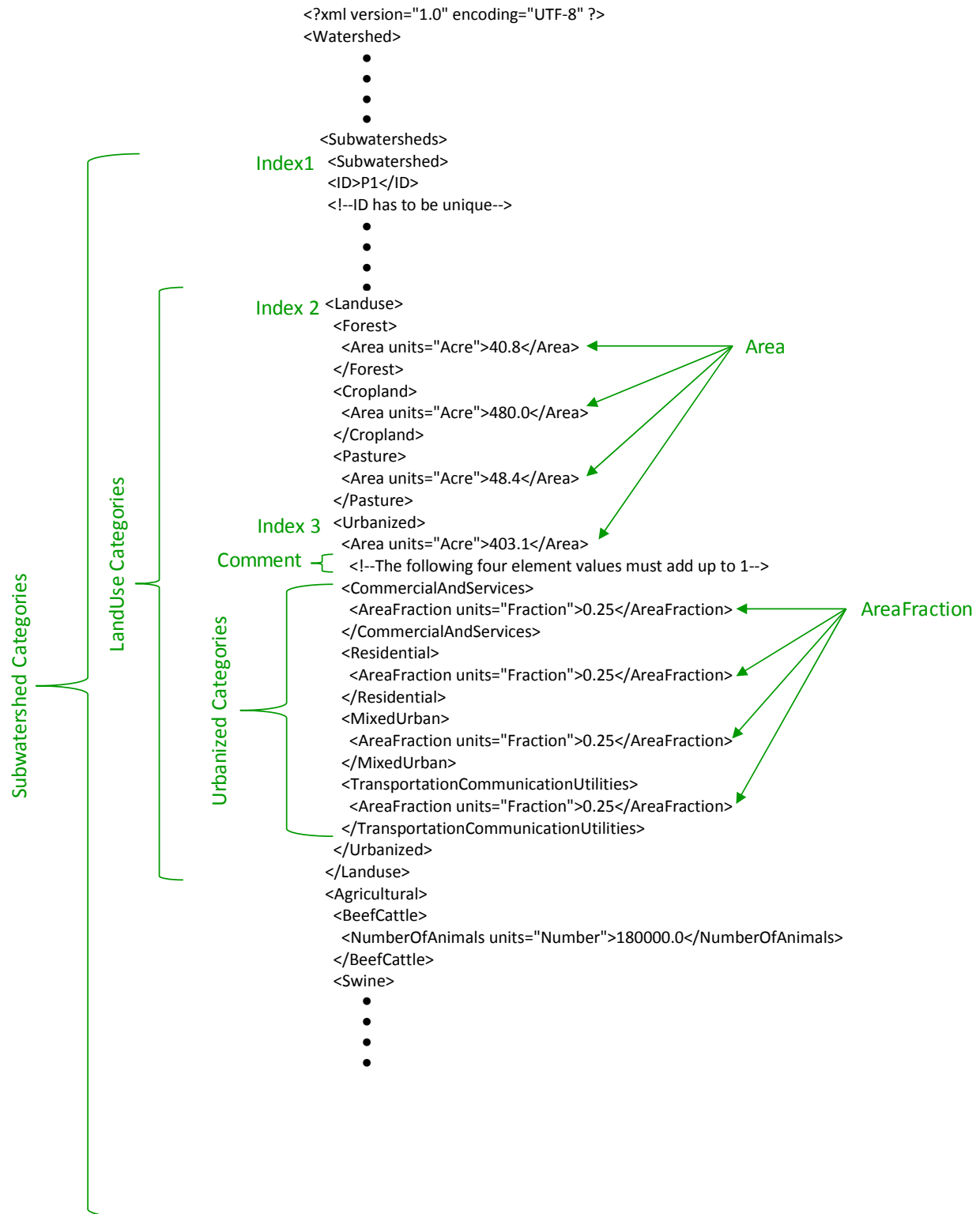


Figure 13. Mappings to the XML document (See Appendix A) of the Metadata for Input Variables “Area” and “AreaFraction” (see Table 6)

REFERENCES

- Ames, D. 2010. MapWindow: Moving MapWindow 6 to DotSpatial.
<<http://www.mapwindow.org/phorum/read.php?13,20244>> (last accessed 21.09.14).
- Aqua Terra. 2011. SARA Timeseries Utility
<<http://www.aquaterra.com/resources/downloads/saratsutility.php>> (last accessed 12.08.15).
- API. 2015. Application Programming Interface.
<http://en.wikipedia.org/wiki/Application_programming_interface> (last accessed 04.02.15).
- ASAE (American Society of Agricultural Engineers) Standards. 2005. Manure production and characteristics. ASAE D384.2. St. Joseph, MI USA. <http://extension.psu.edu/animals/dairy/nutrient-management/certified-dairy/tools/manure-prod-char-d384-2.pdf> (last accessed 07.10.2016).
- Babendreier, J.E. 2010. QA/QC Management Protocols for iemTechnologies Software Development: Summary of Guiding Principles and Best Practices for Quality Assurance in Modeling. U.S. Environmental Protection Agency, National Exposure Research Laboratory, Ecosystems Research Division, Athens, GA (March).
- Babendreier, J.E., Castleton, K.J. 2005. Investigating uncertainty and sensitivity in integrated, multimedia environmental models: Tools for FRAMES-3MRA. *Environ Model Softw* 20 (8), 1043-1055.
- Bicknell, B.R., Imhoff, J.C., Kittle, Jr., J.L., Jobes, T.H., Donigian, Jr., A.S. 2005. HSPF Version 12.2 User's Manual. AQUA TERRA Consultants, Mountain View, CA.
- Bicknell, B.R., Imhoff, J.C., Kittle, J.L., Donigian, A.S., Jr., Johanson, R.C. 1997. Hydrological simulation program – FORTRAN, user's manual for version 11. EPA/600/R-97/080, U.S. Environmental Protection Agency, Athens, GA, 755 p.
- Blaustein, R.A., Pachepsky, Y.A., Hill, R.L., Sheldon, D.R., Whelan, G. 2013. Escherichia coli survival in waters: Temperature dependence. *Water Res* 47 (2), 569–578.
- Boehm, A.B., Grant, S.B., Kim, J.H., Mowbray, S.L., McGee, C.D., Clark, C.D., Foley, D.M., Wellman, D.E. 2002. Decadal and shorter period variability of surf zone water quality at Huntington Beach, California. *Environ Sci Technol* 36 (18), 3885-3892.
- Butler, D.M., Randells, N.N., Franklin, D.H., Poore, M.H., Green Jr., J.T. 2008. Runoff quality from manured riparian grasslands with contrasting drainage and simulated grazing pressure. *Agri Ecosyst Environ* 126, 250-260.
- CodePlex. 2014. DotSpatial. <http://dotspatial.codeplex.com> (last accessed 23.02.14.).
- Crane, S.R., Moore, J.A. 1986. Modeling enteric bacterial die-off: A review. *Water Air Soil Poll* 27 (3/4), 411-439.

C⁴Site Factory. 2011. Model-Independent Parameter ESTimation and Uncertainty Analysis (PEST). <<http://www.pesthomepage.org/>> (Last accessed 12.01.11)

DCE. 2010. FRAMES Downloads, EarthDomainDocumentation.Zip. Data Client Editor (DCE)_0710.ppt. <http://iemhub.org/resources/133/supportingdocs> (last accessed 21.03.15).

Difference Between. 2014. Difference Between XML and XSD. Difference Between.net <<http://www.differencebetween.net/technology/difference-between-xml-and-xsd/#ixzz38KfkFB44>> (last accessed 03.08.14).

Difference Between. 2012. Difference Between Variable and Parameter <<http://www.differencebetween.com/difference-between-variable-and-vs-parameter/>> (last accessed 15.03.15).

Doherty J. 2005. PEST: Model-independent parameter estimation user manual. 5th edition, Watermark Numerical Computing.

Dorner, S.M., Anderson, W.B., Gaulin, T., Candon, H.L., Slawson, R.M., Payment, P., Huck, P.M. 2007. Pathogen and indicator variability in a heavily impacted watershed. *J Water Health* 5 (2), 241-257.

Doyle, M.P., Erickson, M.C. 2006. Closing the door on the fecal coliform assay. *Microbe* 1 (4), 162-163.

Duda, P.B., Hummel, P.R., Donigian Jr., A.S., Imhoff, J.C. 2012. BASINS/HSPF: Model use, calibration, and validation. *T ASAE*. 55 (4), 1523-1547.

Elag, M.M., Goodall, J.L. 2013. An ontology for component-based models of water resource systems. *Water Resour Res*, 49 (8), 5077–5091.

Elag, M.M., Goodall, J.L. 2012. Design and Application of an Ontology for Component-Based Modeling of Water Systems, Abstract IN11B-1470. Presented at 2012 Fall Meeting, AGU, San Francisco, Calif., December 3-7.

EPA (U.S. Environmental Protection Agency). 2000. Bacterial Indicator Tool: User's guide. EPA-823-B-01-003. Office of Water, Washington, D.C.

EPA (U.S. Environmental Protection Agency). 2001a. Better assessment science integrating point and nonpoint sources, BASINS 3.0, user's manual. EPA-823-B01-001, Washington, DC, 343 pp.

EPA (U.S. Environmental Protection Agency). 2001b. Protocol for Developing Pathogen TMDLs, first edition. EPA 841-R-00-002. Office of Water, Washington, DC USA. <https://nepis.epa.gov/Exe/ZyPDF.cgi/20004QSZ.PDF?Dockey=20004QSZ.PDF> (last accessed 04.10.2016).

EPA (U.S. Environmental Protection Agency). 2009. Review of published studies to characterize relative risks from different sources of fecal contamination in recreational water. EPA 822-R-09-001. Office of Water. Washington, DC USA.

EPA (U.S. Environmental Protection Agency). 2010. Quantitative microbial risk assessment to estimate illness in freshwater impacted by agricultural animal sources of fecal contamination, EPA 822-R-10-005, Office of Water, Washington DC.

EPA (U.S. Environmental Protection Agency). 2013a. Data for Environmental Modeling (D4EM). Office of Research and Development, Athens, GA. <http://www.epa.gov/AthensR/research/d4em.html> (last accessed 10.02.14.).

EPA (U.S. Environmental Protection Agency). 2013b. BASINS/HSPF Training, Exercise 10 – Bacterial and temperature modeling. <http://water.epa.gov/scitech/datait/models/basins/upload/Exercise-10-Bacteria-and-Temperature.pdf> (last accessed 23.02.14.).

EPA (U.S. Environmental Protection Agency). 2013c. BASINS user information and guidance, BASINS tutorials and training. <http://water.epa.gov/scitech/datait/models/basins/userinfo.cfm#tutorials> (last accessed 23.02.14.).

EPA (U.S. Environmental Protection Agency). 2013d. Exposure assessment models: HSPF. <http://www2.epa.gov/exposure-assessment-models/hspf> (last accessed 23.02.14.).

EPA (U.S. Environmental Protection Agency). 2013e. BASINS (Better Assessment Science Integrating point & Non-point Sources). <http://water.epa.gov/scitech/datait/models/basins/> (last accessed 23.02.14.).

EPA (U.S. Environmental Protection Agency) and USDA (U.S. Department of Agriculture/Food Safety and Inspection Service). 2012. Microbial risk assessment guideline: Pathogenic microorganisms with focus on food and water. EPA/100/J-12/001; USDA/FSIS/2012-001. Washington, DC.

Fraser, R.H., Barten, P.K., Pinney, D.A.K. 1998. Predicting stream pathogen loading from livestock using a geographical information system-based delivery model. *J Environ Qual* 27, 935-945.

Garnett, L.W. 2012. 5 High density housing concepts. <http://www.probuilder.com/5-high-density-housing-concepts> (last accessed 07.10.16).

Geldreich, E.E. 1978. Bacterial populations and indicator concepts in feces, sewage, stormwater and solid wastes. In: Indicators of Viruses in Water and Food, Ed. G. Berg. Ann Arbor Science Publishers, Inc., Ann Arbor, MI USA 4, 51-97.

GitHub. 2014. What is RDF and what is it good for? <<https://github.com/JoshData/rdfabout/blob/gh-pages/intro-to-rdf.md#>> (last accessed 02.08.14).

Gu, Y., Molina, M., Bohrmann, T., Cyterski, M., Whelan, G., Purucker, T. 2012 Estimation of decay rates for fecal indicator bacteria and bacterial pathogens in agricultural field-applied manure. Annual Meeting of the Association of Southeastern Biologists, Athens, GA, April 04 - 06.

Haas, C.N., Rose, J.B., Gerba, C.P. 1999. Quantitative microbial risk assessment. John Wiley & Sons, Inc. New York. 449 p.

Hunter, P.R., Payment, P., Ashbolt, N., Bartram, J. 2003. Chapter 3. Assessment of risk. In: Ronchi, E., Bartram, J. (Eds.). Assessing microbial safety of drinking water: Improving approaches and methods. OECD/WHO guidance document. OECD/WHO, Paris, pp. 79-109.

IBM. 2006. What is a software architecture?

<<http://www.ibm.com/developerworks/rational/library/feb06/eeles/>> (Last accessed 17.07.14).

IEEE. 2000. IEEE Recommended Practice for Architectural Description of Software-Intensive Systems: IEEE Std 1472000. IEEE Computer Society.

JCGM. 2008. International vocabulary of metrology — Basic and general concepts and associated terms (VIM), JCGM 200, Joint Committee for Guides in Metrology, Bureau International des Poids et Mesures, Sevres Cedex, France. <<http://www.bipm.org/en/publications/guides/vim.html>> (last accessed 22.03.15.)

Johnston, J. M., McGarvey, D.J., Barber, M.C., Laniak, G.F., Babendreier, J.E., Parmar, R., Wolfe, K., Kraemer, S.R., Cyterski, M., Knightes, C., Rashleigh, B., Suarez, L., Ambrose, R. 2011. An integrated modeling framework for performing environmental assessments: Application to ecosystem services in the Albemarle-Pamlico basins (NC and VA, USA). *Ecol. Model.* 222 (14), 2471-2484.

Kashyap, V., Sheth, A.P. 2000. Information Brokering across Heterogeneous Digital Data: A Metadata-based Approach, Kluwer Academic Publishers, Boston, MA

Kim, K., Price, K., Whelan, G., Galvin, M., Wolfe, K., Duda, P., Gray, M., Pachepsky, Y. 2014. Using remote sensing and radar meteorological data to support watershed assessments comprising integrated environmental modeling. In: Ames, D.P., Quinn, N. (Eds.), Proceedings of the 2014 International Congress on Environmental Modelling and Software, San Diego, CA USA.

Kim, K., Whelan, G., Molina, M., Purucker, S.T., Pachepsky, Y., Guber, A., Cyterski, M., Franklin, D., Blaustein, R.A. 2016. Rainfall-induced release of microbes from manure: Model development, parameter estimation, and uncertainty evaluation on small plots. *J Water Health* DOI: 10.2166/wh.2016.239.

Laniak, G.F. 2012. Environmental Software Reuse and Interoperability: An Initial Review and Recommendations for the Office of Research and Development. Internal Report, Ecosystems Research Division, National Exposure Research Laboratory, Office of Research and Development, U.S. Environmental Protection Agency, Athens, Georgia, December 31, 2012 (Final Draft).

Laniak, G.F., Olchin, G., Goodall, J., Voinov, A., Hill, M., Glynn, P., Whelan, G., Geller, G., Quinn, N., Blind, M., Peckham, S., Reaney, S., Gaber, N., Kennedy, R., Hughes, A. 2013. Integrated Environmental Modeling: A Vision and Roadmap for the Future. *Environ Modell Softw* 39, 3-23.

Lahlou, M., Shoemaker, L., Choundhury, S., Elmer, R., Hu, A., Manguerra, H., Parker, A. 1998. Better assessment science integrating point and nonpoint sources—BASINS. Rep. EPA-823-B-98-006, U.S. Environmental Protection Agency, Washington, DC.

MapWindow. 2011. MapWindow 4. <<http://www.mapwindow.org/>> (last accessed 15.04.13.).

MapWindow. 2013. MapWindow 6. <<http://mapwindow6.codeplex.com/>> (last accessed 15.04.13.).

Martinez, G., Pachepsky, Y.A., Sheldon, D.R., Whelan, G., Zepp, R., Molina, M., Panhorst, K. 2013. Using the Q10 model to simulate E. coli survival in cowpats on grazing lands. *Environ Int* 54, 1-10.

McFarland, A., Adams, T. 2014. Characterizing potential bacteria loads for the Leona River watershed using the Spatially Explicit Load Enrichment Calculation Tool (SELECT). Technical Report No. 1305. Texas Institute for Applied Environmental Research. Tarleton State University, Stephenville, Texas.

McKee, K., Karthikeyan, R., Gregory, L. 2011. Modeling support for Buck Creek watershed protection plan development. Texas Water Resources Institute Technical Report. Texas A&M University System, College Station, Texas.

Meals, D.W., Braun, D.C. 2006. Demonstration of methods to reduce E. coli runoff from dairy manure application sites. *J Environ Qual* 35, 1088–1100.

Meersman R. Mark, L. (eds). 1997. Database Application Semantics. Chapman and Hall.

Moore, J.A., Smith, J., Baker, S., Miner, J.R. 1988. Evaluating coliform concentrations in runoff from various animal waste management systems. Special Report 817. Agric. Exp. Station, Oregon State University, Corvallis USA
https://ir.library.oregonstate.edu/xmlui/bitstream/handle/1957/5558/SR%20no.%20817_ocr.pdf?sequence=1 (last accessed 07.10.16).

Moore, J.A., Smyth, J.D., Baker, E.S., Miner, J.R., Moffitt, D.C. 1989. Modeling bacteria movement in livestock manure systems. *T ASAE* 32 (3), 1049-1053.

Morsey, M.M., Goodall, J.L., Bandaragoda, C., Castronova, A.M., Greenberg, J. 2014. Metadata for Describing Water Models. International Environmental Modelling and Software Society (iEMSs), 2014 International Congress on Environmental Modelling and Software, Bold Visions for Environmental Modeling, Sixth Biennial Meeting, San Diego, CA USA.

NRC (National Research Council). 2004. Indicators for Waterborne Pathogens. National Academies Press: Washington, DC.

Oladeinde, A., Bohrmann, T., Wong, K., Purucker, S.T., Bradshaw, K., Brown, R., Snyder, B., Molina, M. 2014. Decay of fecal indicator bacterial populations and bovine-associated source tracking markers in freshly deposited cowpats. *Appl Environ Microb* 80 (1), 110-118.

Overcash, M.J., Humenik, F.J., Miner, J.R. 1983. Livestock Waste Management. Vol. 1. CRC Press, Inc.
Parajuli, P.B. 2007. SWAT bacteria sub-model evaluation and application. Doctoral Dissertation. Kansas State University.

Parmar, R., Wolfe, K., Whelan, G., Laniak, G.F., Galvin, M., Goodall, J.L. 2016. Documenting Models for Interoperability and Usability. In: Sauvage, S., Sánchez-Pérez, J-M., Rizzoli, A. (Eds.), Proceedings of the International Environmental Modelling and Software Society (iEMSs), 8th International Congress on Environmental Modelling and Software, Toulouse, France.

Pelton, M.A. 2009. Requirements, Design, and Specifications for Excel Dictionary Registration Tool (DRT). PNWD-4141, Battelle, Pacific Northwest Division, Richland, WA.

Price, R. 2004. What Is An RDF Triple?

<http://www.robertprice.co.uk/roblog/2004/10/what_is_an_rdf_triple_-shtml/> (last accessed 02.08.14).

Protégé. 2014. Protégé 4 User's Guide. <http://protegewiki.stanford.edu/wiki/Pr4_UG> (last accessed 22.03.15.)

REST (Representational State Transfer). 2015.

<http://en.wikipedia.org/wiki/Representational_state_transfer> (Last accessed 01.05.15).

Riebschleager, K.J., Karthikeyan, R., Srinivasan, R., McKee, K. 2012. Estimating potential E.coli sources in a watershed using spatially explicit modeling techniques. *J A Water Resour As* 48 (4):745–761.

Schoen, M.E., Ashbolt, N.J. 2010. Assessing pathogen risk to swimmers at non-sewage impacted recreational beaches. *Environ Sci Technol* 44, 2286-2291.

Sheth, A.P. 2001. Changing Focus on Interoperability in Information Systems: From System, Syntax, Structure To Semantics. In: M.F. Goodchild, M.J. Egenhofer, R. Fegeas, and C.A. Kottman (eds). *Interoperating Geographic Information Systems*, Kluwer Academic Publishers, 25 pp.

SOA. 2015. Service-oriented architecture <http://en.wikipedia.org/wiki/Service-oriented_architecture> (last accessed 04.02.15).

Soller, J.A., Eisenberg, J.N.S. 2008. An evaluation of parsimony for microbial risk assessment models. *Environmetrics* 19 (1), 61e78.

Soller, J.A., Schoen, M.E., Bartrand, T., Ravenscroft, J., Ashbolt, N.J. 2010. Estimated human health risks from exposure to recreational waters impacted by human and non-human sources of faecal contamination. *Water Res* 44 (16), 4674-4691.

Soller, J.A., Seto, E., Olivieri, A.W. 2008. Microbial risk assessment interface tool: User documentation. Water Environmental Research Foundation, Alexandria, VA.

Soller, J.A., Olivieri, A.W., Eisenberg, J.N.S., Sakaji, R., Danielson, R. 2004. Evaluation of microbial risk assessment techniques and applications. 00-PUM-3. Water Environmental Research Foundation, Alexandria, VA.

Soller, J., Bartrand, T., Ravenscroft, J., Molina, M., Whelan, G., Schoen, M., Ashbolt, N. 2015. Estimated human health risks from recreational exposures to stormwater runoff containing animal fecal material. *Environ Modell Softw* 72, 21-32.

SSPA. 2010. PEST. S.S. Papadopoulos & Associates, Inc. <<http://www.sspa.com/pest/>> (Last accessed 12.01.11).

Teague, A., Karthikeyan, R., Babbar-Sebens, M., Srinivasan, R., Persyn, R.A. 2009. Spatially explicit load enrichment calculation tool to identify potential *E. coli* sources in watersheds. *T ASABE* 52 (4), 1109-1120.

USDA (U.S. Department of Agriculture). 2016. Census of Agriculture: 2012 Ag Census Data. <https://www.agcensus.usda.gov/> (last accessed 07.10.16).

Uschold, M., Gruninger, M. 1996. Ontologies: Principles, methods and applications, *Knowl Eng Rev* 11 (2), 93–136.

Walker, S.E., Mostaghimi, S., Dillaha, T.A., Woeste, F.E. 1990. Modeling animal waste management practices: Impacts on bacteria levels in runoff from agricultural lands. *T ASAE* 33 (3), 807-817.

Wang, L., Mankin, K.R., Marchin, G.L. 2004. Survival of fecal bacteria in dairy cow manure. *Trans ASAE* 47 (4), 1239-1246.

Wang, W.G., Tolk, A., Wang, W.P. 2009. The levels of conceptual interoperability model: applying systems engineering principles to modeling and simulation. In: Proceedings of the Spring Simulation Multiconference. Society for Modeling & Simulation International (SCS), San Diego, CA. <<http://arxiv.org/ftp/arxiv/papers/0908/0908.20191.pdf>> (last accessed 15.03.15).

Watry, G., Ames, D.P. 2008. A practical look at MapWindow GIS. First Edition, 316 p. <<http://www.lulu.com/us/en/shop/gary-watry-and-dan-ames/a-practical-look-at-mapwindow-gis-1st-edition/paperback/product-3650292.html>> (last accessed 15.04.13).

Whelan, G., Tenney, N.A., Pelton, M.A., Coleman, A.M., Ward, D.L., Droppo, J.G., Jr., Meyer, P.D., Dorow, K.E., Taira, R.Y. 2009. Techniques to access databases and integrate data. PNNL-18244, Pacific Northwest National Laboratory, Richland, WA. <http://www.pnl.gov/main/publications/external/technical_reports/PNNL-18244.pdf> (last accessed 04.04.13).

Whelan, G., Kim, K., Pelton, M.A., Castleton, K.J., Laniak, G.F., Wolfe, K., Parmar, R., Galvin, M., Babendreier, J. 2014a. Design of a Component-based Integrated Environmental Modeling Framework. *Environ Modell Softw* 55, 1-24.

Whelan, G., Kim, K., Pelton, M.A., Soller, J.A., Castleton, K.J., Molina, M., Pachepsky, Y., Ravenscroft, J., Zepp, R. 2014b. An Integrated Environmental Modeling Framework for Performing Quantitative Microbial Risk Assessments. *Environ Modell Softw* 55, 77-91.

Whelan, G., Kim, K., Parmar, R., Wolfe, K., Galvin, M., Duda, P., Gray, M., Molina, M., Zepp, R., Pachepsky, Y., Ravenscroft, J., Prieto, L., Kitchens, B. 2014c. Using IEM to Automate a Process-based QMRA. International Environmental Modelling and Software Society (iEMSs), 2014 International Congress on Environmental Modelling and Software, Bold Visions for Environmental Modeling, Sixth Biennial Meeting, San Diego, CA USA.

Wikipedia. 2014. Software architecture. <http://en.wikipedia.org/wiki/Software_architecture> (Last accessed 17.07.14).

Wikipedia. 2015a. Single-family detached home. https://en.wikipedia.org/wiki/Single-family_detached_home (last accessed 07.10.16).

Wikipedia. 2015b. https://en.wikipedia.org/wiki/Multi-family_residential (last accessed 07.10.16).

Wolfe, K., Parmar, R., Whelan, G., Laniak, G.F., Galvin, M., Kim, K., Molina, M., Zepp, R., Duda, P., Keiser, D.D. 2016. A workflow to model microbial loadings in watersheds. In: Sauvage, S., Sánchez-Pérez, J-M., Rizzoli, A. (Eds.), Proceedings of the International Environmental Modelling and Software Society (iEMSs), 8th International Congress on Environmental Modelling and Software, Toulouse, France.

Wolfe, K.L., Parmar, R.S., Laniak, G.F., Parks, A.B., Wilson, L., Brandmeyer, J.E., Ames, D.P., Gray, M.H. 2007. Data for environmental modeling (D4EM): Background and example applications of data automation. Presented at the International Symposium on Environmental Software Systems, Prague, Czech Republic, May 22-25, 2007
<http://cfpub.epa.gov/si/si_public_record_report.cfm?dirEntryId=166789&fed_org_id=770&SIType=PR&TIMSType=&showCriteria=0&address=nerl&view=citation&personID=18011&role=Author&sortBy=pubDateYear&count=100&dateBeginPublishedPresented=> (last accessed 04.04.13).

XML Wikipedia. 2014. XML. <<http://en.wikipedia.org/wiki/XML>>(last accessed 03.08.14).

Zeckoski, R.W., Benham, B.L., Shah, S.B., Wolfe, M.L., Brannan, K.M., Al-Smadi, M., Dillaha, T.A., Mostaghimi, S., Heatwole, C.D. 2005. BSLC: A tool for bacteria source characterization for watershed management. *Appl Eng Agric* 21 (5), 879-889.

APPENDIX A

EXAMPLE MICROBIAL SOURCE MODULE XML DOCUMENT FOR INPUT PARAMETERS/VARIABLES

```
<?xml version="1.0" encoding="UTF-8" ?>
<Watershed>
  <SepticNumberPeople units="Number">2.8</SepticNumberPeople> <!--Optional-->
  <SepticFailureRate units="Fraction">0.025</SepticFailureRate> <!--Optional-->
  <SepticOvercharge units="gal/d/Number">70.0</SepticOvercharge> <!--Optional-->
  <SepticConc units="Cells/L">1000.0</SepticConc> <!--Optional-->
  <SubUrbanizedBuiltUpRate>
    <Commercial units="Cells/Acre/d">6210000.0</Commercial> <!--Optional-->
    <SingleFamilyLowDensity units="Cells/Acre/d">10300000.0</SingleFamilyLowDensity> <!--Optional-->
  >
    <SingleFamilyHighDensity units="Cells/Acre/d">16600000.0</SingleFamilyHighDensity> <!--Optional-->
  ->
    <MultiFamilyResidential units="Cells/Acre/d">23300000.0</MultiFamilyResidential> <!--Optional-->
    <Road units="Cells/Acre/d">200000.0</Road> <!--Optional-->
  </SubUrbanizedBuiltUpRate>
  <MonthID> <!--Optional-->
    <January> <!--Optional-->
      <DieOff units="1/d">0.36</DieOff>
    </January>
    <February> <!--Optional-->
      <DieOff units="1/d">0.36</DieOff>
    </February>
    <March>
      <DieOff units="1/d">0.36</DieOff>
    </March>
    <April>
      <DieOff units="1/d">0.51</DieOff>
    </April>
    <May>
      <DieOff units="1/d">0.51</DieOff>
    </May>
    <June>
      <DieOff units="1/d">0.51</DieOff>
    </June>
    <July>
      <DieOff units="1/d">0.51</DieOff>
    </July>
    <August>
      <DieOff units="1/d">0.51</DieOff>
    </August>
    <September>
```

```

    <DieOff units="1/d">0.51</DieOff>
  </September>
  <October>
    <DieOff units="1/d">0.36</DieOff>
  </October>
  <November>
    <DieOff units="1/d">0.36</DieOff>
  </November>
  <December>
    <DieOff units="1/d">0.36</DieOff>
  </December>
</MonthID>
<Subwatersheds>
  <Subwatershed>
    <ID>P1</ID>
    <!--ID has to be unique-->
    <SepticNumber units="Number of Septics">180000</SepticNumber>
    <MonthID>
      <January>
        <PointFlow units="gal/d">4000.0</PointFlow><!--Average flow from all point sources in the
subwatershed-->
        <PointMicrobeRate units="Cells/L">1.0</PointMicrobeRate><!--Average microbe count in flow
from all point sources-->
      </January>
      <February>
        <PointFlow units="gal/d">4000.0</PointFlow><!--Average flow from all point sources in the
subwatershed-->
        <PointMicrobeRate units="Cells/L">2.0</PointMicrobeRate><!--Average microbe count in flow
from all point sources-->
      </February>
      <March>
        <PointFlow units="gal/d">4000.0</PointFlow><!--Average flow from all point sources in the
subwatershed-->
        <PointMicrobeRate units="Cells/L">3.0</PointMicrobeRate><!--Average microbe count in flow
from all point sources-->
      </March>
      <April>
        <PointFlow units="gal/d">4000.0</PointFlow><!--Average flow from all point sources in the
subwatershed-->
        <PointMicrobeRate units="Cells/L">4.0</PointMicrobeRate><!--Average microbe count in flow
from all point sources-->
      </April>
      <May>
        <PointFlow units="gal/d">4000.0</PointFlow><!--Average flow from all point sources in the
subwatershed-->
        <PointMicrobeRate units="Cells/L">5</PointMicrobeRate><!--Average microbe count in flow from
all point sources-->
      </May>
    </MonthID>
  </Subwatershed>
</Subwatersheds>

```

```

    <June>
      <PointFlow units="gal/d">4000.0</PointFlow><!--Average flow from all point sources in the
subwatershed-->
      <PointMicrobeRate units="Cells/L">6</PointMicrobeRate><!--Average microbe count in flow from
all point sources-->
    </June>
    <July>
      <PointFlow units="gal/d">4000.0</PointFlow><!--Average flow from all point sources in the
subwatershed-->
      <PointMicrobeRate units="Cells/L">7</PointMicrobeRate><!--Average microbe count in flow from
all point sources-->
    </July>
    <August>
      <PointFlow units="gal/d">4000.0</PointFlow><!--Average flow from all point sources in the
subwatershed-->
      <PointMicrobeRate units="Cells/L">8</PointMicrobeRate><!--Average microbe count in flow from
all point sources-->
    </August>
    <September>
      <PointFlow units="gal/d">4000.0</PointFlow><!--Average flow from all point sources in the
subwatershed-->
      <PointMicrobeRate units="Cells/L">9</PointMicrobeRate><!--Average microbe count in flow from
all point sources-->
    </September>
    <October>
      <PointFlow units="gal/d">4000.0</PointFlow><!--Average flow from all point sources in the
subwatershed-->
      <PointMicrobeRate units="Cells/L">10</PointMicrobeRate><!--Average microbe count in flow
from all point sources-->
    </October>
    <November>
      <PointFlow units="gal/d">4000.0</PointFlow><!--Average flow from all point sources in the
subwatershed-->
      <PointMicrobeRate units="Cells/L">11</PointMicrobeRate><!--Average microbe count in flow
from all point sources-->
    </November>
    <December>
      <PointFlow units="gal/d">4000.0</PointFlow><!--Average flow from all point sources in the
subwatershed-->
      <PointMicrobeRate units="Cells/L">12</PointMicrobeRate><!--Average microbe count in flow
from all point sources-->
    </December>
  </MonthID>
  <Landuse>
    <Forest>
      <Area units="Acre">40.8</Area>
    </Forest>
    <Cropland>

```

```

    <Area units="Acre">480.0</Area>
  </Cropland>
  <Pasture>
    <Area units="Acre">48.4</Area>
  </Pasture>
  <Urbanized>
    <Area units="Acre">403.1</Area>
    <!--The following four element values must add up to 1-->
    <CommercialAndServices>
      <AreaFraction units="Fraction">0.25</AreaFraction>
    </CommercialAndServices>
    <Residential>
      <AreaFraction units="Fraction">0.25</AreaFraction>
    </Residential>
    <MixedUrban>
      <AreaFraction units="Fraction">0.25</AreaFraction>
    </MixedUrban>
    <TransportationCommunicationUtilities>
      <AreaFraction units="Fraction">0.25</AreaFraction>
    </TransportationCommunicationUtilities>
  </Urbanized>
</Landuse>
<Agricultural>
  <BeefCattle>
    <NumberOfAnimals units="Number">180000.0</NumberOfAnimals>
  </BeefCattle>
  <Swine>
    <NumberOfAnimals units="Number">70.0</NumberOfAnimals>
  </Swine>
  <DairyCow>
    <NumberOfAnimals units="Number">0.0</NumberOfAnimals>
  </DairyCow>
  <Poultry>
    <NumberOfAnimals units="Number">700.0</NumberOfAnimals>
  </Poultry>
  <Horse>
    <NumberOfAnimals units="Number">48.0</NumberOfAnimals>
  </Horse>
  <Sheep>
    <NumberOfAnimals units="Number">90.0</NumberOfAnimals>
  </Sheep>
  <OtherAgAnimal>
    <NumberOfAnimals units="Number">74.0</NumberOfAnimals>
  </OtherAgAnimal>
</Agricultural>
</Subwatershed>
<Subwatershed>
  <ID>P2</ID>

```

```

<!--ID has to be unique-->
<SepticNumber units="Number of Septics">100.0</SepticNumber>
<MonthID>
  <January>
    <PointFlow units="gal/d">4000.0</PointFlow><!--Average flow from all point sources in the
subwatershed-->
    <PointMicrobeRate units="Cells/L">1.0</PointMicrobeRate><!--Average microbe count in flow
from all point sources-->
  </January>
  <February>
    <PointFlow units="gal/d">4000.0</PointFlow><!--Average flow from all point sources in the
subwatershed-->
    <PointMicrobeRate units="Cells/L">2.0</PointMicrobeRate><!--Average microbe count in flow
from all point sources-->
  </February>
  <March>
    <PointFlow units="gal/d">4000.0</PointFlow><!--Average flow from all point sources in the
subwatershed-->
    <PointMicrobeRate units="Cells/L">3.0</PointMicrobeRate><!--Average microbe count in flow
from all point sources-->
  </March>
  <April>
    <PointFlow units="gal/d">4000.0</PointFlow><!--Average flow from all point sources in the
subwatershed-->
    <PointMicrobeRate units="Cells/L">4.0</PointMicrobeRate><!--Average microbe count in flow
from all point sources-->
  </April>
  <May>
    <PointFlow units="gal/d">4000.0</PointFlow><!--Average flow from all point sources in the
subwatershed-->
    <PointMicrobeRate units="Cells/L">5.0</PointMicrobeRate><!--Average microbe count in flow
from all point sources-->
  </May>
  <June>
    <PointFlow units="gal/d">4000.0</PointFlow><!--Average flow from all point sources in the
subwatershed-->
    <PointMicrobeRate units="Cells/L">6.0</PointMicrobeRate><!--Average microbe count in flow
from all point sources-->
  </June>
  <July>
    <PointFlow units="gal/d">4000.0</PointFlow><!--Average flow from all point sources in the
subwatershed-->
    <PointMicrobeRate units="Cells/L">7.0</PointMicrobeRate><!--Average microbe count in flow
from all point sources-->
  </July>
  <August>
    <PointFlow units="gal/d">4000.0</PointFlow><!--Average flow from all point sources in the
subwatershed-->

```



```

    <PointMicrobeRate units="Cells/L">8.0</PointMicrobeRate><!--Average microbe count in flow
from all point sources-->
  </August>
  <September>
    <PointFlow units="gal/d">4000.0</PointFlow><!--Average flow from all point sources in the
subwatershed-->
    <PointMicrobeRate units="Cells/L">9.0</PointMicrobeRate><!--Average microbe count in flow
from all point sources-->
    </September>
    <October>
      <PointFlow units="gal/d">4000.0</PointFlow><!--Average flow from all point sources in the
subwatershed-->
      <PointMicrobeRate units="Cells/L">10.0</PointMicrobeRate><!--Average microbe count in flow
from all point sources-->
      </October>
      <November>
        <PointFlow units="gal/d">4000.0</PointFlow><!--Average flow from all point sources in the
subwatershed-->
        <PointMicrobeRate units="Cells/L">11.0</PointMicrobeRate><!--Average microbe count in flow
from all point sources-->
        </November>
        <December>
          <PointFlow units="gal/d">4000.0</PointFlow><!--Average flow from all point sources in the
subwatershed-->
          <PointMicrobeRate units="Cells/L">12.0</PointMicrobeRate><!--Average microbe count in flow
from all point sources-->
          </December>
        </MonthID>
      <Landuse>
        <Forest>
          <Area units="Acre">40.8</Area>
        </Forest>
        <Cropland>
          <Area units="Acre">480.0</Area>
        </Cropland>
        <Pasture>
          <Area units="Acre">48.4</Area>
        </Pasture>
        <Urbanized>
          <Area units="Acre">403.1</Area>
          <!--The following four element values must add up to 1-->
          <CommercialAndServices>
            <AreaFraction units="Fraction">0.25</AreaFraction>
          </CommercialAndServices>
          <Residential>
            <AreaFraction units="Fraction">0.25</AreaFraction>
          </Residential>
          <MixedUrban>

```

```

    <AreaFraction units="Fraction">0.25</AreaFraction>
  </MixedUrban>
  <TransportationCommunicationUtilities>
    <AreaFraction units="Fraction">0.25</AreaFraction>
  </TransportationCommunicationUtilities>
</Urbanized>
</Landuse>
<Agricultural>
  <BeefCattle>
    <NumberOfAnimals units="Number">0.0</NumberOfAnimals>
  </BeefCattle>
  <Swine>
    <NumberOfAnimals units="Number">70.0</NumberOfAnimals>
  </Swine>
  <DairyCow>
    <NumberOfAnimals units="Number">0.0</NumberOfAnimals>
  </DairyCow>
  <Poultry>
    <NumberOfAnimals units="Number">700.0</NumberOfAnimals>
  </Poultry>
  <Horse>
    <NumberOfAnimals units="Number">48.0</NumberOfAnimals>
  </Horse>
  <Sheep>
    <NumberOfAnimals units="Number">90.0</NumberOfAnimals>
  </Sheep>
  <OtherAgAnimal>
    <NumberOfAnimals units="Number">74.0</NumberOfAnimals>
  </OtherAgAnimal>
</Agricultural>
</Subwatershed>
</Subwatersheds>
<Agricultural>
  <!--Number of grazing days in a month can be calculated by multiplying the fraction of time animals
  spend grazing by number of days in the month -->
  <BeefCattle>
    <ManureIncorporatedIntoSoil Units="Fraction">0.75</ManureIncorporatedIntoSoil>
    <MicrobialAnimalProductionRates
Units="Cells/d">104000000000.0</MicrobialAnimalProductionRates>
  <MonthID>
  <!-- Sum of monthly applications (i.e., Application) over the year must add to 1.0-->
  <January>
    <Application units="Fraction">0.0375</Application>
    <GrazingDays units="Number">0.0</GrazingDays>
    <TimeSpentInStreams units="Fraction">0</TimeSpentInStreams>
  </January>
  <February>
    <Application units="Fraction">0.0375</Application>

```

```

    <GrazingDays units="Number">0.0</GrazingDays>
    <TimeSpentInStreams units="Fraction">0.0</TimeSpentInStreams>
</February>
<March>
    <Application units="Fraction">0.0375</Application>
    <GrazingDays units="Number">0.0</GrazingDays>
    <TimeSpentInStreams units="Fraction">0.0</TimeSpentInStreams>
</March>
<April>
    <Application units="Fraction">0.05</Application>
    <GrazingDays units="Number">20.0</GrazingDays>
    <TimeSpentInStreams units="Fraction">0.1</TimeSpentInStreams>
</April>
<May>
    <Application units="Fraction">0.05</Application>
    <GrazingDays units="Number">31.0</GrazingDays>
    <TimeSpentInStreams units="Fraction">0.15</TimeSpentInStreams>
</May>
<June>
    <Application units="Fraction">0.0375</Application>
    <GrazingDays units="Number">30.0</GrazingDays>
    <TimeSpentInStreams units="Fraction">.10</TimeSpentInStreams>
</June>
<July>
    <Application units="Fraction">0.0375</Application>
    <GrazingDays units="Number">31.0</GrazingDays>
    <TimeSpentInStreams units="Fraction">0.1</TimeSpentInStreams>
</July>
<August>
    <Application units="Fraction">0.0375</Application>
    <GrazingDays units="Number">31.0</GrazingDays>
    <TimeSpentInStreams units="Fraction">0.1</TimeSpentInStreams>
</August>
<September>
    <Application units="Fraction">0.0375</Application>
    <GrazingDays units="Number">30.0</GrazingDays>
    <TimeSpentInStreams units="Fraction">0.1</TimeSpentInStreams>
</September>
<October>
    <Application units="Fraction">0.3</Application>
    <GrazingDays units="Number">31.0</GrazingDays>
    <TimeSpentInStreams units="Fraction">0.1</TimeSpentInStreams>
</October>
<November>
    <Application units="Fraction">0.3</Application>
    <GrazingDays units="Number">15.0</GrazingDays>
    <TimeSpentInStreams units="Fraction">0.05</TimeSpentInStreams>
</November>

```

```

<December>
  <Application units="Fraction=">0.0375</Application>
  <GrazingDays units="Number">0.0</GrazingDays>
  <TimeSpentInStreams units="Fraction">0.0</TimeSpentInStreams>
</December>
</MonthID>
</BeefCattle>
<DairyCow>
  <ManureIncorporatedIntoSoil Units="Fraction">0.75</ManureIncorporatedIntoSoil>
  <MicrobialAnimalProductionRates
Units="Cells/d">104000000000.0</MicrobialAnimalProductionRates>
<MonthID>
  <!-- Sum of monthly applications (i.e., Application) over the year must add to 1.0-->
  <January>
    <Application units="Fraction=">0.0375</Application>
  </January>
  <February>
    <Application units="Fraction=">0.0375</Application>
  </February>
  <March>
    <Application units="Fraction=">0.0375</Application>
  </March>
  <April>
    <Application units="Fraction=">0.05</Application>
  </April>
  <May>
    <Application units="Fraction=">0.05</Application>
  </May>
  <June>
    <Application units="Fraction=">0.0375</Application>
  </June>
  <July>
    <Application units="Fraction=">0.0375</Application>
  </July>
  <August>
    <Application units="Fraction=">0.0375</Application>
  </August>
  <September>
    <Application units="Fraction=">0.0375</Application>
  </September>
  <October>
    <Application units="Fraction=">0.3</Application>
  </October>
  <November>
    <Application units="Fraction=">0.3</Application>
  </November>
  <December>
    <Application units="Fraction=">0.0375</Application>

```

```

    </December>
  </MonthID>
</DairyCow>
<Horse>
  <ManureIncorporatedIntoSoil Units="Fraction">0.75</ManureIncorporatedIntoSoil>
  <MicrobialAnimalProductionRates Units="Cells/d">420000000.0</MicrobialAnimalProductionRates>
</MonthID>
  <January>
    <GrazingDays units="Number">31.0</GrazingDays>
    <Application units="Fraction">0.0</Application>
  </January>
  <February>
    <GrazingDays units="Number">28.0</GrazingDays>
    <Application units="Fraction">0.0</Application>
  </February>
  <March>
    <GrazingDays units="Number">31.0</GrazingDays>
    <Application units="Fraction">0.0</Application>
  </March>
  <April>
    <GrazingDays units="Number">30.0</GrazingDays>
    <Application units="Fraction">0.1</Application>
  </April>
  <May>
    <GrazingDays units="Number">31.0</GrazingDays>
    <Application units="Fraction">0.1</Application>
  </May>
  <June>
    <GrazingDays units="Number">30.0</GrazingDays>
    <Application units="Fraction">0.0</Application>
  </June>
  <July>
    <GrazingDays units="Number">31.0</GrazingDays>
    <Application units="Fraction">0.0</Application>
  </July>
  <August>
    <GrazingDays units="Number">31.0</GrazingDays>
    <Application units="Fraction">0.0</Application>
  </August>
  <September>
    <GrazingDays units="Number">30.0</GrazingDays>
    <Application units="Fraction">0.0</Application>
  </September>
  <October>
    <GrazingDays units="Number">31.0</GrazingDays>
    <Application units="Fraction">0.4</Application>
  </October>
</November>

```

```

    <GrazingDays units="Number">30.0</GrazingDays>
    <Application units="Fraction">0.4</Application>
  </November>
  <December>
    <GrazingDays units="Number">31.0</GrazingDays>
    <Application units="Fraction">0.0</Application>
  </December>
</MonthID>
</Horse>
<Sheep>
  <MicrobialAnimalProductionRates
Units="Cells/d">12000000000.0</MicrobialAnimalProductionRates>
  <MonthID>
    <January>
      <GrazingDays units="Number">0.0</GrazingDays>
    </January>
    <February>
      <GrazingDays units="Number">0.0</GrazingDays>
    </February>
    <March>
      <GrazingDays units="Number">0.0</GrazingDays>
    </March>
    <April>
      <GrazingDays units="Number">0.0</GrazingDays>
    </April>
    <May>
      <GrazingDays units="Number">0.0</GrazingDays>
    </May>
    <June>
      <GrazingDays units="Number">30.0</GrazingDays>
    </June>
    <July>
      <GrazingDays units="Number">31.0</GrazingDays>
    </July>
    <August>
      <GrazingDays units="Number">31.0</GrazingDays>
    </August>
    <September>
      <GrazingDays units="Number">51.4</GrazingDays>
    </September>
    <October>
      <GrazingDays units="Number">0.0</GrazingDays>
    </October>
    <November>
      <GrazingDays units="Number">0.0</GrazingDays>
    </November>
    <December>
      <GrazingDays units="Number">0.9</GrazingDays>

```

```

    </December>
  </MonthID>
</Sheep>
<Poultry>
  <ManureIncorporatedIntoSoil Units="Fraction">0.96</ManureIncorporatedIntoSoil>
  <MicrobialAnimalProductionRates Units="Cells/d">136000000.0</MicrobialAnimalProductionRates>
  <MonthID>
    <!-- Sum of monthly applications must add to 1.0-->
    <January>
      <Application units="Fraction=">0.0</Application>
    </January>
    <February>
      <Application units="Fraction=">0.0</Application>
    </February>
    <March>
      <Application units="Fraction=">0.0</Application>
    </March>
    <April>
      <Application units="Fraction=">0.10</Application>
    </April>
    <May>
      <Application units="Fraction=">0.10</Application>
    </May>
    <June>
      <Application units="Fraction=">0.0</Application>
    </June>
    <July>
      <Application units="Fraction=">0.0</Application>
    </July>
    <August>
      <Application units="Fraction=">0.0</Application>
    </August>
    <September>
      <Application units="Fraction=">0.0</Application>
    </September>
    <October>
      <Application units="Fraction=">0.40</Application>
    </October>
    <November>
      <Application units="Fraction=">0.40</Application>
    </November>
    <December>
      <Application units="Fraction=">0.0</Application>
    </December>
  </MonthID>
</Poultry>
<Swine>
  <ManureIncorporatedIntoSoil Units="Fraction">0.80</ManureIncorporatedIntoSoil>

```

```

    <MicrobialAnimalProductionRates
Units="Cells/d">10800000000.0</MicrobialAnimalProductionRates>
    <MonthID>
    <!-- Sum of monthly applications must add to 1.0-->
    <January>
        <Application units="Fraction=">0.0</Application>
    </January>
    <February>
        <Application units="Fraction=">0.0</Application>
    </February>
    <March>
        <Application units="Fraction=">0.0</Application>
    </March>
    <April>
        <Application units="Fraction=">0.10</Application>
    </April>
    <May>
        <Application units="Fraction=">0.10</Application>
    </May>
    <June>
        <Application units="Fraction=">0.0</Application>
    </June>
    <July>
        <Application units="Fraction=">0.0</Application>
    </July>
    <August>
        <Application units="Fraction=">0.0</Application>
    </August>
    <September>
        <Application units="Fraction=">0.0</Application>
    </September>
    <October>
        <Application units="Fraction=">0.40</Application>
    </October>
    <November>
        <Application units="Fraction=">0.40</Application>
    </November>
    <December>
        <Application units="Fraction=">0.0</Application>
    </December>
    </MonthID>
</Swine>
<OtherAgAnimal>
    <MicrobialAnimalProductionRates
Units="Cells/d">104000000000.0</MicrobialAnimalProductionRates>
    <MonthID>
    <January>
        <GrazingDays units="Number">0.0</GrazingDays>

```



```

</January>
<February>
  <GrazingDays units="Number">0.0</GrazingDays>
</February>
<March>
  <GrazingDays units="Number">0.0</GrazingDays>
</March>
<April>
  <GrazingDays units="Number">0.0</GrazingDays>
</April>
<May>
  <GrazingDays units="Number">0.0</GrazingDays>
</May>
<June>
  <GrazingDays units="Number">30.0</GrazingDays>
</June>
<July>
  <GrazingDays units="Number">31.0</GrazingDays>
</July>
<August>
  <GrazingDays units="Number">31.0</GrazingDays>
</August>
<September>
  <GrazingDays units="Number">15.4</GrazingDays>
</September>
<October>
  <GrazingDays units="Number">0.0</GrazingDays>
</October>
<November>
  <GrazingDays units="Number">0.0</GrazingDays>
</November>
<December>
  <GrazingDays units="Number">0.0</GrazingDays>
</December>
</MonthID>
</OtherAgAnimal>
</Agricultural>
<Wildlife>
  <Duck>
    <MicrobialWildlifeProductionRates
Units="Cells/d">2430000000.0</MicrobialWildlifeProductionRates>
  <Landuse>
    <Forest>
      <Density units="Number/Acre">0.1</Density>
    </Forest>
    <Cropland>
      <Density units="Number/Acre">1.4</Density>
    </Cropland>
  </Duck>
</Wildlife>

```

```

    <Pasture>
      <Density units="Number/Acre">0.4</Density>
    </Pasture>
    <Urbanized>
      <Density units="Number/Acre">0.1</Density>
    </Urbanized>
  </Landuse>
</Duck>
<Goose>
  <MicrobialWildlifeProductionRates
Units="Cells/d">49000000000.0</MicrobialWildlifeProductionRates>
  <Landuse>
    <Forest>
      <Density units="Number/Acre">0.1</Density>
    </Forest>
    <Cropland>
      <Density units="Number/Acre">0.1</Density>
    </Cropland>
    <Pasture>
      <Density units="Number/Acre">0.1</Density>
    </Pasture>
    <Urbanized>
      <Density units="Number/Acre">0.1</Density>
    </Urbanized>
  </Landuse>
</Goose>
<Deer>
  <MicrobialWildlifeProductionRates
Units="Cells/d">49000000000.0</MicrobialWildlifeProductionRates>
  <Landuse>
    <Forest>
      <Density units="Number/Acre">0.1</Density>
    </Forest>
    <Cropland>
      <Density units="Number/Acre">0.05</Density>
    </Cropland>
    <Pasture>
      <Density units="Number/Acre">0.1</Density>
    </Pasture>
    <Urbanized>
      <Density units="Number/Acre">0.1</Density>
    </Urbanized>
  </Landuse>
</Deer>
<Beaver>
  <MicrobialWildlifeProductionRates
Units="Cells/d">49000000000.0</MicrobialWildlifeProductionRates>
  <Landuse>

```

```

<Forest>
  <Density units="Number/Acre">0.1</Density>
</Forest>
<Cropland>
  <Density units="Number/Acre">0.1</Density>
</Cropland>
<Pasture>
  <Density units="Number/Acre">0.05</Density>
</Pasture>
<Urbanized>
  <Density units="Number/Acre">0.05</Density>
</Urbanized>
</Landuse>
</Beaver>
<Raccoon>
  <MicrobialWildlifeProductionRates
Units="Cells/d">49000000000.0</MicrobialWildlifeProductionRates>
  <Landuse>
    <Forest>
      <Density units="Number/Acre">0.1</Density>
    </Forest>
    <Cropland>
      <Density units="Number/Acre">3.0</Density>
    </Cropland>
    <Pasture>
      <Density units="Number/Acre">0.1</Density>
    </Pasture>
    <Urbanized>
      <Density units="Number/Acre">0.05</Density>
    </Urbanized>
  </Landuse>
</Raccoon>
<OtherWildlife>
  <MicrobialWildlifeProductionRates
Units="Cells/d">49000000000.0</MicrobialWildlifeProductionRates>
  <Landuse>
    <Forest>
      <Density units="Number/Acre">0.1</Density>
    </Forest>
    <Cropland>
      <Density units="Number/Acre">0.05</Density>
    </Cropland>
    <Pasture>
      <Density units="Number/Acre">0.01</Density>
    </Pasture>
    <Urbanized>
      <Density units="Number/Acre">0.08</Density>
    </Urbanized>
  </Landuse>

```

</Landuse>
</OtherWildlife>
</Wildlife>
</Watershed>