Jun 16th, 9:00 AM - 10:20 AM

Using integrated environmental modeling to automate a process-based Quantitative Microbial Risk Assessment

Gene Whelan  
_U.S. Environmental Protection Agency_, Whelan.Gene@epa.gov

Keewook Kim  
_U.S. Environmental Protection Agency, U.S. Department of Energy, Oak Ridge Institute for Science and Education_,  
kim.keewook@epa.gov

Rajbir Parmar  
_U.S. Environmental Protection Agency_, parmar.rajbir@epa.gov

Kurt Wolfe  
_U.S. Environmental Protection Agency_, wolfe.kurt@epa.gov

Mike Galvin  
_U.S. Environmental Protection Agency_, galvin.mike@epa.gov

Follow this and additional works at: [https://scholarsarchive.byu.edu/iemssconference](https://scholarsarchive.byu.edu/iemssconference)

Part of the [Civil Engineering Commons](https://scholarsarchive.byu.edu/iemssconference), [Data Storage Systems Commons](https://scholarsarchive.byu.edu/iemssconference), [Environmental Engineering Commons](https://scholarsarchive.byu.edu/iemssconference), [Hydraulic Engineering Commons](https://scholarsarchive.byu.edu/iemssconference), and the [Other Civil and Environmental Engineering Commons](https://scholarsarchive.byu.edu/iemssconference)

Whelan, Gene; Kim, Keewook; Parmar, Rajbir; Wolfe, Kurt; Galvin, Mike; Duda, Paul; Gray, Mark; Molina, Marirosa; Zepp, Richard; Pachepsky, Yakov; Ravenscroft, John; Prieto, Lourdes; and Kitchens, Brenda, "Using integrated environmental modeling to automate a process-based Quantitative Microbial Risk Assessment" (2014). _International Congress on Environmental Modelling and Software_. 57.  
[https://scholarsarchive.byu.edu/iemssconference/2014/Stream-H/57](https://scholarsarchive.byu.edu/iemssconference/2014/Stream-H/57)
Presenter/Author Information
Gene Whelan, Keewook Kim, Rajbir Parmar, Kurt Wolfe, Mike Galvin, Paul Duda, Mark Gray, Marirosa Molina, Richard Zepp, Yakov Pachepsky, John Ravenscroft, Lourdes Prieto, and Brenda Kitchens

This event is available at BYU ScholarsArchive: https://scholarsarchive.byu.edu/iemssconference/2014/Stream-H/57
Using integrated environmental modeling to automate a process-based Quantitative Microbial Risk Assessment


*U.S. Environmental Protection Agency, Athens, GA USA (Whelan.Gene@epa.gov, Kim.Keewook@epa.gov, Parmar.Rajbir@epa.gov, Wolfe.Kurt@epa.gov, Galvin.Mike@epa.gov, Molina.Marirosa@epa.gov, Zepp.Richard@epa.gov, Prieto.Lourdes@epa.gov, Kitchens.Brenda@epa.gov)

bU.S. Department of Energy, Oak Ridge Institute for Science and Education, Oak Ridge, TN USA

cAqua Terra Consultants, Decatur, GA USA (PDMDuda@aquaterra.com, markgray@aquaterra.com)

dU.S. Department of Agriculture, Agricultural Research Service, Beltsville, MD USA (Yakov.Pachepsky@ars.usda.gov)

*e.U.S. Environmental Protection Agency, Office of Water, Washington, DC USA (Ravenscroft.John@epa.gov)

Abstract: Integrated Environmental Modeling (IEM) organizes multidisciplinary knowledge that explains and predicts environmental-system response to stressors. A Quantitative Microbial Risk Assessment (QMRA) is an approach integrating a range of disparate data (fate/transport, exposure, and human health effects relationships) to characterize potential health impacts/risks from exposure to pathogenic microorganisms. We demonstrate loosely connected IEM legacy technologies (SDMProjectBuilder, Microbial Source Module, HSPF, and BASINS) to support watershed-scale microbial source-to-receptor modeling, focusing on animal-impacted catchments. The coupled models automate manual steps in standard watershed assessments to expedite the process, minimize resources, increase ease of use, and introduce more science-based processes to the analysis. SDMProjectBuilder accesses, retrieves, analyzes, and caches web-based data. The Microbial Source Module provides estimates of microbial loading rates within a watershed; HSPF simulates flow and microbial fate/transport within a watershed; and BASINS provides a user interface to access/modify HSPF files and provide visualization tools. The assessment performs HUC-12 or pour point analyses; automates watershed delineation and data-collection; pre-populates HSPF input requirements, accounting for snow accumulation/melt, microbial fate/transport, and different time increments (hourly, daily, etc.); assigns NLDAS radar meteorological data automatically to individual HUC-12s when observed data are scarce, incorrect, or insufficient; and processes manure-based source terms to estimate manure/microbial loads on subwatersheds automatically, based on number of animals, septic systems, etc. that correlate to land-use patterns.

Keywords: Integrated Environmental Modeling; BASINS; Microorganisms; Pathogens; Watershed modeling; D4EM; HSPF; QMRA

1. INTRODUCTION

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 among 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 have been 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 the components.

A Quantitative Microbial Risk Assessment (QMRA) is a modeling approach that integrates disparate data including those linked to 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 1), 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. By coupling modeling results with epidemiology studies, policy-related issues (e.g., EPA, 2010; EPA and USDA, 2012) can be explored (Figure 1).

Experience demonstrates that a substantial amount of time is required to set up the CSM, collect data that meet model input requirements, and populate model input files prior to any model execution, leaving less time and resources for actual analysis. Because QMRAs can be considered a microbial version of the existing chemical risk paradigm (Whelan et al., 2014b; EPA 2012a, 2012b), manual procedures of collecting data from standard databases (e.g., NHDPlus, NLCD, NLDAS, etc.) and populating model input files tend to be repeated, lending themselves to automation. This is consistent with Laniak et al. (2013) and Peckham et al. (2013) who noted a near-term need is to automate data-

---

**Figure 1.** 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)
gathering, meeting input model data requirements, and model coupling processes as much as possible (Whelan et al., 2014a). A watershed-scale microbial source-to-receptor workflow is constructed that follows the release of microbes from animal-impacted catchments through its waterbody network to receptors of concern. The objective here is to automate the manual steps associated with this standard QMRA-based watershed assessment, expedite the process, minimize resource requirements, increase ease of use, and bring more science-based processes into the analysis.

2. METHODS AND MATERIALS

Four models and a set of databases are used to form a loosely configured modeling infrastructure that supports watershed-scale microbial source-to-receptor modeling by focusing on animal-impacted catchments: Simulation Data Manager Project Builder (SDMProjectBuilder) with the Data for Environmental Modeling (D4EM), Microbial Source Module (MSM), Hydrologic Simulation Program – FORTRAN (HSPF), and the Better Assessment Science Integrating point & Non-point Sources system (BASINS).

D4EM (Wolfe et al., 2007; AquaTerra, 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, 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; 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 sites, or used with a custom-built user interface such as the SDMProjectBuilder.

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 geographic information system (GIS) capabilities using DotSpatial technology; converts DotSpatial-based project files to Mapwindow-based project files (Mapwindow 2011, 2013; Watry and Ames, 2008); and automatically pre-populates input files of fate and transport models. It automates the watershed delineation process, allowing for HUC-8, HUC-12, or pour point analyses; assigns map-layer features (e.g., slope, soil, land use, microbial sources, NLDAS (Kim et al., 2014) radar meteorological stations) automatically; and accounts automatically for snow accumulation/melt, microbial fate and transport, and user-defined simulation time increments (e.g., hourly, daily, monthly, annually). The ability to define time increments is valuable, especially when hourly storm events drive an assessment.

The Microbial Source Module (MSM) organizes, analyzes, and supplies the necessary data to determine microbial loading rates within a watershed to support HSPF input data requirements, and is based on the HSPF Bacterial Indicator Tool (EPA, 2013b, 2013c). The MSM works with the SDMProjectBuilder to read a pre-populated file with microbial source locations, animal numbers and types [agricultural animals (cattle, swine, poultry, etc.), wildlife (ducks, deer, etc.), manure application schedules, grazing patterns, and point source releases (e.g., septic systems and treatment facilities), where applicable] by latitude-longitude and correlate them with land-use type (i.e., built-up/impervious, pastureland, cropland, forest) and subwatersheds delineated by the SDMProjectBuilder. The MSM develops microbial loadings, adjusted for die-off, to the overland (e.g., counts/ha/d) by land use and month and instream (counts/hr) locations within a watershed, and automatically pre-populates the HSPF User Control Input (UCI) file with microbial loading rates. Time scales for instream, point-source loadings can vary.

HSPF (Bicknell et al., 1997) is a comprehensive package for simulating watershed hydrology and water quality for conventional (e.g., sediment, nutrients) and nonconventional (e.g., toxic organic) pollutants and microbes, with microbes being the focus here; implements a basin-scale analysis that includes hydrology and water quality to allow the integrated simulation of land and soil contaminant runoff processes with instream hydraulic and microbial interactions on a variety of user-defined time
scales (hour, day, month, year); and provides a time history of runoff flow rates and microbial concentrations at any point in a watershed (EPA, 2013d)

BASINS (EPA, 2001; Lahlou et al., 1998) is a multipurpose environmental analysis infrastructure designed to perform 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, displaying maps, tables, or graphics; analyzes landscape information; and integrates and displays relationships among data at a scale chosen by the user (EPA, 2013e).

3. RESULTS AND DISCUSSION

A QMRA infrastructure that automates the manual process of characterizing transport of pathogens and fecal indicator bacteria, from the source of release to a point of exposure, has been developed by loosely configuring a set of modules and process-based models. A design schematic of the workflow, tracking data from their sources to the visualization of simulation results is presented in Figure 2, while Figure 2a shows example national databases accessed by the SDMProjectBuilder using D4EM. The user initiates the process with the SDMProjectBuilder by navigating the United States by state, county, and HUC-8 (Figure 3a), delineating the watershed into HUC-12 components (Figure 3a), and choosing a single HUC-12 or pour point (EPA, 2013f) analysis (Figures 3b and 3c, respectively). The pour point in Figure 3c happens to coincide with a gaging station, which is convenient for simulation comparisons and automated model calibration. The watershed is divided into subwatersheds, with each being assigned its own unique characteristics, if desired, by allowing the user to identify minimum subwatershed sizes, land use areas, and flow lengths, since most models are designed for applications at particular scales. With user-defined simulation output time intervals (hourly, daily, etc.), simulation start and end times, and choices of data sources (Figure 2a), the SDMProjectBuilder manages data acquisition from standard national databases using D4EM, which is based on the BASINS watershed modeling system data-download tool. The result is a delineated watershed containing a suite of GIS map layers that includes NHDPlus flowlines, waterbodies, and catchments (subwatersheds); waterbody network; elevation (e.g., slope); soil types; land use types; MET stations; and multiple legal boundaries (state, county, roads, eco/NAWQA regions, etc.). Many map layers are similar to those found in BASINS.

Figure 2. An automated process-based QMRA: a) Design schematic and (b) Workflow
Working with SDMProjectBuilder, MSM consumes potential microbial source locations/types and overlays them on subwatersheds. Coupled with other acquired data, the SDMProjectBuilder automatically creates input files for various models (in this case, HSPF, as shown in Figure 2a), a collection of geospatial data files, and a Mapwindow project (.mwprj) file (Figure 2b) currently used by BASINS. Using the HSPF Windows interface, non-spatially related data may be modified, and the model is executed (Figure 2b), creating a workflow of overland runoff subwatersheds and stream reaches that results in flows and microbial concentrations. BASINS is used to view the flow and concentration time series at any location within the watershed. If additional data are desired – such as gaging station data for comparison – BASINS can be modified to register these datasets by updating the mwprj file (Figure 2b), where data can be viewed alongside simulation results.

We recognize that automation can give birth to “integronsters” (Voinov and Shugart 2013) – constructs that are perfectly valid as software products, but are ugly or even useless as modeling paradigms, similar to concepts surrounding misuse of individual models. Integronsters can be confederations of models that are inappropriately linked, or modeling constructs that misrepresent the problem (Whelan et al. 2014a). To avoid this undesirable result, we are in the process of automating a calibration procedure between observed monitoring data and model simulations, using inverse modeling, to ensure that essential aspects are initially tethered to reality, recognizing that any calibration process is ultimately the responsibility of the analyst and must be interrogated and approved by the analyst prior to acceptance.

![Figure 3](image_url)

**Figure 3.** Schematic illustrating an automated watershed delineation by (a) HUC-8, (b) HUC-12, and (c) Pour Point with gaging station location, near Manitowoc, WI, USA, using the SDMProjectBuilder

4. **SUMMARY**
A loosely coupled software system was developed that connects IEM legacy technologies (SDMProjectBuilder/D4EM, MSM, HSPF, and BASINS) to support a watershed-scale QMRA source-to-receptor assessment, focusing on animal-impacted catchments, although point sources are also considered. SDMProjectBuilder/D4EM accesses, retrieves, analyzes, and caches web-based data, and automatically pre-populates input needs of the fate and transport model. The Microbial Source Module determines microbial loading rates within a watershed; HSPF simulates flow and microbial fate/transport within a watershed; and BASINS provides a user interface to access/modify HSPF files and provide visualization tools. This IEM system (1) focuses on watershed-based assessments; (2) allows for HUC-8, HUC-12, or pour point analyses; (3) automates the watershed delineation process; (4) automates data collection; (5) processes manure-based source terms automatically to estimate microbial loads throughout the watershed, based on point sources and the number and type of animals, land-use patterns, and timing of loadings; (6) automates pre-population of HSPF input-data requirements, accounting for snow accumulation/melt, microbial fate and transport, and a user-defined simulation time increments (e.g., hourly, daily, monthly, annually); (7) uses modeling tools and systems that are familiar to HSPF/BASIN users; (8) allows users to inspect and interrogate water flows and microbial concentration time series at multiple, user-defined locations throughout the watershed; (9) allows batch simulations; (10) allows manual manipulation of input data for more refined and boutique assessments; (11) allows software components to operate in tandem or as stand-alone products; and (12) operates on a personal computer. The objective of coupling these models and software systems is automation of manual steps in a standard watershed assessment to expedite the process, minimize resources, increase ease of use, and introduce more science-based processes to the analysis.

Future efforts include linking many of these components within an IEM infrastructure, such as FRAMES (Whelan et al., 2014a). This would provide access to additional functionalities, like sensitivity/uncertainty tools [e.g., PEST (Doherty, 2004), SuperMUSE (Babendreier and Castleton, 2005)] to help capture parameter variability, risk models [e.g., MRA-IT (Soller et al., 2008, 2004)], other QMRA modeling paradigms [e.g., CAFO model (Pelton et al., 2010)], and microbial databases with the ability to contain information on physico-microbial properties [e.g., half-lives (Bolster et al., 2011); manure release rates (e.g., Guber et al., 2013; Bradford and Schijven, 2002); prevalence of infection; density of pathogens in feces of infected animals; and related information (e.g., animal fecal excretion rates, fraction of failed septic systems, etc.)].

5. ACKNOWLEDGMENTS

This research was supported in part by an appointment to the Research Participation Program at the EPA, Office of Research and Development, administered by the Oak Ridge Institute for Science and Education through Interagency Agreement No. DW8992298301 between the United States Department of Energy and the EPA. This paper has been reviewed in accordance with EPA's peer and administrative review policies and approved for presentation and publication.

6. REFERENCES


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), 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.


Pelton, M.A., Castleton, K.J., Clopton, S.J., 2010. Application requirements, design, and specifications for the integration of CAFO in FRAMES. PNWD-4240, Battelle, Pacific Northwest Division, Richland, WA.


