READ ME FILE
Approaches for Quantitative Use of Surface Water Monitoring Data in Pesticide Drinking Water Assessments

Description

EPA is presenting Approaches for Quantitative Use of Surface Water Monitoring Data in Pesticide Drinking Water Assessments to a FIFRA Scientific Advisory Panel on November 19-22, 2019. The primary documents (listed below) are available at https://www.regulations.gov in docket EPA-HQ-OPP-2019-0417.

- Approaches for Quantitative Use of Surface Water Monitoring Data in Pesticide Drinking Water Assessments (White Paper)
- Framework for Conducting Pesticide Drinking Water Assessments for Surface Water
- SEAWAVE-QEX Standard Operating Procedure
- Drinking Water Assessment Case Study Example for Acute Exposure Concern (Case Study 1)
- Drinking Water Assessment Case Study Example for Chronic and Cancer Exposure Concern (Case Study 2)

Access to SEAWAVE-QEX and associated documentation is available through USGS. There are several supplemental files that are provided in a single zipped file to facilitate access so that readers and reviewers can easily explore the supporting data and analysis for the White Paper (see Nov_2019_SAP_supplemental_documents.zip).

Download and Unzip Instructions

1. Hover over the file name, right-click the file link.
2. Save file to a local directory following displayed instructions.
3. To unzip all the contents of the zipped folder, right-click the zip file, select Extract All, and then follow the instructions.
4. When accessing additional zip files within the subfolders, step 3 will need to be repeated.
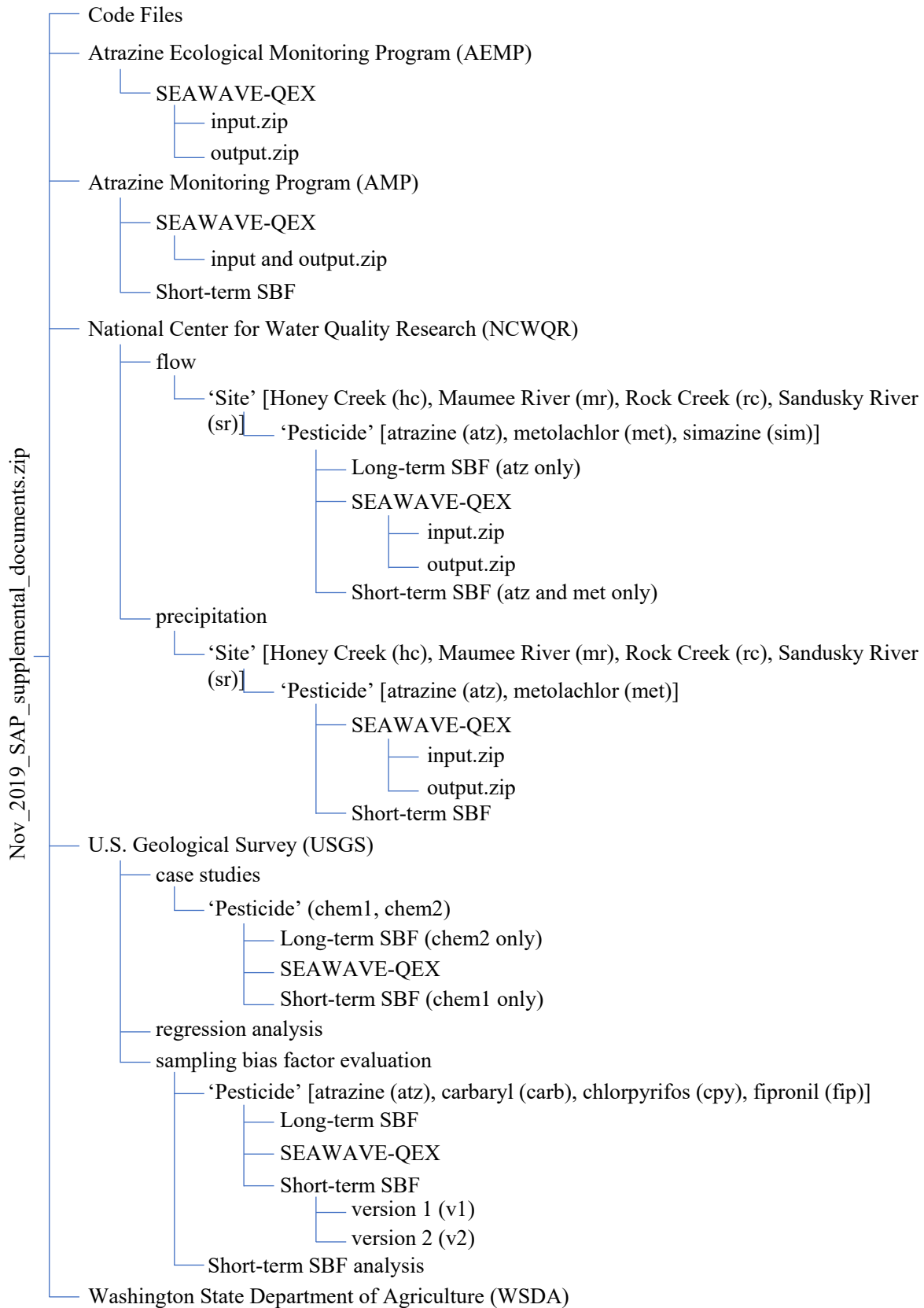
File Structure and Naming Convention

Within the .zip file, there are several individual folders and subfolders that organize the contents, as shown in the **Figure 1** and described below.

The first tier of folders is named for each of the five water monitoring datasets utilized in the White Paper and Case Studies, as well as a file for codes. The monitoring datasets include:

- Atrazine Ecological Monitoring Program (AEMP)
- Atrazine Monitoring Program (AMP)
- National Center for Water Quality Research (NCWQR)
- U.S. Geological Survey (USGS)
- Washington State Department of Agriculture (WSDA)

Disclaimer

Several code scripts developed by EPA are included in the supplemental documents. Code is provided on an "as is" basis and the user assumes responsibility for its use. EPA has relinquished control of the information and no longer has responsibility to protect its integrity, confidentiality, or availability.

Nov_2019_SAP_supplemental_documents.zip

- Code Files
- Atrazine Ecological Monitoring Program (AEMP)
  - SEAWAVE-QEX
    - input.zip
    - output.zip
- Atrazine Monitoring Program (AMP)
  - SEAWAVE-QEX
    - input and output.zip
  - Short-term SBF
- National Center for Water Quality Research (NCWQR)
  - flow
    - 'Site' [Honey Creek (hc), Maumee River (mr), Rock Creek (rc), Sandusky River (sr)]
      - 'Pesticide' [atrazine (atz), metolachlor (met), simazine (sim)]
        - Long-term SBF (atz only)
        - SEAWAVE-QEX
          - input.zip
          - output.zip
        - Short-term SBF (atz and met only)
  - precipitation
    - 'Site' [Honey Creek (hc), Maumee River (mr), Rock Creek (rc), Sandusky River (sr)]
      - 'Pesticide' [atrazine (atz), metolachlor (met)]
        - SEAWAVE-QEX
          - input.zip
          - output.zip
        - Short-term SBF
- U.S. Geological Survey (USGS)
  - case studies
    - 'Pesticide' (chem1, chem2)
      - Long-term SBF (chem2 only)
      - SEAWAVE-QEX
      - Short-term SBF (chem1 only)
  - regression analysis
  - sampling bias factor evaluation
    - 'Pesticide' [atrazine (atz), carbaryl (carb), chlorpyrifos (cpy), fipronil (fip)]
      - Long-term SBF
      - SEAWAVE-QEX
      - Short-term SBF
        - version 1 (v1)
        - version 2 (v2)
    - Short-term SBF analysis
- Washington State Department of Agriculture (WSDA)

Figure 1. Supplemental File Structure and Naming Convention

Code Files

This folder contains Python scripts for creating both short-term and long-term sampling bias factors, summarizing the output from each, and summarizing SEAWAVE-QEX output. Additionally, user guides are provided for each tool.

Atrazine Ecological Monitoring Program (AEMP)

These data were used to evaluate SEAWAVE-QEX in the White Paper (**Section 9.3**). This folder contains zipped input files and results files for SEAWAVE-QEX, named according to site ID (LA for LA-04 and TX for TX-01). The input files include precipitation for the site (dd.txt; transformed according to the SEAWAVE-QEX SOP), a sites ID file, and concentration files for each subsample replicate. The concentration files are named by site ID, subsampling strategy, and replicate number. For example, LA04rand13r1.txt is from LA-04, randomly sampled with 13 samples/year, and is replicate number 1 of 5. TX01strat14r5.txt is from TX-01, sampled 12 times per year with 14-day stratified sampling, and is replicate number 5 of 5. The zipped results folder contains the SEAWAVE-QEX output files from all runs, including conditional simulation data (CSIMS.txt), SEAWAVE-QEX model fitting parameters (demofitoutput.csv), diagnostic plots (Plots.pdf), and other chemograph summary statistics (STATS.txt). Other parts of file names are consistent with the concentration file nomenclature.

The AEMP pesticide concentration data used in this work is available through EPA Office of Pesticide Programs or the public docket EPA-HQ-OPP-2003-0367. Daily streamflow data was provided to EPA separately.

Atrazine Monitoring Program (AMP)

These data were used to evaluate SEAWAVE-QEX and sampling bias factors (SBFs) in the White Paper (**Section 9.5 and Section 10.1.3**). This folder contains zipped input files and output files for SEAWAVE-QEX; each folder within this zipped folder is labeled by site IDs (**Table 9.3 of the White Paper**) and contains a concentration input file (input.txt), a precipitation file for that site (dd.txt), conditional simulation output data (CSIMS.txt), diagnostic plots (Plots.pdf), and other chemograph summary statistics (STATS.txt). Each file name also contains the site ID. The short-term SBFs that were developed on these data are contained in a single comma separated file in the Short-term SBF folder. These were developed based on the chemographs from the final released version of SEAWAVE-QEX (denoted v2 for version 2 in the file name) and using the random log-linear imputation technique in the short-term sampling bias factor program.

The AMP pesticide concentration data used in this work is available through EPA Office of Pesticide Programs or the public docket EPA-HQ-OPP-2003-0367.

National Center for Water Quality Research (NCWQR)

These data were used to evaluate SEAWAVE-QEX and SBFs in the White Paper (**Section 9.3, Section 9.4, Section 10.1, and Section 10.2**). An Excel file is provided in the first folder that provides a list of final subsample datasets for each site and covariate combination (KEY_Final datasets_full and partial.xlsx). All results for full year and partial year restriction SEAWAVE-QEX runs are provided for comparison; the key provides the list of the final runs used in the analysis presented in the White Paper. The sites input file for all NCWQR SEAWAVE-QEX runs is also provided in the first folder (heidsites.txt).

The data are separated by the covariate that was used in SEAWAVE-QEX (flow or model-derived precipitation). Each of these contains folders for the four NCWQR sites: Honey Creek (hc), Maumee River (mr), Rock Creek (rc), and Sandusky River (sr). Within each site are folders for the three pesticides evaluated with these data: atrazine (atz), metolachlor (met), and simazine (sim). Simazine was only evaluated with flow and contains only SEAWAVE-QEX input and output files for the full record dataset.

Atrazine and metolachlor have SEAWAVE-QEX input and output files for the full record and subsampled datasets, and also have short-term SBFs. Input files include concentrations files (.txt) and precipitation data (dd.txt), as appropriate. The concentration files are named by site, chemical, subsampling frequency, and subsample number. For example, HoneyATZ7_1 is the first of five replicates of atrazine at Honey Creek, using a 7-day subsampling strategy and flow as a covariate. When model-derived precipitation is used as a covariate, ptot is added to the front of the file name (i.e., ptotHoneyATZ7_1). Full record samples are named by the convention Pesticide10th Site. For example, Atrazine10th Honey is the full record data for atrazine at Honey Creek. Included in the input files are also images saved while using SEAWAVE-QEX, including the measured concentration observations image (obs), the mid-term flow anomaly (mtfa), and short-term flow anomaly (stfa). These are also named by the same conventions; note that the flow anomalies will be the same for all samples at each site for the same years and so one example is provided. SEAWAVE-QEX output files include conditional simulation data (CSIMS.txt), SEAWAVE-QEX model fitting parameters (demofitoutput.csv), diagnostic plots (Plots.pdf), and other chemograph summary statistics (STATS.txt). Other parts of file names are consistent with the concentration file nomenclature.

Short-term SBFs are provided for three sampling strategies within the short-term sampling bias factor program, as denoted in the file name and described in **Section 4.4.1.3** of the White Paper. These are random, stratified_yes (referred to as stratified in the White Paper), and stratified_no (referred to as stratified-random in the White Paper). In addition to short-term SBFs, long-term SBFs were calculated for atrazine at all sites using flow as a covariate.

The pesticide concentration data used in this work was obtained via personal communication with the NCWQR laboratory (https://ncwqr.org/) and an accompanying quality assurance plan was provided to EPA.

U.S. Geological Survey (USGS)
These data were used to evaluate SBFs and develop watershed regression equations in the White Paper (**Section 10.1, Section 10.2, Section 11.3, and Section 11.4**). They were also used in the two case studies.

The case study folders are labeled by chemical name (chem1, chem2) and each contain subfolders for SBFs and SEAWAVE-QEX. The SEAWAVE-QEX folder contains concentration input files (conc.txt or Chemdata.txt), site input files (sites.txt), conditional simulation output data (CSIMS.txt), diagnostic plots (Plots.pdf), and other chemograph summary statistics (STATS.txt), named by USGS gage station number. Note that there are multiple site files for chem1. Chem1 includes a folder with short-term SBF results and chem2 contains a folder with long-term SBF results. The short-term SBF folder includes one file (chem1_stbf_reslts_random_log_linear_no_csv) with all SBFs and two summary files of these results: one (sbfa_site_chem1.csv) that contains the summary statistics for each year and site across all of the realizations and another (sbfa_summary_chem1.csv) that further summarizes across the years of simulation. For the long-term SBFs, there are two files with the raw results using random and stratified sampling in the SBF generator (sbfc_results_stratified chem2.csv; sbfc_results_random chem2.csv). These files are similarly summarized for each year and site across all of the realizations (sbfc_sites rand chem2.csv; sbfc_sites strat chem2.csv) and further summarized across years of the simulation (sbfc_summary rand chem 2.csv; sbfc_summary strat chem2.csv).

The watershed regression folder contains several files relevant to that analysis. This includes SAS code to evaluate correlation among watershed and catchment variables for all USGS sites included in the analysis (Watershed_var_correlation.sas), the corresponding output file from this code (Watershed_var_correlation.pdf), and a text file of watershed and catchment data for the evaluated USGS sites (usgs_sites_122018-clean final03.21.19.txt). There is also SAS code to evaluate regression models

for each of the four chemicals, which includes steps and notes for excluding or including outliers and potential regressor variables (SBF-reg-evaluate models.sas) and an output file from this code (SBF-reg-evaluate models.pdf). The SAS code for the final regression models for each of the four pesticides is provided (SBF-reg-final models.sas), as well as the output file from this code (SBF-reg-final models.pdf) and a file of watershed and catchment data for sites included in the final regression analysis that includes site summary SBFs for 14-day sampling intervals with random sampling and log-linear interpolation (SBF_reg_input_data.csv).

In the sampling bias factor evaluation folder, each of the four pesticides used to develop SBFs in the White Paper (atrazine, carbaryl, chlorpyrifos, and fipronil) is abbreviated (atz, carb, cpy, and fip). Each pesticide folder contains a separate folder for data relating to SEAWAVE-QEX, short-term sampling bias factors, and long-term sampling bias factors.

The SEAWAVE-QEX folder contains conditional simulation output data (CSIMS.txt), diagnostic plots (Plots.pdf), SEAWAVE-QEX model fitting parameters (fitoutput.csv), and other chemograph summary statistics (STATS.txt), named by USGS gage station number.

The short-term SBFs are divided into two folders based on the version of SEAWAVE-QEX that was used during their development. Version 1 (v1) of SEAWAVE-QEX is a beta version of the model that produced different diagnostic plots than the final release version; the SBFs developed from this version of SEAWAVE-QEX were used only to optimize parameters in the SBF program (i.e., imputation method and sampling strategy) for developing SBFs with version 2. Version 2 (v2) is the publicly available version of SEAWAVE-QEX; these SBFs were used in the rest of the White Paper and case studies.

The long-term SBFs are provided in three files. The first file (results.csv) provides all SBFs across the 100 SEAWAVE-QEX chemographs and all years for each site; the second file (sites.csv) provides summary statistics for each year at each site (collapsing across SEAWAVE-QEX realizations); the third file (summary.csv) provides the final summary of SBFs for each site (collapsing across years). Distinctions between random or stratified sampling in the development of these SBFs is made in the file name.

USGS pesticide concentration data used in this work was obtained through the SEAWAVE-QEX data release (Data Files to Support SEAWAVE-QEX Model for Simulating Concentrations of Selected Pesticides in the Continental United States, 1992–2012).

Additionally, within the sampling bias factor evaluation folder is a folder containing files relevant to the short-term sampling bias factor optimization and analysis. There are nine data files of atrazine short-term bias factors with random, stratified (indicated with "stratified" and "yes" in the file name), or stratified-random (indicated with "stratified" and "no" in the file name) sampling and one of three infilling techniques (linear, log-linear, stair). The files are named similar to the following: "FINAL_bf_results_**random/stratified_linear/stair/log_linear_yes/no**". There is also a file (FINAL_rmse_atr_041519.sas) of SAS code that imports those nine files, merges and sorts data, removes data as needed (e.g., sites without a full year of values), and conducts a Generalized Linear Mixed Model (GLMM) analysis with a lognormal error distribution on the 50th percentile of the root mean square error (RMSE). SAS is required to run this file. Path names that are provided in the code to the data files will need to be updated to reflect their current locations when the program is run on a different computer. Output from the SAS code is provided in the file named FINAL_rmse_atr_041519output.docx. Three additional files are provided with short-term sampling bias factor summary data supporting the White Paper for the four pesticides (atrazine, carbaryl, fipronil, and chlorpyrifos) using random sampling and log-linear infilling. One file (BFdata_all_chem_siteyr.csv) summarizes SBFs over realizations within each site and year combination including the mean, skewness, kurtosis, and the median for the 1-day, 4-

day, and 21-day SBF and for the 50th percentile of the RMSE calculated by the SBF program. This file includes some sites that was ultimately removed due to insufficient data. A second file (BFdata_all_chem_site.csv) further reduces the previous file by providing median SBFs over years for each site and removing sites with insufficient data and a third file (FINAL_chem_summary.csv) gives a summary of SBFs across sites including the median, 25th percentile, 75th percentile, mean, and standard deviation for the 1-day, 4-day, and 21-day SBFs.

Washington State Department of Agriculture (WSDA)
These data were used to evaluate SEAWAVE-QEX in the White Paper (**Section 4.6**). The SEAWAVE-QEX input and output files are contained within this folder for atrazine at the site Sulphur Creek Wasteway and metolachlor at Bertrand Creek. These sites are named SU-1 and BC-1 in the file names, respectively. The file names are also labeled according to the covariate used to develop the file: streamflow (flow), stream stage (stage), or WSDA-developed precipitation based on meteorological data (precip). The input files include concentration files for each site (conc.txt) and covariate data (dd.txt). The output files include conditional simulation output data (CSIMS.txt), diagnostic plots (Plots.pdf), and other chemograph summary statistics (STATS.txt).

The pesticide and precipitation data used in this work was obtained via personal communication with WSDA and an accompanying quality assurance plan was provided to EPA.