GWERD Quality Assurance Project Plan for

Surface Water Transport of Hydraulic Fracturing-Derived Waste Water

Task No.: 26278

QA ID No. G-16495

QA Category: 1

HF Project #5a

Date Original QAPP submitted: 7/26/2011

Number of Pages: 42

Revision No: #2 submitted 01/22/2013 (see p. 39 for Revision History)

/S/ 3/19/2013

James W. Weaver, Principal Investigator — Date

APPROVALS:

/S/ 3/19/2013

David Jewett, Branch Chief — Date

/S/ 3/19/2013

Stephen Kraemer, HF Modeling Technical Research Lead — Date

/S/ 3/19/2013

Steve Vandegrift, GWERD/HF Project QA Manager — Date

Disclaimer

EPA does not consider this internal planning document an official Agency dissemination of information under the Agency's Information Quality Guidelines, because it is not being used to formulate or support a regulation or guidance; or to represent a final Agency decision or position. This planning document describes the overall quality assurance approach that will be used during the research study. Mention of trade names or commercial products in this planning document does not constitute endorsement or recommendation for use.

The EPA Quality System and the HF Research Study

EPA requires that all data collected for the characterization of environmental processes and conditions are of the appropriate type and quality for their intended use. This is accomplished through an Agency-wide quality system for environmental data. Components of the EPA quality system can be found at http://www.epa.gov/quality/. EPA policy is based on the national consensus standard ANSI/ASQ E4-2004 *Quality Systems for Environmental Data and Technology Programs: Requirements with Guidance for Use*. This standard recommends a tiered approach that includes the development and use of Quality Management Plans (QMPs). The organizational units in EPA that generate and/or use environmental data are required to have Agency-approved QMPs. Programmatic QMPs are also written when program managers and their QA staff decide a program is of sufficient complexity to benefit from a QMP, as was done for the study of the potential impacts of hydraulic fracturing (HF) on drinking water resources. The HF QMP describes the program's organizational structure, defines and assigns quality assurance (QA) and quality control (QC) responsibilities, and describes the processes and procedures used to plan, implement and assess the effectiveness of the quality system. The HF QMP is then supported by project-specific QA project plans (QAPPs). The QAPPs provide the technical details and associated QA/QC procedures for the research projects that address questions posed by EPA about the HF water cycle and as described in the *Plan to Study the Potential Impacts of Hydraulic Fracturing on Drinking Water Resources* (EPA/600/R-11/122/November 2011/www.epa.gov/hydraulic fracturing). The results of the research projects will provide the foundation for EPA's 2014 study report.

This QAPP provides information concerning the Wastewater Treatment and Waste Disposal Stage of the HF water cycle as found in Figure 1 of the HF QMP and as described in the HF Study Plan. Appendix A of the HF QMP includes the links between the HF Study Plan questions and those QAPPs available at the time the HF QMP was published.

A2 Contents

A3 Distribution List....4
A4 Project Organization....5
Organization Chart....6
A5 Problem Definition and Background....7
A6 Project/Task Description....8
Tracer-based Empirical Transport Estimation....10
A7 Quality Objectives and Criteria....18
Data with Known Quality....18
Data Analysis Methods....18
Documenting Code Development....19
Simulation Results....19
Model Uncertainty....19
Document Retention....21
A8 Special Training/Certification....21
A9 Documents and Records....21
Contractor (Shaw Environmental) Directory Structure....25
B1-B6, B8 Sampling and Measurement Requirements....26
B7 Sampling and Measurement Requirements....26
B10 Data Management and Hardware/Software Configuration....29
C1 Assessments and Response Actions....31
C2 Reports to Management....34
D1 Data Review, Verification, and Validation....35
D2 Verification and Validation Methods....36
D3 Reconciliation with User Requirements....36
References....37
REVISION HISTORY:....39
Certification....42

A3 Distribution List

Jim Weaver, NRMRL, GWERD

Susan Mravik, NRMRL, GWERD

David Jewett, NRMRL, GWERD

Steve Vandegrift, NRMRL, GWERD

Stephen Kraemer, NERL, ERD

Contract/Coop project participants:

- Victor Murray, Shaw Environmental
- Jonathan Shireman, Shaw Environmental
- Kay Pinley, SEEP program

A4 Project Organization

Stephen Kraemer, Research Hydrologist, National Exposure Research Laboratory (NERL), Ecosystems Research Division (ERD), Regulatory Support Branch (RSB), Athens, GA, and ***Modeling Technical Research Lead***, ORD Hydraulic Fracturing study. Responsibilities: Review and approval of QAPP, project coordination, and review of draft deliverables .

Steve Vandegrift, Quality Assurance Manager NRMRL, GWERD, Ada, OK. Responsibilities: QA review and approval of QAPP and final report, QA guidance, and management of QA audits.

Jim Weaver, Research Hydrologist, NRMRL, GWERD, SRB, Ada, OK. Responsibilities: task oversight, scenario development, modeling, code development, literature review, QAPP preparation and implementation, document authoring, ensuring the project adheres to the QAPP, and implementation of corrective actions identified during audits and reviews.

Susan Mravik, Soil Scientist, NRMRL, GWERD, SRB, Ada, OK. Responsibilities: Scenario development, data collection, model application, contractor oversight.

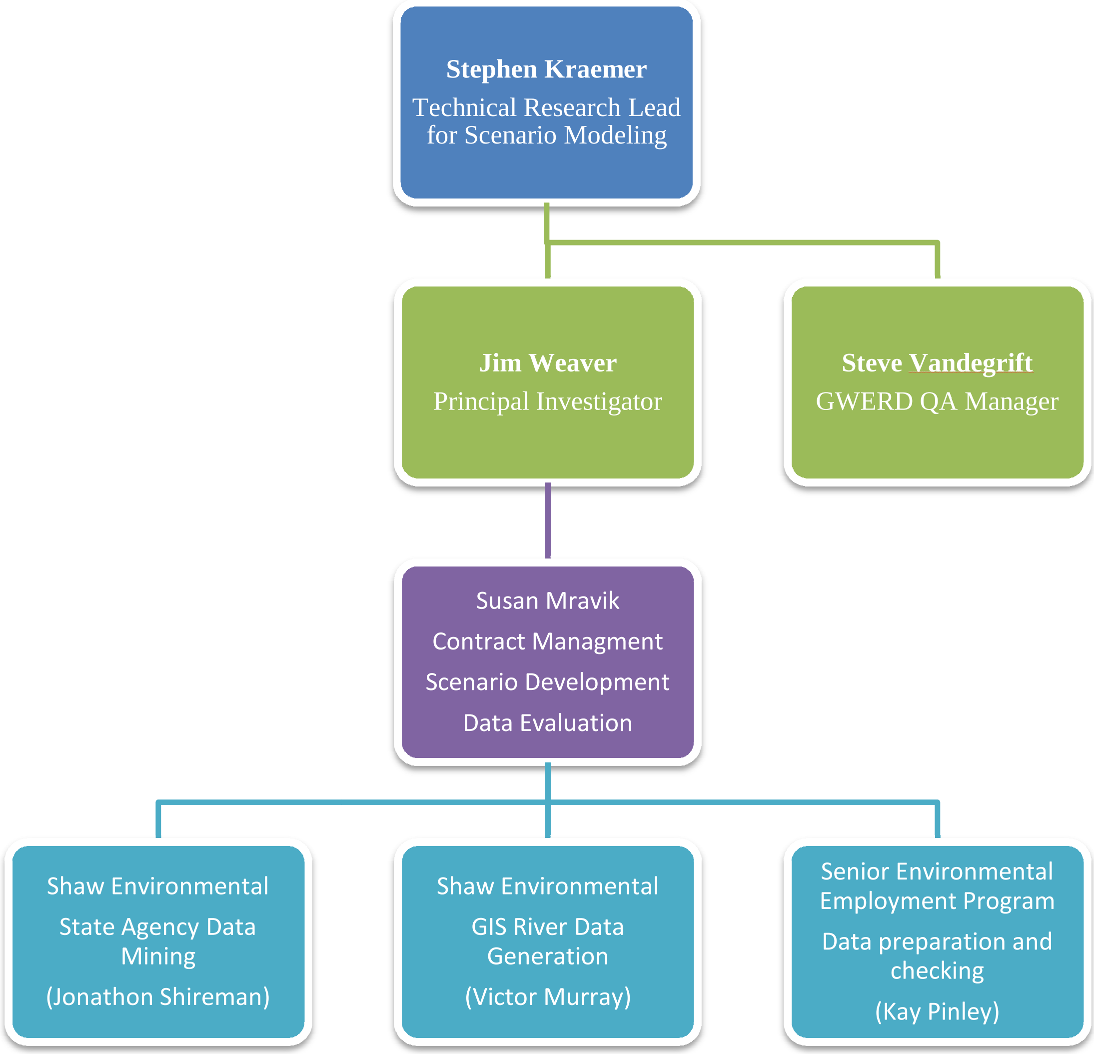
Victor Murray, Shaw Environmental, Map preparation, river parameter determination

Jonathon Shireman, Shaw Environmental, data mining from state agencies

Kay Pinley, Senior Environmental Employee Program participant, data preparation and checking

All project participants shall read and document that they have read this QAPP Revision No. 2 using the form on p. 39.

Organization Chart



A5 Problem Definition and Background

Production water management in the United States was studied by Clark and Veil (2009) who showed that most production water is disposed of by injection, although some was treated in waste water treatment plants. Disposal of hydraulic fracturing (HF) flow-back and production water to treatment plants and subsequent release to surface waters presents potential drinking water contamination problems. These might result from release of naturally occurring radioactive materials (NORMs), elevated concentrations of chloride and bromide, which lead to disinfection byproduct formation in water treatment plants, and possibly other compounds that are present in treated waste waters. NORMs have been characterized as simply gross alpha decay or by identification of specific nuclides, including radium and uranium 235. Increased chloride and bromide present potential problems as they increase the formation of both regulated and unregulated disinfection byproducts. Bromide, particularly, could be a problem when chlor-amination is used for treatment and can lead to a family of unregulated disinfection byproducts. If sufficient in magnitude, HF wastes might expand this problem to areas with low naturally occurring bromine.

Transport in rivers, from discharges to drinking water intakes, is potentially influenced by a number of processes that include river discharge, longitudinal and transverse mixing, turbulent diffusion, volatilization, sorption and decay. River discharge varies from point-to-point and day-to-day as it responds to changing rainfall and runoff. Waste is discharged to rivers at various locations, rates and compositions. Publically owned treatment works (POTWs) and industrial treatment plants may accept HF wastes on an intermittent basis and the composition, volume and frequency of discharge from these is expected to be variable. There may be blending of HF waste with other waste water to reduce impacts to receiving waters.

HF flowback and production water disposal methods vary around the U.S (Clark and Veil, 2009). In some locations deep well injection is used, while others, notably Pennsylvania, allow treatment of wastes at waste water treatment plants. Recycling of HF flow-back and production water for use in additional HF operations decidedly plays a role in the quantity and quality of waste water. These issues have been widely publicized so that future changes in disposal practices are likely. In so far as possible, this project will be designed to provide the most usable results.

Objectives

The primary objective of this project is to use models to illustrate the conditions under which disposal of hydraulic fracturing wastes might cause negative impacts on drinking water resources.

This objective will be met through two sub-objectives:

1) Simulating a generic river situation using the most accurate descriptors as possible to provide a first order view of problematic conditions.
2) Simulating one or more river networks to show potentially problematic conditions where given the actual locations of water intakes.

The project QAPP will be updated after the first set of results (the generic river simulation) is produced. This will allow for "lessons learned" to be incorporated into the QAPP for the watershed simulations. Since the modeling approach may be changed for the watershed-specific simulations, different tests will be undertaken for model sensitivity and uncertainty.

A6 Project/Task Description

Scenarios are to be developed to address surface water disposal of treated HF wastes. Definition of the scenarios provides the conceptual model for evaluation. The conceptual model includes definition of the river system, location of discharges and drinking water intakes, flow rates, discharge rate and composition, transport and transformation processes, required dimensions, and others. Since one focus should be on long-term impacts, the analysis could start with a baseline analysis. This baseline could be defined as both a steady flow in the river network and a steady discharge of treated HF waste, which represent a specific type of release into representatively-flowing river. Deviations from this baseline can address impacts at low flow or drought conditions where discharge might decrease and water demand might increase and conversely at high flow conditions. Waste disposal involving varying volumes or numbers of discharges (both increase and decrease), time-dependent loadings of discharges and varying concentration of effluents are a second set of factors influencing the scenarios. These two sets of factors generate a series of potential impacts for consideration of impacts at drinking water intakes.

Data requirements

The Monongahela, Allegheny, and Susquehanna River networks have been used for disposal of treated HF waste waters, and as such will be the focus of data collection and are likely candidates for development of scenario analysis. Data associated with these rivers that are needed for this study include the geometry of the river network, data on flows and bathymetry, where available from the U.S.G.S. river monitoring network. USGS tracer studies performed on these rivers (e.g., studies on the Susquehanna) may be useful for estimating travel time and effective diffusion coefficients. Data on the quantity and quality of HF discharges are needed. Sources include EPA, DOE and State Agency reports. Existing discharge/drinking water intake data will be sought to test the modeling approach for specific situations.

Model selection

The characteristics of the chosen scenarios, data and availability of model codes will be used to select the appropriate code or codes for simulation. Some simple transport calculations or analytical solutions to the transport equation may be useful in a rough screening analysis. With the need to accommodate increasingly complex features of river networks, numerical models are typically used (for example, the US EPA WASP model, QUAL2K, US COE RMA4, and others). All environmental models are dependent on their input data (see e.g., Oreskes, 2003), so the overall level of improvement in simulation results depends on model capabilities and also availability of input data. Major uncertainties exist in estimating travel times (equivalently transport velocities) and turbulent dispersion coefficients and tracer experiments are advocated as a means to determine their values (Jobson, 1996). For an analysis, such as this, where a generic approach is taken to determine where conditions may exist that are problematic, realistic estimation of model quantities is a critical consideration. To minimize these problems an approach will be taken in this work that relies on tracer data, and its empirical analysis coupled with numerical simulation of the two types described below. Because the empirical approach does not eliminate uncertainties in velocity and dispersion, nor eliminate uncertainty in other parameters, uncertainty analyses will be integrated into the calculations.

The empirical/statistical approaches pioneered by Holley and Jirka (1986) and Jobson (1996) will be evaluated for use in the generic river simulation phase of the project. These models are based on compilations of tracer data. The advantage of these approaches for generic screening is that they 1) are based on rivers from around the U.S., 2) use tracer data from actual experiments, so that they do not require assumptions on travel times and dispersion coefficients and 3) are complimentary to the simulation programs described above.

Jobson's technique was developed for application to instantaneous releases in single reaches, which are characterized by single values of slope, discharge, average annual discharge, and drainage area. Thus the method will be generalized for 1) rivers with varying reach properties, 2) branching river networks, 3) continuous injections of specified duration. The code will be implemented in Java and tested against available tracer data. Comparisons will be made and documented to one or more of the models mentioned above.

Tracer-based Empirical Transport Estimation

Jobson (1996) developed an empirically-based approach to estimate travel time and longitudinal dispersion in rivers and streams. The method relies on compiled tracer data so that the result is largely based on observation of transport in real systems. The motivation for this approach is stated by Jobson

> "In general there are no reliable methods of determining prediction dispersion coefficients (mixing rates) from commonly available hydraulic information. Stream velocities, typically predicted by use of a flow model, generally require very detailed channel geometry and flow resistance coefficients, which are seldom available. The availability of reliable input information is, therefore, almost always the weakest link in the chain of events needed to predict the rate of movement, dilution, and mixing of pollutants in rivers and streams."

Much of this statement remains true fifteen years later, although advances have been made in predicting longitudinal dispersion coefficients (see below). The data-limitation problem can be overcome by using tracer data, as noted by Jobson:

> "Measured tracer-response curves produced from the injection of a known quantity of soluble tracer provides an efficient method of obtaining the data necessary to calibrate and verity pollutant transport models."

Jobson's (1996) procedure relies on a series of regression formulas he developed from tracer data. They represent, collectively, the response of rivers and streams to solute injection experiments. In order to compare data from rivers of diverse sizes and injections of various amounts, the data are normalized by the mass of injection, flow rate, and mass lost to sorption or degradation. The remaining variable, the longitudinal dispersion, is assumed to be comparable given this normalization (Jobson, 1996). Jobson (1996) then used data from 60 rivers, 109 tracer injections, and 422 cross sections to develop the regression equations. The river discharges ranged from a mean annual discharge of 1.3 m^3/s in a small creek to 11,000 m^3/s in the Mississippi River. The slopes ranged from 36.0 m/km in the creek to 0.01 m/km in the Mississippi River.

From the 422 cross sections that had data on annual mean flow, Jobson found that the unit peak concentration was represented by

$$C_{up} \qquad 1025 \;\; {}_{p}^{\;0.}$$

where C_{up} is the peak unit-concentration [sec^{-1}] and T_p is the time to peak concentration in hours. A unit concentration, C_u [T^{-1}] , is determined from

$$C_u \quad 1 \quad 10 \quad \frac{C}{}\,\text{—}$$

where C is the concentration [M/L^3], R_r is the recovery ratio [dimensionless], Q is the stream discharge [L^3/T], and M_i is the mass injected [M]. The recovery ratio is defined as the mass passing a cross section to the mass injected. Although called a concentration, the unit concentration is actually partially non-dimensional mass flux (mass flux per unit mass of injected solute), which retains the time unit in the denominator. The factor of 1×10^6 is a convenience. Jobson refined the estimate of peak unit-concentration, by including the ratio of river discharge, Q [L^3/T] to mean annual river discharge Q_a [L^3/T]. The resulting equation is

$$C_{up} \qquad 857 \;\; {}_{p}^{\;0.} \qquad \text{—}$$

In several cases, data for a river show dependence on the relative discharge (Q/Q_a), although this is not always the case (Jobson, 1996, figures 4 through 7).

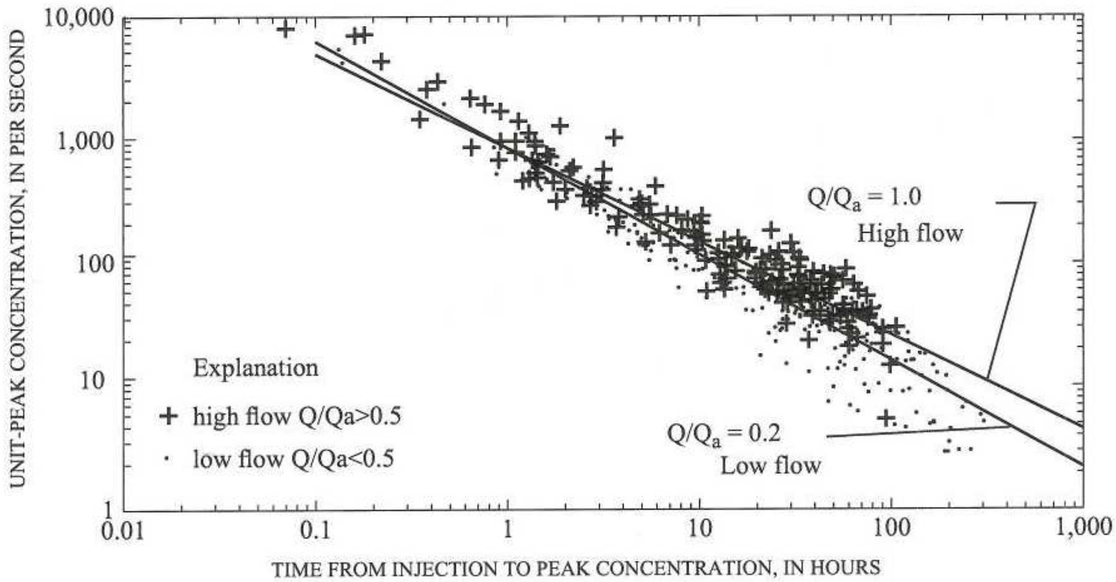


Figure 1 unit-peak concentration regression formulas developed by Jobson (1996).

Variables that influence the transport time include the drainage area, D, [L^2], the reach's slope, *S*, [dimensionless], the mean annual discharge, Q_a, [L^3/T], and the discharge at time of measurement, Q, [L^3/T]. These were used in forming regression formulas for travel time of the peak concentration. The most accurate equation for the peak velocity, V_p, [m/s] contained all variables:

$$_p \quad 0.094 \quad 0.0143 \cdot \quad ^{0.} \quad ^{0.} \quad -$$

where *D'* is the dimensionless drainage area defined b $^{0.} \quad -$, *g* is the acceleration of gravity [L/T^2], and *Q'* is the dimensionless discharge defined by Q/Q_a. Discharge *Q* is expressed in m^3/s, and drainage, *D*, in m^2 for this and the following regression equations. Alternate regressions were developed for situations where some of these variables are unavailable. When slope is not available,

$$_p \quad 0.020 \quad 0.051 \cdot \quad ^{0.} \quad -$$

When slope and mean annual discharge are unavailable,

$$p \quad 0.152 \quad 8.1 \quad \cdot \quad -$$

where the dimensionless drainage area was redefined by $\cdot \quad -$.

Bounding estimates that give velocity values greater than 99 percent of the data points, p , were given for each of these situations:

$$p \quad 0.25 \quad 0.02 \quad \cdot \quad 0. \quad 0. \quad -$$

$$p \quad 0.2 \quad 0.093 \quad 0. \quad 0. \quad -$$

$$p \quad 0.2 \quad 40.0 \quad 0. \quad -$$

The time for the leading edge of the concentration distribution, T_l [hours], was found to be highly correlated to the peak arrival time, following

$$0.890 \quad p$$

The time for the trailing edge of the concentration distribution, T_{d10} [hours], was estimated from

$$2 \quad 10 \quad p$$

an equation which is based on the assumption that the area under the unit concentration curve is 1×10^6, and that half of the mass lies between the peak concentration and a point where the concentration is one tenth the peak value.

Because limited data from Pennsylvania streams and rivers were included in Jobson's formulas, Reed and Stuckey (2002) evaluated the Jobson equations for use in the Susquehanna, Delaware, and Lehigh River basins. They found that the equations show good agreement with time-of-travel studies at low and moderate flow rates. At high flow rates, the Jobson equations over-predicted travel times, and so, Reed and Stuckey (2002) developed a modified equation. Reed and Stuckey (2002) recommend using the Jobson equation for low to moderate flow rates, where $Q/D^{0.73}$ is less than about 2. For higher values of $Q/D^{0.73}$, the modified equation that should be used is

$$0.6067$$

where the velocity, , is expressed in ft/s, the discharge, , in ft^3/s and the drainage area, , in $miles^2$.

The empirical approach eliminates two problems of transport modeling: estimating travel times and turbulent dispersion coefficients. Normally, these are not well estimated from simple measurements made at gauging stations (Jobson, 1996, Deng et al., 2002). Thus, the primary value of the empirical modeling methodology is to eliminate the need for calibration to observed transport velocities (equivalently travel times) and dispersion coefficients. Since the empirical equations were originally developed for application to a single uniform reach, a procedure will be implemented in a Java model code to allow the calculations to be performed over a series of connected reaches. The downstream output of each reach will be used as the upstream input of the next reach. By allowing multiple reaches to connect at one point, calculation can be performed over a river network. The approaches for developing this code and its testing will be documented in laboratory notebooks, internally to the model code, and in electronic documents as appropriate.

The regression equations developed by Jobson are not without scatter which implies uncertainty in values determined from these equations (see Figure 1). The code will be used in an uncertainty mode to incorporate the uncertainty resulting from the regression formulas and any other uncertain parameters.

As a check on quality, two applications of numerical models shall be performed. First, a numerical solution of the transport equation will be embedded into the empirical model code. This will allow the empirical results to be duplicated on a reach by reach basis, using a numerical model. All of the assumptions concerning the parameters and linkages between the reaches will be the same; only the calculation method will differ.

From this point, the numerical model results will be extended in two ways. First, the dispersion theory developed by Aris (1956) will be used to determine the value of an effective turbulent diffusion coefficient for the empirical model results. In effect, the Aris (1956) theory provides a means to determine what value of a dispersion coefficient would produce the results observed in the empirical data set. Because of the scatter in the regression formulas, these dispersion coefficients vary and a probability distribution can be constructed from Monte Carlo simulation applied to the empirical model. Similarly, the travel times (equivalently transport velocities) vary due to regression formula scatter. Compiling these from a Monte Carlo simulation applied to the empirical model provides the input needed for Monte Carlo simulation using the embedded numerical model. Second, the Deng et al. (2002) model provides an advanced method for approximating turbulent diffusion coefficients. As shown by Deng et al. (2002) a closer fit to values estimated from data was achieved than through any previous theory. Deng's model also then provides the means to estimate turbulent dispersion coefficients.

The second use, of a numerical model, will be to confirm results from the empirical calculations using an established external numerical model (the US EPA WASP model, Ambrose et al., 1983, Ambrose et al., 2009, DiToro et al, 1981). This model has been applied to numerous EPA surface water transport projects and provides an independent check of the calculations from the empirical model and the embedded numerical model.

When the first (and subsequent versions) of the code is completed, a version number will be established and assigned to the java code. All of the input files, source codes, class files, software version information, and other necessary information will be archived in the directory:

> MyDocuments/research/HF/SurfaceWaterScenario/Weaver/100-Modeling/workspace-Date-Version

("Date-Version" will be added as appropriate)

("workspace" is the default used by the Java Eclipse development environment)

Below workspace in the hierarchy is:

directory	**Contents**
/RiverModel/.settings	Eclipse development environment settings files
/RiverModel/bin	Java class files (required for execution
/RiverModel/src	Java source files
/RiverModel/results	Archived results from test problems and applications

A division-wide versioning software system will be researched for improved software versioning.

An electronic document will be prepared to describe the required input for the model. The document will be stored in the hierarchy described in section A9.

Definition and Testing of a Simplified Scenario (Objective 1).

A simplified generic scenario will be developed to assess the general characteristics of releases of treated water to surface waters. The conceptual model will consist of an idealized river section with generalized inputs and receptors. The inputs, however, will be generated from as realistic information as possible, given the constraints of time, required high-level quality assurance and data availability. The scenarios will be developed based on locations where discharges actually occur. Data on oil and gas waste disposal in Pennsylvania will be mined to generate these locations. These selections, in turn, determine the size and properties (slope, drainage area, annual discharge) of the river network. For example, Williamsport Pennsylvania was the location of HF waste discharges during the first half of 2011 (https://www.paoilandgasreporting.state.pa.us/publicreports/Modules/Welcome/Welcome.aspx). Williamsport is located on the West Branch of the Susquehanna River where data from USGS station 01551500 are available.

From this type of input, the model is expected to be able to generate a general guide to releases of treated HF wastes that allows exploration of the ranges of parameters which generate or mitigate drinking water exposure. To make the simulation results realistic, actual locations will be used in the generic simulations, which present a departure from the original plan for the project. The reason for this change is that 1) specific locations are needed to drive the Jobson (1996) empirical modeling approach, and 2) estimates of treated wastewater discharges are highly localized. By looking at specific locations where HF wastes are/were disposed, the results will be most defensible, because some drainage-ways in areas of intensive HF activity are not used for treated wastewater disposal.

Selection of Test Watershed and Definition of Scenarios (Objective 2).

The watersheds will be prioritized by the amount of available data. The most data-rich watershed will be selected first for development of a simulation and establishment of scenarios. The scenarios will include varying of the variables described above to develop watershed-specific versions of the simplified scenario described above, but with constraints built in from the location and nature of specific facilities. The results of the watershed scenarios will be compared against the generic simulations to determine the ability of the generic constraints to capture the watershed characteristics. Required details for the application to the watersheds will be added in a revision to the QAPP, once the generic simulations and model development is completed.

I. Expected products (outputs) and impact (outcomes)

output:

Draft Paper on empirical/numerical methodology, April 2013, describing the tracer data, empirical model, extensions to the empirical model, numerical model confirmation and monte carlo results

Draft Paper on generic scenario modeling, July 2013, definition of scenarios, review of data collected, model results

Draft Model User's guide, Sept 2013. review of methodology, input guide for model, example results, sensitivity and uncertainty results.

Other outputs to be determined

outcome:

Assessment of conditions that make surface water discharge of treated HF waste problematic.

II. Milestones and status

Aug 1, 2011: Completed and approved original QAPP.

Oct 1, 2011: Draft summary of literature on surface water transport (intended as background material for 2012 report to Congress)

Dec 1, 2011: Transport model and associated analytical solutions selected; Generic scenario model described in draft document

April 15, 2013: Completion of draft paper describing modeling approach and containing comparisons against tracer data

July 15, 2013: Completion of draft of generic results from western Pennsylvania-derived data.

Sept 30, 2013: Completion of draft user's guide for model

A7 Quality Objectives and Criteria

In this project the quality objectives are:

- To obtain data to support modeling studies that are of known quality
- To document the correct application of data interpretation and analysis methods
- To document steps in model code development
- To perform simulation results where the trail from input data to model outputs is as transparent as possible
- To assess uncertainty in the model results
- To retain records that document the activities of the project

Data with Known Quality

Data form the basis of inputs for the scenario modeling. These will be drawn from published peer-reviewed journal papers, federal agency reports, and state agency-accepted data. Information in these documents will be used to judge the quality of the data. Where available "supplemental information" from the papers will be used and saved as part of the quality documentation. See Section B9 for more detail.

Data Analysis Methods

Where necessary to interpret or manipulate data from various sources, the methods used will be documented in the lab notebook. Documentation will include the methods and their sources, example results with correctness verification and location of any spreadsheets or other resources used in calculation. The empirical methods developed by Holley and Jirka, 1986, and Jobson, 1995 are examples of where these approaches are likely to be used.

Documenting Code Development

Development of any code to implement a model shall be documented in laboratory notebooks of the project participants. A Java language code is anticipated to implement varying reach properties, tree-searching, uncertainty analysis, and multiple inputs for the Jobson (1996) empirical method. The code itself shall contain internal documentation to describe the functioning of the model. Model outputs and inputs shall be stored electronically according to the structure described in section A9. Test problems shall be referenced in laboratory notebooks (locations of electronic files) and described along with relevant results.

Simulation Results

The basis of simulation results will be documented by drawing a path from the input data to specific model results. This will be largely documented in the laboratory notebooks of the participants. The lab notebooks will be unique to this project. The documentation will include the development of a conceptual model for the transport scenarios, documentation of the sources of inputs, model results, any complication of model results—as in a spreadsheet, and the source for interpretation of the results.

All model results must be within parameters set by the numerical model developers (i.e., within mass balance targets) to be accepted.

As an alternative to direct simulation the use of statistical models will be explored in this project. Because a generic applicability is sought in the first phase of the work, statistical models (Jobson, 1996, Kilpatrick and Taylor, 1986, and Holley and Jirka, 1986) will be explored for their usefulness in this work. These provide an approach based on data analysis from around the U.S., although each might use a differing underlying data set. The assessment of these approaches can come from comparison against each other's results, other analytical models of transport and expert judgement.

Model Uncertainty

Since most of the model inputs are anticipated to have variability or uncertainty associated with them, the model is not expected to produce one single-valued result. Typically, only ranges or probability distributions of model outputs are justified from environmental simulation models. As such a set of appropriate scenarios will be constructed to illustrate the appropriate uncertainties in the model results. Characteristics of the problem and the model results will point to the appropriate scenarios. It is anticipated, however, that high, medium and low flow situations will be of interest. Others will be defined as appropriate.

In addition to variation in flows, parameters describing the release of treated waste water and its transport in river networks have uncertainties associated with them. Monte Carlo simulation will be used to perform an uncertainty analysis on the model results. For the empirical model, such simulation is necessary because of scatter in the regression formulas for travel time and concentration. Other uncertain parameters include the timing, discharge and concentration of the release. These parameters will be modeled as having ranges of uncertainty as defined by data from actual treatment plant discharges (NPDES permit data). At a minimum the range of values will be used to generate a uniform distribution for simulation (Weaver et al., 2002, Weaver, 2004, Tillman and Weaver, 2006), where possible an enhanced distribution will be created from the observed data and used as an empirical cumulative probability distribution for these quantities.

Statistical analysis

Monte Carlo analysis produces probability distributions as outputs and statistical analysis will be used for two purposes. First, non-parametric tests will be used to determine when a sufficient number of Monte Carlo simulations have been performed. The tests are non-parametric because an output distribution is not assumed. Candidate tests include the Kolmogorov-Smirnov test (or the related Mann-Whitney test). The tests will be performed as part of the output from the EPA river model and the results checked against standard software, namely MINITAB release 14.13. In terms of formal statistical analysis, a null hypothesis will be defined as the likelihood that an output developed from a fixed number of Monte Carlo simulations is statistically the same as an output developed from a greater number of simulations. If the null hypothesis is true, then a sufficient number of simulations has been performed.

The second use of statistics is to determine how the Monte Carlo output distributions are related to the concentration of concern. An approach to this determination is to state a null hypothesis that the true mean of the Monte Carlo results is exactly equal to the contaminant level of concern (Moore et al., 2009, pp 382-383). The alternate hypotheses are that the mean of the Monte Carlo results is higher or lower than the level of concern.

Document Retention

Model inputs and results that will form the basis of outputs from this project will be adequately documented for future tracing. The names and an outline of the contents of the files will be recorded in the lab notebooks of the participants. See Section A9 for more detail.

During development, the code resides on an EPA computer, external backup drives and EPA shared drives. Once final version of the code is developed, they will be archived along with input data sets and results files. We are currently (11-19-2012) investigating a configuration management system.

A8 Special Training/Certification

No special training is anticipated at the time of this writing.

A9 Documents and Records

All project documents will be stored in electronic form on Agency computers. The local "MyDocuments" synchronization feature will be used for storage and backup. The documents will be divided into two broad categories: records and non-records. "Records" will be used for all work produced by this project. "Non-records" will be used for information copies of documents.

The project is expected to produce, non-records that consist of informational copies of journal papers, agency reports and others. The records produced for the project will consist of data used in simulations, reduced data used in simulation and methods of data reduction, definition of model scenarios, model input files, model output files, interim reports (milestones) and a final report.

The project plan will be saved as a record under the directory and title:

MyDocuments/Research/HF/SurfaceWaterScenario/Weaver/Records/001ProjectPlan/

The QAPP will be saved as a record under the directory and title:

MyDocuments/Research/HF/SurfaceWaterScenario/Weaver/Records/002QAPP/Weaver-QAPP-surface water scenario modeling-revXX.docx

The “XX” will reflect the version number of the QAPP, beginning with “00” The version number is anticipated only to change if approved changes and additions are made to the initially-approved QAPP.

Each project participant will be supplied with the copy of the QAPP. Additionally the QAPP will be continuously available from the ORD O: drive. Each EPA participant will establish a similar directory structure for storage of their documents. They shall replace the Directory

MyDocuments/Research/HF/SurfaceWaterScenario/Weaver

with a similar directory containing their name.

Data for this project are planned to be obtained from journal papers, published reports and other appropriate sources. These will be considered to be non-records. Electronic copies will be stored as described above. For example:

MyDocuments/Research/HF/SurfaceWaterScenario/Weaver/005NonRecords/Agency Reports

MyDocuments/Research/HF/SurfaceWaterScenario/Weaver/005NonRecords/ConsultantReports

MyDocuments/Research/HF/SurfaceWaterScenario/Weaver/005NonRecords/Literature

MyDocuments/Research/HF/SurfaceWaterScenario/Weaver/005NonRecords/OutsideCommunications

MyDocuments/Research/HF/SurfaceWaterScenario/Weaver/005NonRecords/StateData

MyDocuments/Research/HF/SurfaceWaterScenario/Weaver/005NonRecords/University

General documentation developed for publication will be saved in

MyDocuments/Research/HF/SurfaceWaterScenario/Weaver/050-Documentation

Model inputs are not considered fundamental data sources, but are to be documented with any associated model outputs. Electronic files shall be named so that the model used, date and characteristics of the input can be briefly identified in appropriately designed directories, AND associated with the corresponding model outputs. For example, for the model “100”

MyDocuments/Research/HF/SurfaceWaterScenario/Weaver/Records/100-Modeling/aaa-ModelName

Where aaa is an arbitrary sequence number.

Supporting information for the model runs will be saved under appropriate directory titles, for example for the Jobson model:

MyDocuments/Research/HF/SurfaceWaterScenario/Weaver/Records/100-Modeling/001-Jobson

For the Jobson empirical model, and

MyDocuments/Research/HF/SurfaceWaterScenario/Weaver/Records/100-Modeling/001-Jobson/001-ReedAndStuckey

MyDocuments/Research/HF/SurfaceWaterScenario/Weaver/Records/100-Modeling/001-Jobson/002-SusquehannaTimeOfTravel

MyDocuments/Research/HF/SurfaceWaterScenario/Weaver/Records/100-Modeling/001-Jobson/003-Susquehanna Flow

MyDocuments/Research/HF/SurfaceWaterScenario/Weaver/Records/100-Modeling/001-Jobson/004-RiverMileData for Tioga-Chemung-Susquehanna

MyDocuments/Research/HF/SurfaceWaterScenario/Weaver/Records/100-Modeling/001-Jobson/005-RiverTracerData

Model inputs and outputs will also be documented in a laboratory notebook, giving the sense of the simulations performed and the locations of electronic computer files in the directories as indicated above. Models and versions used will be documented in the laboratory notebook.

For the separate scenarios that will be developed in this project they will be numbered and catalogued in the lab notebook. For example:

MyDocuments/Research/HF/SurfaceWaterScenario/Weaver/Records/100-Modeling/001-Jobson/Scenario1

MyDocuments/Research/HF/SurfaceWaterScenario/Weaver/Records/100-Modeling/001-Jobson/Scenario2

Overall the electronic data scheme will follow:

MyDocuments/Research/HF/SurfaceWaterScenario/Weaver/ 001-ProjectPlan

MyDocuments/Research/HF/SurfaceWaterScenario/Weaver/ 002-QAPP

MyDocuments/Research/HF/SurfaceWaterScenario/Weaver/005NonRecords

MyDocuments/Research/HF/SurfaceWaterScenario/Weaver/005NonRecords/Literature

MyDocuments/Research/HF/SurfaceWaterScenario/Weaver/005NonRecords/AgencyReports

MyDocuments/Research/HF/SurfaceWaterScenario/Weaver/005NonRecords/

MyDocuments/Research/HF/SurfaceWaterScenario/Weaver/

MyDocuments/Research/HF/SurfaceWaterScenario/Weaver/100-Modeling

MyDocuments/Research/HF/SurfaceWaterScenario/Weaver/100-Modeling/aaa-ModelName/Scenario1

MyDocuments/Research/HF/SurfaceWaterScenario/Weaver/100-Modeling /aaa-ModelName/Scenario2

Subdirectories from these major directories will be created as needed.

Model source code shall be stored in directories as follows:

MyDocuments/Research/workspace/RiverModel/src/riverModel

Java Class files shall be stored in:

MyDocuments/Research/workspace/RiverModel/bin/riverModel

Results from test problems shall be stored as follows:

MyDocuments/Research/workspace/RiverModel/results

This directory structure follows the requirements of the Eclipse development environment (see section B10) used for this project.

Contractor (Shaw Environmental) Directory Structure

To match project 5b and other requirements, Shaw Environmental will follow a different directory structure:

C:\Projects\HydraFrac\

Then because of the multiple sites around the US it is more specific

C:\Projects\HydraFrac\PA_SRB\

And finally your TD Data path is

C:\Projects\HydraFrac\PA_SRB\RiverMile

Then the backup goes to the Kerr Facility IT archive drive path of

L:\Lab\CSMOS\8CS - Shaw Option 3 TD - 2011\8HF - HydraFrac

Then to the specific TD(s)

L:\Lab\CSMOS\8CS - Shaw Option 3 TD - 2011\8HF - HydraFrac\8HF116HF

Data files received from Shaw (by EPA) will be saved according to the structure described in the beginning of section A9.

Because this project is assigned a level 1 QA Category, all paper project records require permanent retention per Agency Records Schedule 501, *Applied and Directed Scientific Research.* Records will be stored in room 211 (Weaver's office) in the GWERD until they are transferred to GWERD's Records Storage Room. At some point in the future, paper records will be transferred to a National Archive facility.

All documentation shall provide enough detail to allow for reconstruction of the project activities. Documentation practices shall adhere to ORD PPM 13.2, "Paper Laboratory Records."

Records will be moved to the HF project O: drive when work is completed.

B1-B6, B8 Sampling and Measurement Requirements

The following list of sampling and measurement requirements appears in "EPA Requirements for Quality Assurance Project Plans " (EPA QA/R-5, EPA/240/B-01/003). These items were considered for this plan but were judged non-applicable to a literature, data evaluation, and modeling study.

B1 Sampling Process Design

B2 Sampling Methods

B3 Sample Handling and Custody

B4 Analytical Methods

B5 Quality Control

B6 Instrument/Equipment Testing, Inspection, and MaintenanceB8 Inspection/Acceptance of Supplies and Consumables

B7 Sampling and Measurement Requirements

B7 Instrument/Equipment Calibration and Frequency

Calibration. The Jobson (1996) empirical approach is uncalibrated. Because the underlying dataset used to develop the approach used rivers of all sizes in the US (including the Mississippi) the empirical model applies anywhere in the US. A special study determined that the methods applied to Pennsylvania (Reed and Stuckey, 2002), with some exceptions, which will be used in this project.

For use of the empirical model, the appropriate testing procedure is to demonstrate that the uncalibrated model results match data from a tracer experiment. Seven experiments that cover a range of flow conditions are being considered for testing the empirical model (Antietam Creek, Monocacy River, Tangipohoa River, Red River, Wind River, Mississippi River and the Yellowstone River). The last of these (Yellowstone River) was not used in generating the empirical equations—so it provides a test of the predictive capability of the method. Application to the other rivers is essentially equivalent to calibrating a numerical model, as the model is forced to match the experimental data. Calibrated models can only be said to represent the data to which they were calibrated—no extrapolation is demonstrated by calibration. The Yellowstone River application, in contrast, is an extrapolation of the method beyond its calibration data set. Thus model results which demonstrate correct simulation of the Yellowstone experiment would demonstrate a higher level of model testing than analogous calibration of a numerical model.

Goodness-of-fit for the empirical model will generally be taken by visual observation of concentration histories (breakthrough curves) at sampling locations, through professional judgment. No calibration is warranted for this approach. Because of scatter in the regression equations, tracer data are only expected to be fit in a statistical sense (see section A6). For the unlikely possibility that the empirical model is not able to match the observed data, recourse will be made to numerical simulation using estimated dispersion coefficients and travel times. Ultimately, calibration to the tracer data can be used to fit the numerical model to specific waterways if necessary.

For models that require calibration testing, data are available from numerous USGS tracer experiments (Nordin and Sabol, 1974). In these cases, goodness-of-fit will be determined from quantitative measures (i.e, least squares). To support use of the empirical model, numerical models will be fitted to the same tracer data sets. Uncertainty analyses will be applied where appropriate.

Sensitivity analysis: Here sensitivity analysis is defined as determining the unit change in model response to a unit change in parameter input. A sensitivity analysis will be performed for the model, but by recognizing that unit sensitivities are often misleading, because parameters interact to produce the model output. A sensitivity analysis that is based on interaction of all parameters will be produced for the user's guide.

B9 Non-direct Measurements

The data needed for this project all fall under the category of non-direct measurements. These are discussed in items "A7 Quality Objectives and Criteria" and "D1 Data Review, Verification, and Validation".

The data anticipated include:

- data on flows in specific rivers of interest that are available from USGS gages
- data on watershed characteristics
- USGS tracer study data

- data on discharges from publicly-owned and commercial waste water treatments that treat HF wastes
- data on concentration of typical flowback and produced water (FB/PW)
- data on concentrations of disinfection byproduct producing chemicals that created potential impacts on drinking water resources
- data on background concentrations of all chemicals of concern

Data Sources:

Four major data sources will be used for the project:

1) USGS data on flows and watershed characteristics are presented in a finalized, reviewed from on their web site (http://waterdata.usgs.gov/nwis/sw). These will be taken for selected river courses of interest and the following information will be gathered: average annual discharge, monthly average discharge, drainage area, location (latitude-longitude), gage elevation.
2) USGS tracer data, published in USGS reports will be accepted as being of acceptable quality. An early compilation (Nordin and Sabol, 1974) was used by Jobson (1996) as one part of the data for developing his regression equations. Other studies conducted since 1996 have potential usefulness for independent testing of the equations. One such experiment was conducted by McCarthy (2009) in the Yellowstone River.
    a. An internal EPA review will be made of published USGS tracer data. Data will be checked for consistency and errors.
3) Data submitted to state agencies or US EPA in fulfillment of legal requirements (i.e., national pollutant discharge elimination system (NPDES) required monthly monitoring; Pennsylvania mandated reporting of quantities of oil and gas waste). Data are available for Pennsylvania on treatment of oil and gas wastewaters and their disposal methods and locations. These shall be collected by Shaw Environmental according to the procedures and QAPP developed for HF project 5b on water acquisition.
    a. NPDES data obtained from industrial waste water treatments plants will be reviewed in accordance to QA data supplied with the NPDES reports.
4) Data on flowback and produced water for the Marcellus Shale will be used to estimate concentrations in wastewater, as the Pennsylvania data contains only waste volumes. For example, Rowan et al., 2011 compiled data on the Northern Appalachian basin of the US. These data are directly applicable to generating input conditions for this modeling. Some of the needed data include: disinfection byproduct-generation (i.e.,

Krasner et al., 2006) and background concentration data for bromide, total dissolved solids and naturally occurring radioactive materials.

Data from the four above-mentioned sources will be accepted as being of sufficient quality for this project: USGS data streamflow data, data from USGS reports, data reported to state agencies for compliance, and peer-reviewed, published literature.

Since the purpose of the modeling is to assess generic disposal conditions, input parameters describing the discharge flow, discharge concentration, and discharge timing will be treated as uncertain, but based on the NPDES permit data. For example, if the discharge concentration from a plant varies over several models from x mg/L to y mg/L, the discharge on any given day could be constructed randomly from a uniform distribution with end points of x mg/L and y mg/L.

Procedures developed for minor calculation/manipulation of data will be described and documented in the participant's laboratory notebooks. Subsequent usage of these calculations will be referenced to appropriate pages of the notebooks.

B10 Data Management and Hardware/Software Configuration

Data for this project will be stored as described in section "A9 Documents and Records," which includes electronic input and output files, spreadsheets and laboratory notebooks.

The PI is responsible for maintaining data files, including their security and integrity. All files will be stored (electronic) and labeled to identify this project.

Laboratory notebooks of the researchers will be the primary key to all data used in the project. The PI's (Weaver) notebook will summarize all data, models and model applications for the project. A spreadsheet/word document will be developed to summarize all available data. This spreadsheet will contain a description of the item, source, and location of computer files containing more information (if applicable). This spreadsheet will be continuously available to all project members by using the ORD O: drive. Ultimately the spreadsheet will become part of the project report.

The data management spreadsheet will be stored in

> MyDocuments/Research/HF/SurfaceWaterScenario/Weaver/100-Modeling/aaa-ModelName/000-DataSummary

And will be named

SurfaceWaterScenarioDataManagement.xlsx

Similarly model runs will be catalogued in a spreadsheet/word document

The data management spreadsheet will be stored in

MyDocuments/Research/HF/SurfaceWaterScenario/Weaver/Records/Modeling

And will be named

SurfaceWaterScenarioModelRunManagement.xlsx

Data in hard copy form will be manually entered into Excel spreadsheets or model input files on designated GWERD staff computer and will given to the PI.

A minimum of ten percent of electronic data in spreadsheets or in model input files will be spot-checked to ensure accuracy of the transfer. If errors are detected during the spot-check, the entries will be corrected. Detection of an error will prompt a more extensive inspection of the data, which could lead to a 100% check of the data set being entered at that time if multiple errors are found. The checks shall be documented in lab notebooks to demonstrate that appropriate checking has been performed, and that corrections have been made. Spreadsheet cells that are corrected shall be colored to show where changes were made.

Model inputs and model outputs will be validated by initial and final reviews: This will include checking to assure that the model input files contain the intended input values, and after completing model runs, that model outputs correspond to the correct sets of inputs. When compiled for presentation, compilations (likely to be in spreadsheets) will be checked against actual output files, using the 10%/100% checking criteria described above.

USGS stream gage data will be used to generate parameter estimates needed for use in the Jobson equations. Treatment of these data shall be as follows:

Discharges and drainage areas will be obtained from the database developed under project 5b and its QA procedures.

Distances along rivers: Distances along rivers are calculated in ESRI software using built-in procedures. The calculated distances will be spot checked (minimum of 10%; 100% check if multiple errors found) using appropriate techniques (i.e., subdividing the river into segments, then adding the segments to assure that the results from both approaches are consistent). Checks will be documented in lab notebooks.

Slopes: Slopes are calculated from gage elevations and distances along the rivers. A hand check of calculated river slopes will be performed, using the criteria for data checking (minimum of 10%; 100% check if multiple errors found).

Hardware and Software Configuration

Hardware: Calculations will be performed with agency standard CTS and ORD computers. Two immediately available machines are in the possession of the PI: CTS Dell #CTS008316 and ORD Dell Latitude D630, decal #002507.

Software: Both of these machines use Windows XP. Standard software will be used in this project: Microsoft Office Word and Microsoft Office Excel are planned for data evaluation and generic screening calculations. Modeling codes such as WASP, QUAL2K, or others will be employed for the watershed simulations.

Java program development shall be undertaken using the Java Version 1.6 or higher and the Eclipse Development Environment. Eclipse is documented at http://www.eclipse.org/.

C1 Assessments and Response Actions

Model Performance Testing

The performance of all models used in this task will be tested against available USGS tracer data. These data will be used for ground truthing the model application as they will be used to answer the question of "How well does the model represent actual conditions in the field". These assessments will be documented and reported as part of the project results.

USGS tracer data are available for numerous rivers, streams and creeks around the US (e.g., Nordin and Sabol, 1974, P.M. McCarthy, 2009). These experiments generally consist of release of a known mass of tracer dye into a flowing river. Concentration-versus- time data are collected at a number of downstream locations. These tracer data provide the means of testing numerical models, because the data directly incorporate travel time and turbulent dispersion of the tracer. The empirical calculations of Jobson (1996) permit calculation of travel time and concentration without calibration. Thus, a test of this calculation method is its ability to replicate tracer experiment data. Several data sets will be selected from USGS tracer literature and used to test Jobson empirical results. Results, from these tests, will be documented in laboratory notebooks. Since numerous tracer experiments exist, a set of data from Nordin and Sabol, 1974 will be used for the testing. Rivers/creeks will be chosen to cover the entire range of flow rates in the Nordin and Sabol data set. One or more data sets will be selected that were not used in development of the Jobson equations (Yellowstone River, McCarthy, 2009). Since models are inherently dependent on the choices made for their inputs, the response of the model to variation in inputs will be determined. Sensitivity analysis seeks to determine the response of the model to a unit change in each of its inputs. The results of sensitivity analysis allow an importance ranking of the parameters to be established, with the implication that the most important parameters should receive the most attention—i.e., additional data collection and refinement of estimated values.

Uncertainty analysis brings to the assessment, the combined influence of sensitivity to input parameters and the range of values seen in model application. In the planned work the parameters to be varied are:

Discharge concentration, flow rate, duration,

River flow: high, medium or low,

River characteristics: slope, drainage area, average annual discharge

Distance to nearest receptor.

Results from repeated simulation will determine if a consistent pattern of parameter importance exists. If so, this will be documented as a result of the simulations.

The model results will be based on mass balance, travel time and peak concentration. For each tracer experiment the mass injected, travel time to each measurement station and peak concentration are available and will be used to judge model results.

Mass balance errors are expected to be controlled to be less than 10% between the empirical model results and a specific tracer experiment. Because the empirical data analysis was not developed with a specific mass balance constraint, 10% error is considered appropriate.

Because the regression equations used in the empirical model have scatter, the travel time and peak concentration are not expected to always match individual tracer data. Tracer data are expected, however, to fall within the boundaries of statistical, Monte Carlo results from the model. A selection of tracer experiments will be used to demonstrate the model performance for conditions covering the range of observed flows. When the numerical model is used in an uncalibrated manner with predicted transport velocities and dispersion coefficients, a similar approach will be used: the statistical parameter of the Monte Carlo output will be expected to include the tracer experiment results.

Numerical models can be calibrated to tracer data. The match between the calibrated model and a tracer experiment will be quantified with a squared error criterion. In calibration the goal will be to minimize the squared error:

ˆ

where are the concentrations observed at a measuring location, and ˆ are the modeled concentrations.

QA Audits

A Technical systems audit (TSA) was conducted on March 1, 2012, early enough in the project to allow for identification and correction of any issues that may affect data quality. Detailed checklists, based on the procedures and requirements specified in this QAPP, related SOPs, and EPA Policies were prepared and used during the audit. A QA assessment (comparable to an Audit of Data Quality on measurement projects) will be conducted on a representative sample of data. This assessment and its timing was discussed during the TSA. This assessment will trace data from its source, through the modeling process to its output and compare it with that generated during the project. These audits will be conducted with contract support , with oversight by Steve Vandegrift, QAM.

See Section C2 for how and to whom assessment results are reported.

Assessors do not have stop work authority; however, they can advise the PI if a stop work order is needed in situations where data quality may be significantly impacted, or for safety reasons. The PI makes the final determination as to whether or not to issue a stop work order.

For assessments that identify deficiencies requiring corrective action, the audited party must provide a written response to each Finding and Observation to the QA Manager, which shall include a plan for corrective action and a schedule. The PI is responsible for ensuring that audit findings are resolved. The QA Manager will review the written response to determine their appropriateness. If the audited party is other than the PI, then the PI shall also review and concur with the corrective actions. The QA Manager will track implementation and completion of corrective actions. After all corrective actions have been implemented and confirmed to be completed, the QA Manager shall send documentation to the PI and his supervisor that the audit is closed. Audit reports and responses shall be maintained by the PI in the project file and the QA Manager in the QA files, including QLOG.

At the conclusion of a TSA, a debriefing shall be held between the auditor and the PI or audited party to discuss the assessment results. Assessment results will be documented in reports to the PI, the PIs first-line manager, and the Technical Research Lead. If any serious problems are identified that require immediate action, the QAM will verbally convey these problems at the time of the audit to the PI.

The PI is responsible for responding to the reports as well as ensuring that corrective actions are implemented in a timely manner to ensure that quality impacts to project results are minimal.

C2 Reports to Management

Progress reports will be made at the monthly project conference calls. These will include information on quality assurance and documentation.

All final audit reports shall be sent to the first-line manager of the PI, the Technical Research Lead, and copied to the PI. Audit reports will be prepared by the QA Manager with input from the QA support contractor where audit performance was delegated. Specific actions will be identified in the reports.

D1 Data Review, Verification, and Validation

Data review, verification, and validation will focus initially on the acceptability of literature data for simulation purposes. This review will rely on expert judgment, criteria presented in Section B9, and a broad knowledge of the literature on several topics including generalized transport in rivers, disinfection byproduct research and chemical transport.

Model verification, defined here as the determination that the model solves its equations correctly, will be demonstrated through specific documented tests of the software. These include tests of the geometry of river network connections, mass balance (error $<10\%$ considered acceptable), and tests of other internal algorithms. The latter include testing the construction of the triangular distributions used to create the empirical distributions, the empirical peak concentration calculation, determining the average concentration in empirical concentration distributions, determining the peak-to-peak transport time, and other algorithms seen to be critical to the results. Post-processing of the data is used to generate basic statistics and histograms from the results. These will be tested by comparison to results obtained from the output statistical distributions using functions in Microsoft Excel. Where internal statistical tests are used their results will be compared against results obtained from MINITAB (Version 14.13.)

Model validation, defined here as the demonstration that the model correctly serves its purpose, will be conducted primarily through comparison with published tracer experiments, the internal numerical model, and external numerical model(s) as was described in section A6. The numerical models have the capacity to be calibrated to the tracer data sets through parameter adjustment; therefore using the squared error (see section C1) provides a metric for judging the ability of the model to match the measured tracer distribution. Matches demonstrate the ability of numerical models to represent actual contaminant distributions, subject to the choice of parameter values. Because in this project there are not data to which to calibrate the models, the validation step is dependent on the use of the empirical model and the numerical model using the Deng et al. (2002) estimate of the dispersion coefficient. The use of these two approaches ties the current modeling to the data-based approaches which were used to develop these methods. By using only Monte Carlo results, with their incorporated uncertainty, the best use of the underlying empirical data will be made.

Data or model results will not be released outside of the Robert S. Kerr Environmental Research Center until they have been reviewed, verified and validated as described below. The PI is responsible for deciding when project data can be shared with interested stakeholders in conjunction with NRMRL Management's approval.

D2 Verification and Validation Methods

Quantitative comparisons will be used when allowed by data availability. These will be used to develop a metric, say least squares, that can provide an objective fitting parameter. These are expected to allow for the actual tracer data to be shown to be within boundaries predicted by the scenario model results. See more discussion in section D1

D3 Reconciliation with User Requirements

The project leader is likely to be a part of the writing team for the 2012 and 2014 reports to congress. Through this and the leadership of the theme lead (Stephen Kraemer), the model results will remain focused on the appropriate objectives. Dr. Kraemer will serve as a reviewer at each critical stage of the project.

References

Ambrose, R.B, Hill, S.I., Mulkey, L.A., 1983, User's Manual for the Chemical Transport and Fate Model (TOXIWASP), Version 1, EPA 600/3-83-005

Ambrose, R.B, Wool, T.A. , 2009, WASP7 Stream Transport Model Theory and User's Guide, 2009, EPA/600/R-09/100.

Aris, 1956, On the dispersion of a solute in a fluid flowing through a tube, Proceedings of the Royal Society of London, Series A, Mathematical and Physical Sciences, Vol. 235, No. 1200, April 10, 67-77.

Clark, C.E., Veil, J.A., 2009, Produced Water Volumes and Management Practices in the United States, U.S. Department of Energy, National Energy Technology Laboratory, Contract DE-AC02-06CH11357, 60 pp.

Deng, Z.-Q., Bengtsson, L., Singh, V.P., Adrian, D.D., 2002, Longitudinal Dispersion Coefficient in Single-Channel Streams, American Society of Civil Engineers, Journal of Hydraulic Engineering, October, 901-916, DOI: 10.1061/(ASCE)0733-9429(21002)128:10(901).

DiToro, D.M., Fitzpatrick, J.J. , Thomann, R.V., 1981, Documentation for Water Quality Analysis Simulation Program (WASP) an Model Verification Program (MVP), EPA 600/3-81-044.

Holley, E.R and G.H. Jirka, 1986, Mixing in Rivers, U.S. Army Corps of Engineers, Environmental and Water Quality Operational Studies, Technical Report E-86-11, Contract No. DACW83-M-0737.

Jobson, H.E., 1996, Prediction of Traveltime and Longitudinal Dispersion in Rivers and Streams, United States Geological Survey, Water-Resources Investigations Report 96-4013.

Kilpatrick, F.A., and K. R. Taylor, 1986, Generalization and application of tracer dispersion data, Water Resources Bulletin, 22(4), 537-548.

Krasner, S.W., Weinberg, H.S., Richardson, S.D., Pastor, S.J., Chinn, R., Sclimenti, M.J., Onstad, G.D., Thruston, A.D., 2006, Occurrence of a new generation of disinfection byproducts, Environmental Science and Technology, 40, 7175-7195.

McCarthy, P.M., 2009, Travel Times, Streamflow Velocities, and Dispersion Rates in the Yellowstone River, Montana ,United States Geological Survey, Scientific Investigations Report 2009-5261.

McClave, J.T., Sincich, T., 2003, Statistics, 9th ed, Prentice-Hall, Upper Salld River, New Jersey, 822 pp.

Moore, D.S., McCabe, G.P., Craig, B.A., 2009, Introduction to the Practice of Statistics, 6th ed, extended version, W.H. Freeman and Company, NY, 766 pp + appendices.

Nordin, C.F., Sabol, G.V., 1974, Empirical Data on Longitudinal Dispersion in Rivers, United States Geological Survey, Water-Resources Investigations 20-74.

Oreskes, N., 2003, The Role of Quantitative Models in Science, in Models in Ecosystem Science, C.D. Canham, J. J. Cole, and W. K. Lauenroth, eds., Princeton University Press, 13-31.

Reed, L.A. Stuckey, M.H., 2002, Prediction of Velocities for a Range of Streamflow Conditions in Pennsylvania, United States Geological Survey, Water-Resources Investigations Report 01-4214.

Rowan E.L., M.A. Engle, C.S. Kirby, T.F. Kraemer, 2011, Radium Content of Oil- and Gas-Field Produced Water in the Northern Appalachian Basin (USA): Summary and Discussion of Data, United States Geological Survey, Scientific Investigations Report 2011-5135.

Tillman, Fred D and James W. Weaver, 2006, Uncertainty from Synergistic Effects of Multiple Parameters in the Johnson and Ettinger (1991) Vapor Intrusion Model, Atmospheric Environment, 40(22) 4098-4112.

Weaver, James W., 2004, Characteristics of Spilled Oils, Fuels, and Petroleum Products: 3a. Simulation of Oil Spills and Dispersants Under Conditions of Uncertainty, United States Environmental Protection Agency, Washington, D.C., EPA 600/R-04/120.

Weaver, James W., Caroline L. Tebes-Stevens, Kurt Wolfe, 2002, Uncertainty in Model Predictions Plausible Outcomes from Estimates of Input Ranges, Brownfields 2002, Charlotte, North Carolina, Nov 13-15, 2002.

REVISION HISTORY:

Revision Number	Date Approved	Revision
0	9/10/11	New document
1	2/22/12	• A3, A4, Organization Chart: additional project personnel added • A6, Model Selection: additional text included on Jobson's technique • A6, Definition and Testing of a Simplified Scenario: text added on scenario development explanation of departure from original plan on generic simulations • Added description of Jobson (1996) equations • Updated timing of milestones on p. 10 • A7, deleted second paragraph in "Data with Known Quality" • A7, added section on "Documenting Code Development" • A7, added second paragraph to "Model Uncertainty" • A9, provided more detail on file structure for storing electronic copies and file naming convention • Added detail for Shaw Environmental data file structure • B7, added description of calibration and goodness-of-fit • B9,added detailed description of data that will be used and is considered acceptable • B10, added description of checks for each data type • C1, "Model Performance Testing," added discussion on the use of tracer data to test Jobson empirical results • C1 subsurface example switched to surface water example • C1, "QA Audits," provided clarification on ADQs • References, additional references added
2	2/28/2013	TSA audit responses: • Branch Chief name changed to David Jewett • Kay Pinley added as project participant • Distribution list updated • Signature sheet added for project participant awareness • A4 Documentation of minor calculation procedures added to section A4 • A6 Discussion of model choice included directly in QAPP: use of the empirical model, backed by an internal numerical approach and external confirmation using EPA wasp. • A6 Text added to describe usefulness of empirical approach, and the role for numerical models • Reference to Deng, 2002 added • A6 Confirmatory use of WASP model added • References to Ambrose et al. 1983, Ambrose et al., 2009 and DiToro et al., 1981 were added • Reference to Aris (1956) added • A6 Milestones revised • A9 code versioning described

		• A7 code input document described • Statement added that QAPP revision for watershed simulation will be added once model development and generic simulation completed. • B9 Acceptance criteria for USGS and NPDES data described • B9 Discussion of uncertainty in model inputs was added. • C1, added paragraph on model performance testing. EPA directed response to API sponsored review comments: • A5, additional production water management discussion added, although I note that the version 0 and version 1 QAPP already contained such information • Reference to Clark and Veil (2009) added • A5, Uncertainty analysis discussion added in sections B9,(review comments incorrectly identified uncertainty discussion in A5) • A6 model selection discussion clarified in section A6 (see above) • A6 the more intensive focus on the empirical/statistical approach is described in the revised QAPP (see above) • A6 as we have not moved to the watershed modeling section of the project, only a short clarification section has been added to the QAPP. • A6 funding section was deleted. • A6 description of products expanded and made more specific, although the option for including future products remains. • A7 enhanced information uncertainty analysis has been included in sections A7 and B9. • A7 the section on use of data has been expanded. • A9 planned products are included in section A6, additional information on their contents provided • A9 configuration management has been added • B7 the discussion on calibration has been reviewed and slightly updated. The reviewer fails to understand that the empirical model is not calibrated. • B7 corrective action for the case of the empirical model failing statistical representation of tracer data is described. • B7 the role of sensitivity analysis has been added to the QAPP • C1 enhanced discussion of uncertainty analysis and sensitivity analysis has been added to sections B9 and A7 • C1 reference to hydraulic conductivity and porosity has been removed, but the point made by their inclusion is valid. • Disclaimer and EPA Quality description added. • A7 Statistical analysis of Monte Carlo results added. Response to PQAM comments • Milestones were updated with current dates • TBD contract participants (SSC) have been removed, due to delay in awarding contract. QAPP will be revised when SSC contractor available

		• C1 Assessment of the modeling process, input, output and data included. • D1 Data review/verification/validation requirements were clarified

Certification

GWERD Quality Assurance Project Plan for

Surface Water Transport of Hydraulic Fracturing-Derived Waste Water

Revision #2

I certify that I have read, understand and will comply with the requirements set forth in this QAPP. I understand that my work for this project is subject to periodic audits and evaluation by EPA personnel.

Name:

Title:

Affiliation: