FEB 1 8 2020

OFFICE OF
CHEMICAL SAFETY AND
POLLUTION PREVENTION

Memorandum

SUBJECT: Transmittal of Meeting Minutes and Final Report for the Federal Insecticide, Fungicide and Rodenticide Act, Scientific Advisory Panel (FIFRA SAP) Meeting held on November 19-21, 2019

TO: Richard Keigwin
Director
Office of Pesticide Programs

FROM: Tamue Gibson, MS
Designated Federal Official, FIFRA SAP Scientific Staff
Office of Science Coordination and Policy

THRU: Steven Knott, MS
Executive Secretary, FIFRA SAP
Office of Science Coordination and Policy

Hayley Hughes, DrPH, MPH, CSP
Director
Office of Science Coordination and Policy

Please find attached the meeting minutes for the FIFRA Scientific Advisory Panel open public meeting held in Arlington, Virginia on November 19-21, 2019. This report addresses a set of scientific issues being considered by the U.S. Environmental Protection Agency regarding the Office of Pesticide Programs' Approaches for Quantitative Use of Surface Water Monitoring Data in Pesticide Drinking Water Assessments.

Attachment

cc:
Alexandra Dunn
David Fischer
Cheryl Dunton
Richard Keigwin
Anna Lowit
Marietta Echeverria
Dana Spatz
OPP Docket

FIFRA Scientific Advisory Panel Members

Robert E. Chapin, PhD
George B. Corcoran, PhD
Sonya K. Sobrian, PhD
Clifford P. Weisel, PhD
Raymond S.H. Yang, PhD

FQPA Science Review Board Members

Claire Baffaut, PhD
Veronica Berrocal, PhD
Terry Councell, MS
Timothy Green, PhD
Ian Kennedy, PhD
Rebecca Klaper, PhD
Andrew Miglino, PhD
Lisa Nowell, PhD
Thomas Potter, PhD
Kenneth Portier, PhD (Retired)
John Rodgers, Jr., PhD
James Sadd, PhD
Xuyang Zhang, PhD

FIFRA Scientific Advisory Panel
Meeting Minutes and Final Report
No. 2020-01

Peer Review of the Approaches for Quantitative Use of Surface Water Monitoring Data in Pesticide Drinking Water Assessments

November 19-21, 2019
FIFRA Scientific Advisory Panel Meeting

Held at

Holiday Inn Rosslyn, At Key Bridge
1900 Fort Myer Drive
Arlington, Virginia 22209

1

NOTICE

The Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA), Scientific Advisory Panel (SAP) is a Federal advisory committee operating in accordance with the Federal Advisory Committee Act (FACA) and established under the provisions of FIFRA as amended by the Food Quality Protection Act (FQPA) of 1996. The FIFRA SAP provides advice, information, and recommendations to the U.S. Environmental Protection Agency (EPA or Agency) Administrator on pesticides and pesticide-related issues regarding the impact of regulatory actions on health and the environment. The SAP serves as a primary scientific peer review mechanism of the EPA, Office of Pesticide Programs (OPP), and is structured to provide balanced expert assessment of pesticide and pesticide-related matters facing the Agency. The FQPA Science Review Board members serve the FIFRA SAP on an *ad hoc* basis to assist in reviews conducted by the FIFRA SAP. The meeting minutes and final report are provided as part of the activities of the FIFRA SAP.

The FIFRA SAP carefully considered all information provided and presented by the Agency, as well as information presented by the public. The minutes represent the views and recommendations of the FIFRA SAP and do not necessarily represent the views and policies of the Agency, nor of other agencies in the Executive Branch of the federal government. Mention of trade names or commercial products does not constitute an endorsement or recommendation for use.

The meeting minutes and final report do not create nor confer legal rights nor impose legally binding requirements on the EPA or any other party. The meeting minutes and final report of the November 19-21, 2019, FIFRA SAP meeting represent the SAP's consideration and review of scientific issues associated with "The Approaches for Quantitative Use of Surface Water Monitoring Data in Pesticide Drinking Water Assessments." Steven Knott, MS, FIFRA SAP Executive Secretary, reviewed the minutes and final report. Robert E. Chapin, PhD, FIFRA SAP Chair, and Tamue Gibson, MS, FIFRA SAP Designated Federal Official, certified the minutes and final report, which is publicly available on the SAP website http://www.epa.gov/sap under the heading of "Meetings" and in the public e-docket, Docket No. EPA-HQ-OPP-2019-0417, accessible through the docket portal: http://www.regulations.gov. Further information about FIFRA SAP reports and activities can be obtained from its website at http://www.epa.gov/sap. Interested persons are invited to contact Tamue L. Gibson, MS, SAP Designated Federal Official, via e-mail at gibson.tamue@epa.gov.

TABLE OF CONTENTS

TABLE

FIFRA Scientific Advisory Panel
Meeting Minutes and Final Report
No. 2020-01

Peer Review of the Approaches for Quantitative Use of Surface Water Monitoring Data in Pesticide Drinking Water Assessments

November 19-21, 2019
FIFRA Scientific Advisory Panel Meeting

Held at

Holiday Inn Rosslyn, At Key Bridge
1900 Fort Myer Drive
Arlington, Virginia 22209

Robert E. Chapin, PhD
FIFRA SAP Chair
FIFRA Scientific Advisory Panel

RFChapin

Date: 18 Feb 2020

Tamue L. Gibson, MS
Designated Federal Official
Office of Science Coordination and Policy

Date: *February 18, 2020*

4

Federal Insecticide, Fungicide, and Rodenticide Act
Scientific Advisory Panel Meeting
November 19-21, 2019

Peer Review of the Approaches for Quantitative Use of Surface Water Monitoring Data in Pesticide Drinking Water Assessments

PARTICIPANTS

FIFRA SAP Chair

Robert E. Chapin, PhD, Independent Consultant,
Preston, Connecticut

Designated Federal Official

Tamue L. Gibson, MS
Biologist and Designated Federal Official, Office of Science Coordination and Policy, U.S. Environmental Protection Agency, Washington, District of Columbia, Tel: (202) 564-7642, Email:gibson.tamue@epa.gov

FIFRA Scientific Advisory Panel Members

George B. Corcoran, PhD, Chairman and Professor, Department of Pharmaceutical Sciences, Wayne State University, Detroit, Michigan

Sonya K. Sobrian, PhD, Associate Professor, Howard University College of Medicine, Howard University, Washington, District of Columbia

Clifford P. Weisel, PhD, Professor, Rutgers University,
Piscataway, New Jersey

Raymond S.H. Yang, PhD, Professor (Emeritus), College of Veterinary Medicine and Biomedical Sciences, Colorado State University, Fort Collins, Colorado

FQPA Science Review Board Members

Claire Baffaut, PhD, Research Hydrologist, United States Department of Agriculture, Agricultural Research Service, Cropping System and Water Quality Research Unit, Columbia, Missouri

Veronica Berrocal, PhD, Associate Professor, Department of Statistics, University of California, Irvine, California

Terry Councell, MS, Coordinator, Total Diet Study, United States Food and Drug Administration, College Park, Maryland

Timothy Green, PhD, Research Agricultural Engineer/Hydrologist at the United States Department of Agriculture, Agricultural Research Service, Center for Agricultural Resources Research, Fort Collins, Colorado

Ian Kennedy, PhD, Senior Evaluation Officer, Health Canada, Pest Management Regulatory Agency, Ottawa, Ontario, Canada

Rebecca Klaper, PhD, Professor, School of Freshwater Sciences, University of Wisconsin-Milwaukee, Milwaukee, Wisconsin

Andrew Miglino, PhD, Physical Scientist, United States Food and Drug Administration, Center for Veterinary Drugs, Rockville, Maryland

Lisa Nowell, PhD, Research Chemist, United States Geological Survey, Sacramento, California

Thomas Potter, PhD, President and Principal Scientist, Innovative Systems for Education and Environmental Technology

Kenneth Portier, PhD, Consulting Biostatistician (formerly of the National Cancer Society), Athens, Georgia

John Rodgers, Jr., PhD, Professor of Aquatic Toxicology and Ecotoxicology, Clemson University, Clemson, South Carolina

James Sadd, PhD, Professor of Environmental Science, Department of Geology and Environmental Science, Occidental College, Los Angeles, California

Xuyang Zhang, PhD, Senior Environmental Scientist (Specialist), California Department of Pesticide Regulation, Environmental Monitoring Branch, Sacramento, California

LIST OF ACRONYMS AND ABBREVIATIONS

AEMP	Atrazine Ecological Monitoring Program
AMP	Atrazine Monitoring Program
APEX	Agricultural Policy/Environmental eXtender
AR	Autoregressive
CCI	Conductivity Claypan Index
CEAP	Conservation Effects Assessment Program
CPHEA	The Center for Public Health and Environmental Assessment
CS	Compound Symmetry
CTS	Correlation Time Scales
CWS	Community Water System
CWS DWI	Community Water System Drinking Water Intake
DT_{50}	Variability of substance degradation
DW	Drinking Water
DWA	Drinking Water Assessment
DWLOC	Drinking Water Level of Concern
DWS	Drinking Water System
EDWC	Estimated Drinking Water Concentration
EFED	Environmental Fate and Effects Division
EJSCREEN	Environmental Justice Screening and Mapping Tool
ELEXNET	Electronic Laboratory Exchange Network
EPA	Environmental Protection Agency
FACA	Federal Advisory Committee Act
FDA	Food and Drug Administration

FERN	Food Emergency Response Network
FFDCA	Federal Food, Drug and Cosmetics Act
FIFRA	Federal Insecticide, Fungicide, and Rodenticide Act
FQPA	Food Quality Protection Act
GIS	Geographic Information System
GLMM	Generalized Linear Mixed Model
HA	Health Advisories
HUC	Hydrologic Unit Code
K_{oc}	Sorption
LAMPS	Land Use and Agricultural Management Practices Web-Service
MCL	Maximum Contaminant Level
MCMC	Markov Chain Monte Carlo
MODIS	Moderate Resolution Imaging Spectroradiometer
MTFA	Mid-Term Flow Anomaly
NASQAN	National Stream Quality Accounting Network
NASS	National Agricultural Statistics Service Information
NAWQA	National Water-Quality Assessment
NCWQR	National Center for Water Quality Research
NLCD	National Land Cover Database
NRCS	National Resources Conservation Service
NWALT	Wall-to-Wall Anthropogenic Land Use Trends
NWIS	National Water Information System
OCSPP	Office of Chemical Safety and Pollution Prevention
OPP	Office of Pesticide Programs
PC	Principal Component

PCA	Principal Component Analysis
PWC	Pesticide in Water Calculator
QA	Quality Assurance
QC	Quality Control
RMSE	Root Mean Square Error
SAM	Spatial Aquatic Model
SAP	Scientific Advisory Panel
SAS	Statistical Analysis System
SBF	Sampling Bias Factor
SEAWAVE-QEX	Seasonal Wave with Streamflow Adjustment with Extended Capability
SOP	Standard Operating Procedure
SSD	Seasonal Standard Deviation
SSURGO	Soil Survey Geographic Database
STATSGO	State Soil Geographic Database
STFA	Short-Term Flow Anomaly
SVI	Soil Vulnerability Index
SWAT	Soil and Water Assessment Tool
USDA	United States Department of Agriculture
USGS	U.S. Geological Survey
WARP	Watershed Regressions for Pesticides
WSDA	Washington State Department of Agriculture
WQP	Water Quality Portal

INTRODUCTION

The Federal Insecticide, Fungicide, and Rodenticide Act Scientific Advisory Panel completed its review of the set of scientific issues being considered by the Environmental Protection Agency regarding the Approaches for Quantitative Use of Surface Water Monitoring Data in Pesticide Drinking Water Assessments (DWAs). Advanced notice of the meeting was published in the Federal Register on August 13, 2019. The review was conducted in an open Panel meeting held in Arlington, Virginia, on November 19-21, 2019. The Agency position paper, charge questions, and related documents in support of the SAP meeting are posted in the public e-docket at http://www.regulations.gov (ID: EPA-HQ-OPP-2019-0417). Robert E. Chapin, PhD, chaired the meeting. Tamue L. Gibson, MS, served as the Designated Federal Official.

In preparing these meeting minutes and final report, the Panel carefully considered all information provided and presented by the Agency presenters, as well as information presented by public commenters. The meeting minutes and final report address the information provided and presented at the meeting, especially the Panel response to the Agency charge.

U.S. EPA presentations were provided during the FIFRA SAP meeting by the following (listed in order of presentation):

Introduction and Welcome – Marietta Echeverria, MS, Director, Office of Chemical Safety and Pollution Prevention (OCSPP)/Office of Pesticide Programs (OPP)/Environmental Fate and Effects Division (EFED)

OPP Technical Presentation – Overview of the Approaches for Quantitative Use of Surface Water Monitoring Data in Drinking Water Assessments – Dana Spatz, MS, Branch Chief, OCSPP/OPP/EFED

Framework for Conducting Pesticide Drinking Water Assessments for Surface Water – Rochelle Bohaty, PhD, Senior Chemist, OCSPP/OPP/EFED

Analyzing and Interpreting Surface Water Pesticide Monitoring Data – Rochelle Bohaty, PhD, Senior Chemist, OCSPP/OPP/EFED

Evaluation of SEAWAVE-QEX as an Imputation Technique for Estimating Daily Pesticide Concentrations from Pesticide Monitoring Data Part 1 – Sarah Hafner, PhD, Chemist, OCSPP/OPP/EFED; Rochelle Bohaty, PhD, Senior Chemist, OCSPP/OPP/EFED; Matthew Bischof, MS, Natural Resource Scientist, Washington State Department of Agriculture

Evaluation of SEAWAVE-QEX as an Imputation Technique for Estimating Daily Pesticide Concentrations from Pesticide Monitoring Data Part 2 – Sarah Hafner, PhD, Chemist, OCSPP/OPP/EFED; Rochelle Bohaty, PhD, Senior Chemist, OCSPP/OPP/EFED; Matthew Bischof, MS, Natural Resource Scientist, Washington State Department of Agriculture

Development and Evaluation of a Sampling Bias Factor Program Part 1: Short-term – Christine Hartless, PhD, Biologist, OCSPP/OPP/EFED; Rochelle Bohaty, PhD, Senior Chemist, OCSPP/OPP/EFED

Development and Evaluation of a Sampling Bias Factor Program Part 2: Long-term – Charles Peck, MS, Senior Environmental Engineer, OCSPP/OPP/EFED; Rochelle Bohaty, PhD, Senior Chemist, OCSPP/OPP/EFED

Watershed Extrapolation and Weight-of-Evidence Approach – Christine Hartless, PhD, Biologist, OCSPP/OPP/EFED; Charles Peck, MS, Senior Environmental Engineer, OCSPP/OPP/EFED James Hook, MS, Ecologist, OCSPP/OPP/EFED

Drinking Water Assessment Case Study – **A Pesticide with Short-term Exposure Consideration** – Katrina White, PhD, Senior Biologist, OCSPP/OPP/EFED

Drinking Water Assessment Case Study – **A Pesticide with Long-term Exposure Considerations** – Jessica Joyce, MS, Physical Scientist, OCSPP/OPP/EFED; Charles Peck, MS, Senior Environmental Engineer, OCSPP/OPP/EFED

Development of Surface Water Monitoring Program – Matthew Bischof, MS, Natural Resource Scientist, Washington State Department of Agriculture; Sarah Hafner, PhD, Chemist, OCSPP/OPP/EFED; Rochelle Bohaty, PhD, Senior Chemist, OCSPP/OPP/EFED

PUBLIC COMMENTERS

Oral statements were presented by:

1) Daniel Perkins, PhD, Senior Consultant, Waterborne Environmental, Incorporated, Champaign, Illinois

2) Paul Mosquin, PhD, Statistician, RTI International, Research Triangle Park, North Carolina

3) Jeremey Aldworth, PhD, Statistician, RTI International, Research Triangle Park, North Carolina

4) Manojit Basu, PhD, Managing Director, Science Policy, Crop Life America

Written statements were provided by:

1) On behalf of CropLife America
   a. Manojit Basu, PhD, Managing Director, Science Policy, Crop Life America

EXECUTIVE SUMMARY

The United States Environmental Protection Agency's (EPA) Office of Pesticide Programs (OPP) uses aquatic exposure models to estimate potential pesticide concentrations in drinking water sources across the country. Approximately 87% of the total population or 282 million people receive drinking water via public-supply systems. Surface water sources of drinking water include rivers, streams, lakes, and impoundments such as reservoirs. Available surface water monitoring data for most pesticides have not been considered robust enough for direct use in pesticide exposure assessments due to infrequent sampling and lack of coverage across the landscape. The models, associated concepts, and assumptions have been vetted through various FIFRA Scientific Advisory Panels (SAPs) and independent groups, as well as in the open literature, and practices continue to evolve. In summary, the previous FIFRA SAPs recommended that EPA:

- Use a modified version of SEAsonalWAVEQ with EXtended capabilities model (SEAWAVE-QEX) to estimate upper-end pesticide concentrations that may occur between sampling events;
- Use more data (i.e., more pesticides and more sites) in the development of Sampling Bias Factors (SBFs); and
- Determine if the variance in the distribution of SBFs can be explained by site watershed characteristics.

The EPA has been working for several years to develop a method to evaluate existing surface water monitoring data for quantitative use in pesticide DWAs. Quantitative use of existing surface water monitoring data in DWAs has been limited by temporal and spatial challenges. That is, effectively capturing peak concentrations and pesticide occurrence in water is highly site-specific due to variable pesticide use patterns, and soil and weather conditions. The EPA has long desired to utilize available surface water monitoring data in pesticide DWAs; however, the data have not been sufficient, and the tools available to assist in understanding the data were not robust enough.

Instead, the EPA has relied largely on aquatic exposure models to estimate pesticide concentrations across time and space. As the desire to integrate surface water monitoring data in DWAs remains and stakeholders continue to criticize the EPA DWAs for not incorporating available surface water monitoring data, the EPA has more recently taken a different approach. Rather than designing a resource-intensive monitoring program to directly measure drinking water exposure, the EPA is looking for approaches that make the best use of the available data. Specifically, EPA would like to be able to reliably estimate, using available monitoring data, protective pesticide concentrations in drinking water at sites with infrequent or no available monitoring data.

The Agency sought advice and recommendations from the FIFRA SAP on scientific issues associated with the evaluation and utility of SEAWAVE-QEX model capabilities, long-term SBFs, watershed regressions, and optimization of short-term SBFs.

The FIFRA SAP addressed four charge questions and provided the following overall summary of the major conclusions and recommendations that are detailed in the report.

Charge Question 1. – ***SEAWAVE-QEX***

The Panel complimented EPA for advancing computational technology in the Agency's scientific and regulatory processes; in this case specifically using the SEAWAVE-QEX for estimating pesticide residues in surface water for DWAs. In discussing the strengths and weaknesses of the SEAWAVE-QEX model, various members of the Panel acknowledged not only the difficulty of the problem that the Agency is attempting to address, but also the elegance and relative simplicity of SEAWAVE-QEX. Some of the model's identified strengths include; 1) SEAWAVE-QEX logically deconstructs a daily pesticide concentration time series into components describing major sources of variation (seasonality, streamflow variability, and long-term trends); 2) allows, through conditional simulations, the user to reproduce the statistical characteristics of daily pesticide concentration over a time period; 3) SEAWAVE-QEX performs well in describing near-complete monitoring data; 4) SEAWAVE-QEX's function (a pulse of inputs) represents the way producers treat fields in a watershed for a given crop; 5) by using flow as a covariate, SEAWAVE-QEX takes into account the hydrologic landscape processes that affect pesticide transport, (i.e., infiltration and runoff generation or lack thereof).

Some of the model's identified weaknesses include: a seasonal wave component for compounds that do not have a clear seasonal pattern of use; it is difficult to assess how much of an improvement using SEAWAVE-QEX makes when one compares estimated maximum concentrations to observed maximum concentrations without knowing the pesticide water calculator (PWC); and finally, running the SEAWAVE-QEX model has a steep learning curve, it is time and resource intensive, and it is difficult to understand by the uninitiated.

Recommendations:
- Consider using eco-hydrologically similar watersheds with a stream gauge or hydrologic models to estimate missing flow values.
- Assess the accuracy of SEAWAVE-QEX for pesticides that have multiple potential windows of use including assessing whether SEAWAVE-QEX has a waveform that would fit multi-application scenarios.
- Compare SEAWAVE-QEX results with those obtained with the PWC.
- Evaluate SEAWAVE-QEX against measured daily or near-daily concentrations across a range of watershed and pesticide biophysical characteristics.
- Consider a Bayesian formulation of the SEAWAVE-QEX model.

Regarding evaluation of the SEAWAVE-QEX model using the subsampled data, the Panel raised several concerns. The Panel noted how the number of replicates considered by the Agency – 5 subsamples – is too small to really assess sample-to-sample variability in the estimates. Several Panel members noted the lack of criteria to determine if sampling makes a large difference or to determine the best sampling strategy. It was unclear to the Panel whether the SEAWAVE-QEX model requires 12 pesticide concentration measurements per year or 12 measurements per seasonal wave.

Recommendations:
- Consider repeating the analysis and increasing the number of subsamples to at least 30, in order to better distinguish between within-sample and between-sample variability in the estimated maximum concentration.

- Set threshold values for under and overestimation frequency (i.e., how often is the concentration under or overestimated) and bias (i.e., by how much is it under or overestimated) as a function of the use of the model.
- Clarify explicitly the requirements in terms of the number of pesticide concentration measurements needed per seasonal wave and per year.

Charge Question 2. – *SAMPLING BIAS FACTORS*

Sampling bias factors are multiplicative factors used to calculate an upper bound prediction interval (i.e., 95th percentile) for a given measured concentration value to account for a less than daily sampling frequency which might miss the concentration peaks. Sampling bias factors are envisioned as an alternative approach to estimating drinking water concentrations when available monitoring data are inadequate for use in SEAWAVE-QEX, because measured data are not available or are too highly censored. Use of SEAWAVE-QEX in SBF computation has addressed a lingering problem in SBF use, i.e., estimated SBFs may be highly uncertain, especially for sites that are infrequently sampled.

The Panel agreed that the use of SEAWAVE-QEX to compute the SBFs for the four reference pesticides (atrazine, carbaryl, chlorpyrifos, and fipronil) was technically sound and could be acceptable as a screen but did not recommend application of these SBFs to other pesticides without further analysis. Panel members agreed that four reference pesticides were too few to draw conclusions on the role of physical/chemical properties or pesticide class and recommended developing SBFs for additional pesticides to better estimate the magnitude of uncertainties in this approach. To select SBFs for an individual pesticide, the Panel suggested that EPA consider pesticide class, physical/chemical properties, use profile, and factors reflecting vulnerability of treated land (such as dominant hydrologic soil group, slope, precipitation, and presence of sub-surface artificial drainage) in addition to those listed in the EPA white paper (table 11.2, page 201; EPA 2019). The Panel noted that the data-intensive sites that were used to determine the SBFs for reference pesticides were all located in the Midwestern U.S., where the setting is agricultural land, the principal crops are corn, wheat, and soybeans, and tile drains are common. This suggests uncertainty as to whether the maximum SBF values for the four referenced pesticides determined from these Midwestern sites would be protective for streams in other environmental settings. The Panel cautioned that further evaluation was needed for application of SEAWAVE-QEX and SBFs to small watersheds with "flashy' streams or to static, non-flowing systems.

Charge Question 3**. – *SPATIAL RELEVANCE OF MONITORING DATA AND SAMPLING BIAS FACTORS WITH WATERSHED AND PESTICIDE CHARACTERISTICS***

The Panel agreed that EPA's proposed use of a weight-of-evidence approach was a useful approach for determining the spatial relevance of monitoring sites to source drinking water analysis, providing relevant monitoring data or SBFs to areas lacking them, aiding in the search for appropriate surrogates, and evaluating the possible utility of nearby monitoring detects at values of concern. Panel members raised a number of important issues to resolve when implementing this approach.

The determination of spatial relevance of monitoring sites to source drinking water could be operationalized in a number of ways - such as clear guidance through a decision tree or an (ArcGIS) ModelBuilder ® model that can be used to compare a given watershed to acceptable SBF watersheds, making this process simpler, standardized and less subject to error from subjective decision making.

In the short-term, EPA needs to define terms so that they can be used consistently and be specific about the requirement for "a site that is spatially relevant" beyond what is currently in the EPA white paper. The reasonableness or utility of this requirement should be demonstrated by analysis or at least a clear rationale.

Recommendations:
- As an intermediate goal, EPA can ease implementation of this approach by arranging for geospatial data that risk assessors and others can easily access.
- Further, The Land Use and Agricultural Management Practices web-Service (LAMPS) can be explored as an option to enhance the weight-of-evidence approach while new analytic tools are being developed.

The Panel agreed that it made sense to relate SBFs to watershed attributes, based on the performance of the regression models. The Panel also agreed with the Agency's conclusion that the short-term SBF regression equation for the four evaluated pesticides has relatively poor predictive ability and, thus should not be used for drinking water evaluation without further improvement. The Panel recommended alternate datasets and statistical approaches that may produce better predictions of SBF variability.

While stepwise regression is a traditional choice for building a predictive model, the Panel believed that more recent approaches and, in particular, regression tree methods (bagging, boosting and random forests), could be employed and produce more insightful results.
The Panel proposed the following varying alternative approaches:

Regression trees: Regression trees are predictive models that predict the value of the outcome variable by splitting several input variables.

Median regression: The linear regression models proposed in the EPA white paper are all applied to predict the median of the SBFs derived at each site. However, it was discussed that for the linear regression models that are used to estimate the mean of a distribution, specifically in the case of skewed distributions, that the median and the mean do not coincide.

Principal component analysis: The Panel suggested that, to improve the predictive power of the linear regression models for SBFs, the Agency could consider applying principal components analysis (or exploratory factor analysis) to the predictors to generate new factors that could be better predictors than the individual watershed properties.

Change the outcome variable: The Panel recommended that the Agency predict the numerator of the SBF ratio with the denominator of the SBF being itself a covariate used in the model along with all the other watershed attributes. Thus, the Agency should attempt to predict the true max as a function of the median-of-the-median value.

The following is a listing of predictors that EPA should use in a regression modeling approach.

Pesticide usage data: Pesticide usage is perhaps the most important variable in determining pesticide concentrations in surface water. Adding this variable could improve the regression analysis. EPA could potentially use the U.S. Geological Survey (USGS) estimated county-level usage data.

Precipitation: Precipitation is another important parameter of pesticide concentration. The Panel recommended that future analyses should use annual precipitation data instead of a 30-year average.

Climate data: A Panel member suggested that the Agency perform the analysis to predict the median SBF for a given pesticide in any year rather than summarizing the SBFs across time at each site, and multiple Panel members advised against the use of climate variables averaged across a 30-year time period as predictors in a regression modeling framework, especially given variations in climate data over time.

Alignment of timing between independent variables and SBFs: The Panel agreed with EPA that future work should try to align the timing of the pesticide sampling with that of the independent variables (pesticide application).

Charge Question 4. – *DRINKING WATER ASSESSMENT CASE STUDIES*

The Panel agreed that the Framework was generally clear, well organized, appropriately tiered, with the most conservative assessments conducted at the lowest tiers (1 and 2), and similar to approaches used in other well-established systems of pesticide risk assessment. The *Introduction* is especially effective, informative, and strategic in defining principles and preparing the reader to understand the context of subsequent sections. Panel members agreed that, after EPA's determination on the scope of use for SBFs, the SEAWAVE-QEX model and weight-of-evidence approach, the Framework will need to be updated to describe the use of these tools under the appropriate tier.

Panel members made several recommendations to improve the DW Framework including the following:
- Clarify the intended target audience(s) for the Framework document (EPA staff, stakeholders, general public) and adjust language accordingly.
- Clarify the conceptual model to show explicit decision criteria and the information required for each decision level.
- Provide links between durations, exposures and effects for each duration in the models (1, 4, 21, 365-days, and 30 years).
- Presently, the Framework emphasizes the potential exposure information (i.e., monitoring data and modeling estimates), while paying relatively little attention to the accuracy of the risk comparator or the effects data (i.e., Drinking Water Level of Concern (DWLOC), Maximum Contaminant Level (MCL), and Health Advisories (HAs)), even though these also contribute to uncertainty in the risk decision.
- Provide more information on how the PWC is used in lower tier assessments.

- Label rates: during the input process, adjust the recommended pesticide label rates so that the maximum allowed rates match the typical use rates used during the evaluation.
- Consider use of additional models in tiers 3 or 4 (i.e., United States Department of Agriculture's (USDA) Agricultural Policy/Environmental eXtender (APEX) or Soil and Water Assessment Tool (SWAT) models).
- The acronym for DWLOC is utilized to represent two different parameters one for Drinking Water Level of <u>Comparison</u> and Drinking Water Level of <u>Concern</u>. This acronym should be avoided in both representations or pick one and change the other.
- The limitations of the SEAWAVE-QEX model should be clearly expressed, including its limited use in non-flowing waters, which are important sources for DW in many agricultural watersheds; variance in the estimated or extrapolated maximum pesticide concentrations that are compared in deterministic fashion to DWLOCs; and the need for explicit statements of underlying assumptions.
- Because the evaluation of environmental fate and degradation is done conservatively, the modeling should generally overpredict environmental concentrations. Refining the evaluation of the environmental fate should be explored as a Tier 3 refinement for modeling.

The case studies were useful in understanding how SEAWAVE-QEX and SBFs fit into the evaluation Framework and illustrate the level of complexity and balancing of factors needed in a pesticide-specific risk assessment. The first case study effectively illustrated an acute exposure assessment for drinking water for "chem1", which has moderate persistence, and appears to be an effective surrogate for pesticides with primarily residential use in areas of little cropland and a high percentage of impervious surface. This case study clearly showed the processes used in moving from screening up through Tier 4, with some exceptions. Panel recommendations included:
- Provide additional guidance on evaluating the goodness of fit of the SEWAVE-QEX model, especially a discussion of weak seasonal waves and the decision criteria for judging acceptability;
- Provide more explicit guidance on where and how to address residential use, given that residential uses were not modeled, and especially in areas where both agricultural and residential uses co-occur (i.e., are they additive?).
- Clarify section 5.4 by first defining the decision criteria for determining exceedance using SBFs, so that the logic in the subsequent SBF analysis will be easier to follow.
- The Panel noted that Figure 4.3 is difficult to follow and recommended revising to indicate why "exceedances" occur at low SBF values.
- Conduct a similar short-term case study for a pesticide with long-term persistence.

DETAILED PANEL DELIBERATIONS AND RESPONSE TO CHARGE

Charge Question 1. – *SEAWAVE-QEX*

Because of the sporadic nature of pesticide concentrations in surface water, monitoring programs with limited sampling frequency often do not provide a reliable estimate of the range of pesticide concentrations relevant to cancer and non-cancer durations of toxicological concern typically considered in pesticide human health risk assessments (i.e., 1-, 4-, 21-, and 365-days). The SAPs in 2010, 2011, and 2012, suggested that EPA look into the use of a seasonal wave regression model (SEAWAVE-Q) developed by the U.S. Geological Survey (USGS) to help interpret surface water monitoring data by generating daily pesticide concentration chemographs. SEAWAVE-QEX, a modified version of SEAWAVE-Q, is designed to estimate extreme (i.e., peak or daily average) pesticide concentrations, using streamflow as a covariate. EPA evaluated SEAWAVE-QEX using high-frequency surface water monitoring data and streamflow data and concluded that SEAWAVE-QEX is a suitable tool for estimating pesticide concentrations for non-sampled days so that upper-end pesticide concentrations may be estimated.

EPA also evaluated alternatives to using streamflow (i.e., precipitation and stream stage) as a covariate. This was done because daily streamflow data are not always complete or readily available and may be seasonal (i.e., periods of each year will not have flow record). This would also allow for sampling sites located in low-flow or no-flow systems where flow would not be a suitable covariate.

> **1.a.** Please discuss the strengths and weaknesses of using SEAWAVE-QEX to estimate short-term (i.e., 1-, 4-, and 21-day) average pesticide concentrations, as well as long-term (365-day) pesticide concentrations.

> **1.b.** EPA subsampled daily or near-daily measured concentration data to generate SEAWAVE-QEX input data that are more reflective of available non-targeted surface water monitoring data. These subsampled data were used as inputs in SEAWAVE-QEX and the results were compared to the original source data. Discuss the soundness of the approach for evaluating the use of SEAWAVE-QEX to develop reliable daily chemographs.

> **1.c.** Please comment on the use of precipitation and stream stage data as inputs into SEAWAVE-QEX to provide reasonable estimates of pesticide concentrations. Also, please discuss the suitability and limits of using methods for infilling missing streamflow data (i.e., waterData R package, Ryberg and Vecchia, 2017).
> **1.d.** Considering that SEAWAVE-QEX was developed from data from flowing systems and that community drinking water systems use drinking water from a variety of surface water sources, including low- or non-flowing systems, please comment on the utility of using SEAWAVE-QEX for low- or non-flowing systems. Is the Panel aware of a better tool for infilling monitoring data for low- or non-flowing systems?

1.e. Section 6.6 of the EPA white paper describes the utility of SEAWAVE-QEX in the context of designing a surface water monitoring program. Please comment on EPA's conclusions regarding how SEAWAVE-QEX can be used to optimize a surface water monitoring program design in order to use monitoring data in pesticide drinking water assessments.

1.f. During EPA's evaluation of SEAWAVE-QEX, we used the following criteria in running the model; 3 years of data, 12 samples per year, and greater than 30% of the samples were detections. However, Vecchia (2018) notes that there is flexibility around the data requirements for input into SEAWAVE-QEX, provided the diagnostic plots indicated that the model assumptions are fulfilled. Please comment on any data characteristics, such as sampling frequency and timing within and across years, that should be considered when exercising flexibility in data requirements.

Panel Response 1a:

a. Panel members complimented EPA for advancing computational technology in the Agency's scientific and regulatory processes; in this case, specifically using the SEAWAVE-QEX for estimating pesticide residues in surface water for DWAs. The Panel also appreciated EPA's efforts in preparing the white paper and other documents for this FIFRA SAP meeting.

SEAWAVE-QEX is a parametric regression model originally built to evaluate long-term trends in sparse monitoring data and later was further refined to approximate seasonal variance of model residuals and serial correlation between measured concentrations. The model was developed with the goal of interpolating sparse daily pesticide concentration time series data by representing the principal sources of variation in daily pesticide concentration. To do so, it uses information on background pesticide concentration, seasonality of pesticide application, stream flow, and variations in stream flow at the medium-range time scale (i.e., at the monthly time scale) and at the short-range time scale (i.e., at the daily or weekly time scale due to rainfall and runoff). To account for additional, residual temporal autocorrelation in daily pesticide concentration, the model also includes autocorrelated error terms, whose addition constitutes the main development compared to the SEAWAVE-Q model previously reviewed by another FIFRA SAP.

EPA analyzed the full record of daily, or near-daily, pesticide occurrence data from four monitoring data sources [the National Center for Water Quality Research (NCWQR), the Washington State Department of Agriculture (WSDA), the Atrazine Monitoring Program (AMP), and the Atrazine Ecological Monitoring Program (AEMP)] to create datasets for use in SEAWAVE-QEX. Estimated chemographs were compared to the original measured pesticide data to evaluate the accuracy of SEAWAVE-QEX estimated chemographs.

In discussing strengths and weaknesses of the SEAWAVE-QEX model, some of the members of the Panel acknowledged not only the difficulty of the problem that the Agency is trying to address but also the elegance and relative simplicity of SEAWAVE-QEX.

In the first use of SEAWAVE-QEX described in the EPA white paper, SEAWAVE-QEX is used to "fill in" pesticide concentration values in time series of pesticide concentrations at selected USGS and National Water-Quality Assessment (NAWQA) sites. This study was designed to

examine differences between observed times series and SEAWAVE-QEX imputed time series when pesticide concentration is observed almost daily.

In the second use of SEAWAVE-QEX reported in the EPA white paper, SEAWAVE-QEX is used to impute pesticide concentration values in situations affected by large gaps of sampling, representing pesticide concentration time series at sites with sparse monitoring data. This study was conceived with the intention of investigating the performance of SEAWAVE-QEX at locations where pesticide concentrations are observed infrequently. To assess SEAWAVE-QEX performance and compare the imputed data with the observed ones, time series of daily pesticide concentrations at frequently sampled sites were subsampled using a sampling protocol designed to mimic typical practices employed by various monitoring programs across the country. Two different sampling protocols were considered: one that corresponds to more intensive sampling (i.e., 7-day sets), and a second one that corresponds to longer subsampling windows (i.e., 14-day sets). Using these two subsampling protocols, five realizations of daily pesticide concentration data recorded on 7-day and 14-day sample time interval were generated, each having a minimum of 12 samples per year to meet SEAWAVE-QEX minimum requirements. Each synthetic monitoring pesticide time series data was input to SEAWAVE-QEX and the imputed time series were compared to the full measured time series.

The second study that the Agency performed to assess the performance of SEAWAVE-QEX was quite extensive. It consisted of imputing daily pesticide concentration time series for five realizations of daily pesticide concentration values recorded on a 7-day and 14-day sampling time interval at four NCWQR sites for two pesticides (Atrazine and Metolachlor) and at two AEMP sites for one pesticide (Atrazine). In both cases, imputed time series were generated using the SEAWAVE-QEX model with two covariates (streamflow and precipitation) used alternatively as predictors in the regression model. Hence, results relative to the second study were based on an assessment of 200 time series analyses.

For each synthetic dataset, SEAWAVE-QEX was used to generate 100 conditional traces from which, for each site-year, 100 estimates of the maximum daily pesticide concentration, 100 estimates of the maximum 4-day average pesticide concentration, 100 estimates of the 21-day average, and 100 estimates of the 365-day average pesticide concentration were derived. From these estimates, the maximum of the 100 values for each time period (i.e. 1-day, 4-day, 21-day and 365-day) were computed and compared to the (true) value computed from the original complete daily pesticide concentration dataset for each site-year. In addition, summary statistics (mean, median, upper, mid, and lower percentiles) of the 100 distributions were examined and summarized as well.

Evaluating the results of these two studies on the ability of SEAWAVW-QEX to estimate pesticide concentrations when near-complete monitoring data for both concentrations and streamflow are available, the Panel identified the following strengths and weaknesses:

Strengths:

1. SEAWAVE-QEX is an elegant model that logically deconstructs a daily pesticide concentration time series into components describing major sources of variation (seasonality, streamflow variability, and long-term trends), and it allows, through conditional simulations,

to reproduce the statistical characteristics of daily pesticide concentration over a time period. Specifically, through the generation of conditional traces, SEAWAVE-QEX allows users to infer upon the distribution of selected summary statistics of daily pesticide concentration from an incomplete time series of pesticide concentration measurements.

The parameters of the model are based on measured pesticide concentrations. They have an interpretation that can be useful for designing future sampling strategies (i.e., the correlation temporal scale, also discussed in charge question 1.f.), and the multi-step model fitting process, described in the EPA white paper, is a logical approach to a difficult parameter estimation problem.

2. SEAWAVE-QEX performs well in describing near-complete monitoring data (Vecchia, 2018). For example,
    a. It seems able to adequately simulate time-series scatter of pesticide concentrations in near complete monitoring data.
    b. It can estimate pesticide concentrations with high temporal resolution, which are critical for risk assessments conducted for short-term exposure concerns.
    c. When fit to (near) complete monitoring time series, it provides estimates of maximum concentrations at the higher percentiles (i.e., >75th percentile of the annual maximum concentrations) for 1-, 4-, 21-, and 365-day averages, which tend not to underestimate the true maxima and which also tends to not overestimate maxima by more than an order of magnitude, compared to actual daily chemographs.

3. The Panel noted that the SEAWAVE-QEX function (a pulse of inputs) represents the way producers treat fields in a watershed for a given crop. The chemical application starts with the first field being treated in a watershed and ends with the last application on the last crop field being treated. The sudden change of slope at the top of the pulse (i.e., the highest asymptotic top of the pulse function) represents the time of highest loss in the watershed. This maximum changes with years, as a function of input and runoff. If all field operations go well and weather cooperates, there is a lot of field work accomplished in a very short time: this is the maximum possible daily input. If rain interrupts periods of fieldwork, the inputs are spread over a longer period.

4. By using flow as a covariate, SEAWAVE-QEX takes into account the hydrologic landscape processes that affect pesticide transport, i.e., infiltration and runoff generation or lack thereof, which drive pesticide transport.

EPA also evaluated the utility of alternatives to streamflow [i.e., precipitation (both modeled and measured) as a covariate in SEAWAVE-QEX, as streamflow data were often not available for all sampling sites either because flow data were not collected, were incomplete, or the sampling site was located in low or no flow systems. One Panel member thought that the performance of SEAWAVE-QEX does not significantly degrade when missing streamflow data are estimated with precipitation data. However, another Panel member called attention to reduced confidence in SEAWAVE-QEX applications described in the EPA white paper when flow data were unavailable and surrogates including stage or precipitation adjusted for time of travel were used in the model in place of flow.

Another Panel member noted that missing flow data are the result of different possibilities, to which different solutions may be applied. The possible causes and corresponding solutions include:

    a. Stage (i.e., water level) is higher than the range of the rating curve. In that case, streamflow estimates can be obtained by extension of the rating curve, first physical principles, or correlation with another trusted stream gage.

    b. Equipment failure causes loss of stage data, and therefore, discharge data. Values can be obtained by correlation with another trusted stream gage.

    c. Streambed went dry for a period of time. Note that there is no transport of pesticide either and those concentrations are zero as well.

    d. There is no stream gage. Streamflow values from an eco-hydrologically similar watershed may be useful as a covariate. This would be particularly useful for smaller watersheds (Hydrologic Unit Code (HUC) 12 or drainage areas <100 km^2) but not likely to work for large ones.

Several Panel members pointed out that discharge values from a hydrologic model may be useful as more and more model applications show good simulation of flow values, even with minimal or sometimes no calibration.

Actionable item: The Panel recommended that EPA consider using eco-hydrologically similar watersheds with a stream gage or hydrologic models to estimate missing flow values.

5. The use of two comparisons for the quantitative evaluation of estimated concentrations of short- and long-term periods (i.e., a comparison of summary statistics from SEAWAVE-QEX output to the maximum annual average concentrations of the measured full record datasets and a comparison of the full distribution of measured concentration data to SEAWAVE-QEX) output data is a strength.

6. The EPA white paper lists the challenges and limitations of using SEAWAVE-QEX, for risk assessment and realizes that "No model for imputing non-daily monitoring data is expected to estimate pesticide concentrations with complete accuracy," and that "SEAWAVE-QEX can both overestimate and underestimate pesticide concentrations on an annual basis." However, the model has demonstrated that underestimation rates are low for point estimates when selecting a summary statistic from all conditional simulations (i.e., maximum of the annual 99th percentile concentrations).

7. SEAWAVE-QEX performs well when applied to sparse monitoring data scenarios (EPA white paper).

    a. When the monitoring data meet general requirements of >3 years data, 12 or more samples/year spread across most of the application season, and ≤70% censored observations, the estimated maxima and 80% error bounds are less than two orders of magnitude greater than the true series maxima.

    b. As expected, the more samples per year, and more uncensored samples, the better are the SEAWAVE-QEX maxima in estimating the true maxima.

8. SEAWAVE-QEX seems to work best when the modeled time period does not significantly extend beyond the minimum and maximum sampling times and when the annual or season-limited time series is best described by one seasonal wave.

9. Even though, as pointed out by some public commenters, there are other methods of imputing missing pesticide concentration values in the near-complete USGS and NAWQA time series, such as kriging, the Panel believes that SEAWAVE-QEX offers a more flexible approach than kriging. The Panel offered several reasons for the superiority of SEAWAVE-QEX over kriging. First, kriging typically assumes a constant variance, while SEAWAVE-QEX allows for a seasonally-varying standard deviation of the residual terms. Secondly the specification of SEAWAVE-QEX, which includes not only a temporal trend, but also streamflow information to characterize a long-term temporal trend (via the Mid-Term Flow Anomaly (MTFA) term) and day-to-day variation (via the Short-Term Flow Anomaly (STFA) term), has several advantages. For example, the specification of the temporal trend components (i.e. the MTFA term) in SEAWAVE-QEX constrains the infilled values to follow the general pattern of pesticide concentrations prior to and following the missing days. Additionally, the day-to-day variation components in SEAWAVE-QEX (i.e. the STFA term) not only allows for the imputed values to exceed pesticide concentrations recorded at other days in the time series, but also constrains the imputed values not to have more variation than what has been observed prior to and following the missing days.

10. SEAWAVE-QEX generates chemographs with daily time-series that can be used to estimate uncertainties in monitoring data.

11. The model has wide applicability, as long as the site meets the data requirements and the output from the model captures the relevant regulatory statistic from less than daily data.

12. The model provides diagnostic plots that help users evaluate goodness-of-fit of the model.

13. The development of the model has been well documented in Vecchia (2018).

14. Based on EPA's evaluation using daily/near-daily monitoring data of atrazine and metolachlor, the model was able to generate daily concentrations from sub-sampled data with reasonable accuracy:
    a. Low percent of under-predicting as shown in Figure 9.18 and 9.19 of the EPA white paper, except for results on the 365-day average period.
    b. Upper percentile summary statistics of annual 99th percentile concentrations (Maximum, 99th, 95th, and 90th) were within the ballpark of measured maximum.
    c. The resulting SBFs developed using SEAWAVE-QEX estimated chemographs were close to those developed using full record of data.

15. When running the model, there is a limited number of parameters that require user input. This ensures consistency in modeling results and minimizes human error.

The Panel highlighted several weaknesses with the SEAWAVE-QEX model, and the methodology used by EPA to evaluate it.

Weaknesses:

1. Several Panel members noted that while the foundation of the SEAWAVE-QEX model is the ability to account for the seasonality of pesticide usage through application of a "seasonal wave" function, a seasonable wave component is not useful for compounds that do not have a clear seasonable pattern of use, i.e., pesticides applied sporadically. [Seasonal wave fit is an integral part of generating simulated chemographs: Refer to page 32, Figure. 3.1 of the EPA white paper]. SEAWAVE-QEX appears to not perform well for intermittently used chemicals. Members of the Panel believe that this can be ascribed to a failure in correctly identifying the season for the chemical in consideration. As currently used, season is broadly defined as the potential window of use but also as the potential window of monitoring data. For intermittent use chemicals, there may be multiple windows of use and high removal of the chemical between uses. The Panel recognized that current available monitoring data might not permit the correct identification of the season for a pesticide used intermittently and different monitoring programs might be needed to really address this issue.

   Additionally, a Panel member noted that the SEAWAVE-QEX assessment did not include pesticides that are applied multiple times to crops during growing seasons. Completing this assessment would increase confidence in SEAWAVE-QEX use. Pesticide application scenarios of this type were not specifically described in the EPA white paper. However, they are common, particularly for fungicides. An example cited was a comparative study of fungicide and herbicide loss during peanut production in the Southeastern USA. The fungicide was applied four times and the herbicide only once during each of four growing seasons within an 8-year period. Overall, fungicide loss in surface runoff was more than 10 times greater than the herbicide even though herbicide application on a mass basis was nearly 50% greater than the season-long fungicide application (Potter et al., 2014). These findings showed that frequent pesticide application may increase runoff risk and that losses may continue at high rates across growing seasons. In turn, the magnitude and timing of discharges to streams and rivers would be impacted.

   Actionable item: The Panel recommended that EPA assess the accuracy of SEAWAVE-QEX for pesticides that have multiple potential windows of use. Furthermore, the Panel recommended that EPA assess whether SEAWAVE-QEX has a waveform that would fit multi-application scenarios.

2. A Panel member noted that, without knowing the PWC, it is difficult to assess how much of an improvement using SEAWAVE-QEX makes when one compares estimated maximum concentrations to observed values (1-day or rolling averages).

   Actionable item: The Panel recommended that the EPA compare SEAWAVE-QEX results with results produced by the PWC.

3. When running the model, few parameters require user input. This ensures consistency in modeling results and minimizes human error. However, and as acknowledged by EPA, the use of SEAWAVE-QEX has a steep learning curve, it is time and resource intensive, and it will be difficult to understand by the uninitiated. For example, the following steps, which are all required when using SEAWAVE-QEX, demand both resources and expertise.

a. Collecting monitoring data (pesticide concentrations and the daily covariate data) for use with SEAWAVE-QEX requires considerable effort on a site-by-site basis, unless the site is associated with a USGS gaging station with daily streamflow measurements.

b. After the data have been evaluated and prepared, the SEAWAVE-QEX minimum data requirements for input require additional time to identify and are not necessarily straightforward.

c. Analyzing the diagnostic plots and verifying a proper model fit, which are essential to ensuring that the resulting concentrations are trustworthy for use in risk assessment, requires effort and experience.

d. Acquiring expertise in model fitting, as well as deciding the level of model refinement on a site-specific basis requires a considerable amount of effort.

e. Although EPA has developed an excellent Standard Operating Procedure (SOP) for use with SEAWAVE-QEX, understanding and interpreting the diagnostics plots takes time and it is possible that different users will interpret the results differently.

The use of SEAWAVE-QEX for only tier 4 assessments (i.e., for the few compounds that need the most refined DWAs) or when it is supported by staff that truly understand the model and the complexities of fitting it may mitigate this weakness.

4. Although SEAWAVE-QEX was developed primarily for sites with sparse monitoring data, it can also be used to fill in occasional gaps or missing pesticide concentration data in surface water monitoring datasets with higher frequencies. However, if all data from high frequency sampling fall within a short timeframe (such that all samples within a given year are clustered), SEAWAVE-QEX may not be able to adequately fit the data to an appropriate wave model to represent all seasons.

5. Collecting samples during the use season(s) is important, although SEAWAVE-QEX was found to perform better when samples spanned a greater period of the year (i.e., 6 months versus 4 months), as evidenced by comparison of the 14-day subsampled NCWQR datasets to the 7-day subsampled datasets.

6. When comparing summary statistics to the maximum of the maxima for the full, log-linearly interpolated record, the underestimation frequency was less than 15%. However,

a. SEAWAVE-QEX has a slightly higher tendency to underestimate when the 7-day sampling strategy subsets are input into the model relative to the 14-day sampling strategy subsets.

b. The 365-day annual average and the 365-day annual maximum show consistent underestimation regardless of whether streamflow or precipitation is used as a covariate. The 365-day averaging period was underestimated more frequently than the other averaging periods across sites and pesticides, with estimated concentrations at a minimum of 50% of the measured maximum concentration across sites and pesticides. This is similar to what was observed for the 4-day average (minimum ~40% of the measured maximum), while the 21-day average generally was above 60%. In contrast, 1-day average estimated concentrations were as low as 20% of the measured maximum at some sites.

Members of the Panel hypothesized that this underestimation is due to the inadequacy of the model to capture correctly the true form of the seasonal wave. Panel members highlighted that even though the SEAWAVE-QEX model includes both MTFA and STFA terms as predictors in the regression, their presence does not fully compensate

for the variability in the data, thus leaving a great deal of unexplained variation to be captured by residual terms.

7. Several Panel members noted that the SEAWAVE-QEX evaluation against measured daily or near-daily pesticide concentrations used only four sites, all located in Northeast Ohio. The drainage areas for two of these sites (Rock Creek and Honey Creek) are nested within a larger watershed (Sandusky River). All have a dominance of agricultural land (72-81%), and the main crops grown in Northeast Ohio are corn and soybean, with some hay and wheat. Most agricultural land in these watersheds is artificially drained with subsurface tiles, which accelerate the transport of water and associated pollutant from the fields to the streams. When it comes to the simulation of hydrologically related variables (here pesticide concentrations), it is important to test the model under different sets of conditions: different weather, different crops, different soils, and different management practices. Biophysical properties of a region have an impact on the vulnerability of land to transport of contaminants. In particular, rainfall amounts, timing, and intensity have strong impacts on contaminant transport from agricultural areas. Crops grown have also an impact.

One Panel member added that only three pesticides were used in the evaluation with similar physio-chemical properties from sites that were close to each other. These datasets, although very valuable, are not representative of the wide array of pesticides that are currently in use in the U.S.

Overall, the dataset does not represent the variability in watershed properties and pesticide use patterns that are observed across the U.S.

Actionable item: The Panel recommended that the EPA evaluate SEAWAVE-QEX against measured daily or near-daily concentrations across a range of watershed and pesticide biophysical characteristics. If datasets with daily data are not available, the Panel recommended considering the next best option, which may be near-daily data.

8. One Panel member noted that SEAWAVE-QEX assumes constant pesticide rates from year-to-year (page 33 of the EPA white paper). This is a large assumption as crop distribution, and thus crop acreage, total pesticide input, and the rate at which these pesticides are applied in the watershed may change from year to year as weather, crop management, and economics change. However, one Panel member observed that this assumption on pesticide rates as written in the EPA white paper differs from the underlying SEAWAVE-Q model, which is used to assess pesticide trends over time. The SEAWAVE-Q model does not assume constant pesticide inputs from year to year.

Yet, even in this case, changes in regulation and consequently in use patterns of pesticide may result in abrupt changes in temporal trends of pesticide concentration. Under this condition, the sample data should be subdivided into different periods and the SEAWAVE-QEX model should be fitted separately for each period.

Actionable item: The Panel recommended that the EPA verify the accuracy of the statement on page 33 of the EPA white paper. If incorrect, EPA should revise the white paper. If correct,

EPA should consider a method by which application rates can vary from year to year in the SEAWAVE-QEX model.

9. For the two-season model, pesticide removal is assumed to be no longer than 60 days, this may not be true for persistent pesticides, which take much longer time to dissipate. Users should avoid using the two-season model for persistent pesticides.

10. Several Panel members noted that EPA did not have criteria to evaluate the strength of similarities between concentrations predicted by the model and measured concentrations. One Panel member reminded the Panel that Vecchia (2018) evaluated SEAWAVE-QEX by comparing the annual maximum daily concentration (i.e., the 1-day maximum concentration) in a dataset of daily values generated by SEAWAVE-QEX to those obtained by applying SEAWAVE-QEX using a sub-sample of those daily values. Vecchia used the following criteria; bias was low if between 10 and 25 percent, moderate between 25 and 50 percent, or high if greater than 50 percent. The SEAWAVE-QEX evaluation does not appear to include well-defined criteria. The panel member encourages EPA to consider the following: what constitutes acceptable results by SEAWAVE-QEX; and what indicates that the peak concentration is not underestimated or grossly overestimated? (i.e., is that 10% difference, 25%, or 50%?) When comparing concentration distributions, EPA should consider what is an acceptable thickness of the colored bands that represents 100 simulations of each subsample (section 12.2 page 232 of the EPA white paper).

    Actionable item: The Panel recommended that the EPA define the criteria for SEAWAVE-QEX evaluations. These criteria should take into consideration the use of the model as recommended by Harmel et al. (2014) and the Engineering Practice on model calibration, validation, and evaluation (ASABE, 2017). Regulatory or health aspects require more stringent goodness of fit than planning purposes.

11. SEAWAVE-QEX is a mix of empirical and theoretical constructs, but at its core it is a "regression" model. As such, it is better at estimating mean patterns and less competent at estimating stochastic (residual) patterns. In addition, the important stochastic patterns may not be fully captured with less than 100 monitoring data spread over multiple years and when 70% of the data points are below detection. However, it is important to note that in these settings any regression model, and more generally, any statistical model, would have difficulty in predicting with any accuracy extreme events such as maxima.

12. As also concluded in the EPA white paper, the performances of the summary statistics derived with SEAWAVE-QEX for characterizing short-term and long-term pesticide concentrations are not easily generalizable as they vary by pesticide and site. Section 9.3, page 141, of the EPA white paper states, "[t]he summary statistic that performs best differed somewhat for each site-pesticide combination and depends on the definition of best – whether the objective is for the subsamples to always overestimate the highest annual maximum, or overestimate as often as underestimate the highest annual maximum or all annual maxima." As such, use of SEAWAVE-QEX can never be "automatic" and it will always require an educated user to apply it properly to observed pesticide concentration data.

13. The minimum data requirements (>= 3 years of data, <=30% censorship) are stringent relative to typical monitoring data that are currently available. Many monitoring datasets do not meet even the lowest requirements (especially the censoring rate) resulting in a very limited use of the model.

14. Significant errors could occur when model assumptions are violated, especially when the pesticide concentration exhibit the following properties:
    a. The variation in flow/concentration during the day is large (grab samples are used in model fitting, which are not representative of the daily peaks; daily average flow is not representative of the peak flow within the day).
    b. The relationship between pesticide transport and flow is highly nonlinear.

15. The model cannot be applied beyond the year span of the sampling data. However, this is a weakness that any model would have.

    Actionable item: Panel members recommended that EPA limit SEAWAVE-QEX use to cases when flow data are available.

Additional comments raised by some Panel members regarding the studies and procedures used to assess the performance of the SEAWAVE-QEX model, are provided below.

One Panel member noted that the Agency's efforts in performing studies to assess the performance of SEAWAVE-QEX could have been balanced to better capture both the imputation variability and sampling variability. The Panel member noted that while the Agency addressed imputation variability by generating 100 equally likely conditional traces, the Agency did not capture sampling variability well, as five replicates are not enough to provide a good understanding of sample-to-sample variability.

One Panel member also noted that with 100 conditional traces available, other percentiles besides the maximum of the maximum could have been investigated by the Agency, as the white paper discusses in Appendix E, such as:
   I. The maximum of the annual 99th percentile concentrations.
  II. The 95th percentile of the annual 99th percentile concentrations.
 III. The 90th percentile of the annual 99th percentile concentrations.

Another Panel member noted that in designing the second study to assess the performance of SEAWAVE-QEX in the case of sparse pesticide concentration measurements, the synthetic datasets were always constructed by the Agency to contain exactly 12 samples per year, which is the minimum sample size requirement for SEAWAVE-QEX, thus testing the performance of SEAWAVE-QEX in the worst case scenario.

Actionable item: For sensitivity analysis, Panel members encourage the Agency to design a study that assesses the performance of SEAWAVE-QEX when the number of pesticide concentration samples per year was slightly larger than the minimum required.

Multiple Panel members noted that the discussion of findings from both the first and the second study could and should have included more details and a more rigorous and logically organized

presentation. For example, as stated before, the accuracy or performance target was not clearly stated when discussing the quality of model fits. The Panel recognized that the EPA white paper presents some diagnostic plots, and the Panel had access to all diagnostic plots for each imputed site and much of the sparse data study outputs.

Some Panel members also noted the unevenness in the amount of details offered and the low level of readability of the EPA white paper in certain parts. Specifically, the white paper offered a lot of detail in some parts, while too little in others. Additionally, because the assessment covered so many different issues, it was difficult to follow the presentation of the results in some cases, while in other cases the key findings were immediately obvious. Nevertheless, the Panel remarked that by reviewing the EPA white paper along with the SOP document and the paper by Vecchia (2018), the strengths and weaknesses of the SEAWAVE-QEX model could be determined. **Actionable item:** The Panel recommended that EPA include the descriptions of the SEAWAVE-Q and SEAWAVE-QEX models in the EPA white paper and include the Vecchia (2018) paper as an Appendix to provide the necessary model information.

Finally, while the Panel was not asked to comment on the specific formulation of the SEAWAVE-QEX model, Panel members offered several suggestions. One Panel member believed that there was one issue with the model's specification that should be discussed as it represented a potential weakness. Specifically, in the model, the t_m term represents the midpoint of the time interval being analyzed. It appeared from the EPA white paper that in the fitting process, the t_m term is fixed and not estimated. That decision carries great consequences as t_m is clearly a function of how the "season" is specified and it centers the time scale for each year. Additionally, the selection of t_m impacts the interpretation of most of the regression parameters and the estimation of their associated standard errors. Because of the role of t_m, one Panel member suggested that the Agency perform studies to better understand how changes in the value of t_m impacts the value of other model parameters and their associated standard errors and develop guidance on how t_m should be set to best support model fitting and parameter interpretation in any particular situation.

As stated by the eminent statistician, George Box: "All models are wrong, but some are useful…,"the inevitable weaknesses in the SEAWAVE-QEX may be reduced by application of Bayesian methods coupled with Markov Chain Monte Carlo (MCMC) simulations to help in the estimation of the parameters in the model through the combination of prior knowledge and information in the data. The fact that there were eight peer-reviews and updates of the use of various versions of the SEAWAVE-QEX already captured the spirit and essence of Bayesian analysis. Further, the available monitoring datasets from many years make it ideal for the application of Bayesian modeling techniques.

Actionable item: One Panel member recommended that EPA consider a Bayesian formulation of the SEAWAVE-QEX model.

Another Panel member recommended evaluation of possibly using pesticide use intensity within watersheds in the model. Use-intensity was the most important factor in estimating atrazine concentration trends in Midwestern streams in rivers using the USGS Watershed Regressions for Pesticides model (WARP) (Stone et al., 2013). A recommended source of pesticide use estimates was the USGS county-level database. (Thelin and Stone, 2013; Baker and Stone, 2015). Currently

29

x

presentation. For example, as stated before, the accuracy or performance target was not clearly stated when discussing the quality of model fits. The Panel recognized that the EPA white paper presents some diagnostic plots, and the Panel had access to all diagnostic plots for each imputed site and much of the sparse data study outputs.

Some Panel members also noted the unevenness in the amount of details offered and the low level of readability of the EPA white paper in certain parts. Specifically, the white paper offered a lot of detail in some parts, while too little in others. Additionally, because the assessment covered so many different issues, it was difficult to follow the presentation of the results in some cases, while in other cases the key findings were immediately obvious. Nevertheless, the Panel remarked that by reviewing the EPA white paper along with the SOP document and the paper by Vecchia (2018), the strengths and weaknesses of the SEAWAVE-QEX model could be determined. **Actionable item:** The Panel recommended that EPA include the descriptions of the SEAWAVE-Q and SEAWAVE-QEX models in the EPA white paper and include the Vecchia (2018) paper as an Appendix to provide the necessary model information.

Finally, while the Panel was not asked to comment on the specific formulation of the SEAWAVE-QEX model, Panel members offered several suggestions. One Panel member believed that there was one issue with the model's specification that should be discussed as it represented a potential weakness. Specifically, in the model, the t_m term represents the midpoint of the time interval being analyzed. It appeared from the EPA white paper that in the fitting process, the t_m term is fixed and not estimated. That decision carries great consequences as t_m is clearly a function of how the "season" is specified and it centers the time scale for each year. Additionally, the selection of t_m impacts the interpretation of most of the regression parameters and the estimation of their associated standard errors. Because of the role of t_m, one Panel member suggested that the Agency perform studies to better understand how changes in the value of t_m impacts the value of other model parameters and their associated standard errors and develop guidance on how t_m should be set to best support model fitting and parameter interpretation in any particular situation.

As stated by the eminent statistician, George Box: "All models are wrong, but some are useful…,"the inevitable weaknesses in the SEAWAVE-QEX may be reduced by application of Bayesian methods coupled with Markov Chain Monte Carlo (MCMC) simulations to help in the estimation of the parameters in the model through the combination of prior knowledge and information in the data. The fact that there were eight peer-reviews and updates of the use of various versions of the SEAWAVE-QEX already captured the spirit and essence of Bayesian analysis. Further, the available monitoring datasets from many years make it ideal for the application of Bayesian modeling techniques.

Actionable item: One Panel member recommended that EPA consider a Bayesian formulation of the SEAWAVE-QEX model.

Another Panel member recommended evaluation of possibly using pesticide use intensity within watersheds in the model. Use-intensity was the most important factor in estimating atrazine concentration trends in Midwestern streams in rivers using the USGS Watershed Regressions for Pesticides model (WARP) (Stone et al., 2013). A recommended source of pesticide use estimates was the USGS county-level database. (Thelin and Stone, 2013; Baker and Stone, 2015). Currently

data are available on-line for nearly all counties in the continental U.S. for the period 1992-2017 (https://water.usgs.gov/nawqa/pnsp/usage/maps/about.php). Another data source for California is the statewide reporting program that has run continuously since 1990 (https://www.cdpr.ca.gov/docs/pur/purmain.htm). A limitation in applying these data is their geographic reporting metrics, county or township boundaries and zip codes. Simply, drainage basins do not necessary follow county, township, and/or zip code boundaries. This was addressed by a USGS group using geographic information system (GIS) based land cover data (Nagaki and Wollock, 2005). They concluded that drainage basin estimates derived from county level data provided consistent and comparable indicators of agricultural pesticide application at watershed scales.

The same Panel member also noted that suggestions to evaluate pesticide application timing in watersheds, which were made during the 2015 SAP meeting that evaluated an early version of the EPA Spatial Aquatic Model (SAM), should be consulted (FIFRA-SAP, 2015). As indicated in the report, numerous published studies have shown that rainfall soon after pesticide application often controls the magnitude of pesticide losses in surface runoff from farm fields. Thus, the use of application timing estimates in SEAWAVE-QEX may also improve model outcomes and increase confidence in results.

Another potentially limiting factor in use of SEAWAVE-QEX for pesticide risk assessment relates to the uncertainties of the model effectiveness in evaluating concentration trends in pesticide degradates. As noted by several Panel members, with some pesticides, degradates play a critical role in evaluating exposure risks. No SEAWAVE-QEX applications to pesticide degradates were provided in the EPA white paper. A published application of SEAWAVE-Q to acetanilide herbicide degradate concentration dynamics in Iowa River samples indicated that its successor, SEAWAVE-QEX, may be suitable for imputing degradate concentrations when daily data are not available (Kalkhoff et al., 2011). The same Panel member recommended further work to demonstrate this.

This Panel member made a final point that the model is constrained by the type of samples and monitoring data that can be used to process. The model focuses on grab samples and does not appear to be equipped to handle composite samples, either flow or time-based, thus limiting the potential use of this valuable approach to water quality sampling. A recent publication by a Swedish research group has shown why composite sampling may be superior to grab sample collection when monitoring pesticides in rivers (Jonsson et al., 2019). They developed a simple low-flow sampler to collect weekly composite samples and in a field study, pesticide concentrations in weekly composites from rivers draining agricultural watersheds were compared to daily grab samples. In composites, an average of 19 pesticides per sample were quantified, compared with nine pesticides per sample with grab sampling. The study also indicated a potential for underestimation of chronic exposure by grab when compared to composite sampling.

Panel Response 1b.

  b. Daily Pesticide concentrations of atrazine and metolachlor measured at monitors within the NCWQR for atrazine and AEMP data program were subsampled by the Agency to perform out-of-sample evaluation of SEAWAVE-QEX. The sampling frequency used to create the daily subsample was either one in seven days or one in 14 days. To

comment and provide an assessment on the soundness of the approach it is necessary that the approach used by the Agency is fully understandable. However, statistical members of the Panel had a hard time grasping the details of the stratified sampling procedure based on the description provided in the EPA white paper, as some key details were glossed over, and two different subsampling procedures are described in the white paper. The statement offered in the EPA white paper on page 111 describes the process of creating a stratified subsample from the NCWQR data as;

> "To meet the suggested SEAWAVE-QEX minimum number of samples, each dataset contained 12 samples/year. The first concentration for each subsampled dataset was from a random date between the first sampling date of a year and the midpoint of the sampling period (i.e., start date limited to spring), and samples were taken every seven or 14 days until 12 samples were subsampled for the year."

This highlights how the sampling procedure was devised to obtain pesticide concentration data that are more representative of a typical sampling schedule for a pesticide monitoring program, with measurements taken every seven or every 14 days.

Presentations by the Agency before the Panel and subsequent discussion clarified to the statistical members of the Panel that the samples generated by the Agency are systematic samples with only the start monitoring date chosen at random within a range of possible dates, comprised between the first sampling date and the midpoint of the pesticide monitoring period (EPA white paper, page 111) and limited to spring.

Five datasets of daily pesticide concentration sampled every seven and every 14 days were generated and used as input to SEAWAVE-QEX for assessing the performance of SEAWAVE-QEX in estimating maximum pesticide concentration, if only weekly or bi-weekly grab samples are available. Specifically, evaluation of SEAWAVE-QEX was based on two different criteria; performance of the model with respect to certain summary statistics, and with respect to the full distribution. These two types of evaluations are appropriate in the context of DWAs. Some Panel members concluded that the rationale of subsampling, generating replicates, removing those samples from the comparison datasets, and measuring the percent underestimated appears to be a reasonable approach. However, the Panel raised several comments on both the approach used to generate the subsamples, and on the model evaluation.

Regarding the subsampling procedure the Panel provided the following comments:

- One Panel member noted that while a sampling interval of one-in-seven or one-in-14 days is representative of real pesticide concentration data, the sparseness in the measurement frequency might be disruptive of the dependence structure in daily pesticide concentration: daily pesticide concentration data are not an independent series of observations, and subsampling observations should be done in a way that preserves the original dependence structure in the data. Vecchia (2018) showed in simulation experiments that bias and 80% percent error bounds were smaller when the pesticide concentration data was sampled with an inter-measurement time interval that is smaller than the concentration time scale, and that the concentration time scale for

many USGS sites was estimated to be below five days (see Figure 16, Vecchia (2018)). One Panel member noted that while the study was designed to ascertain whether SEAWAVE-QEX could be useful in real-world scenarios when only weekly or bi-weekly grab sample are available, the simulations studies by Vecchia (2018) as well as the results discussed in the white paper, might be useful to inform future sampling protocols.

- The current subsampling method, with seven-day and 14-day interval and a forced starting date in spring would result in a lack of data from the winter months. While this may result in a better model fitting as the majority of the samples occurred during the use season, some Panel members noted that this does not reflect pesticide concentrations in winter months. In some situations, winter storms, especially the first flush after the dry period, often generate peaks in pesticide concentration. The current subsampling method resulted in a lack of testing of the SEAWAVE-QEX model for these scenarios.

- The Panel noted that for one site in the AEMP program, a random sampling strategy was also examined, however, not much detail was offered as for the exact strategy employed. Specifically, the white paper (page 114) states that the random process selects 13 "values" but no indication as for how this was done was provided. It is assumed that this meant "selecting 13 dates from each year at random and retaining associated concentrations" given the white paper talks about random samples including neighboring dates. At the same time this would require dropping some samples because of the minimum three-day spacing rule.

- Two Panel members mentioned the possibility of random sampling. In particular one Panel member noted that a random sampling approach should be used to reflect monitoring programs that have irregular sampling schedules, while the second Panel member suggested an alternative random sampling strategy consisting of selecting 13 concentrations from the statistical distribution of concentrations developed from the full dataset for that year, and then assigning those concentrations to randomly selected dates. However, this Panel member acknowledged that two methods have the potential of generating quite different results, and that this should be further clarified.

- An additional component in the process of creating subsampled dataset are the constraints imposed on the subsampled datasets; these datasets should contain three years of data, 12 data points a year, with no more than 70% no-detects. This translates into the smallest usable dataset having as few as 36 sample points with 30% or four to five detects, on which the SEAWAVE-QEX model was then run. The Panel noted that as the SEAWAVE-QEX model has about seven parameters that need to be estimated, even considering the fact that the streamflow time series provides helpful information in estimating some of these parameters, this would still amount to estimating at most seven parameters from four to five detects, which the Panel noted to be a lofty task.

Regarding evaluation of the SEAWAVE-QEX model using the subsampled data, the Panel has raised several concerns.

First, the Panel recognized that the ultimate goal of the subsampling approach was to create synthetic sparse monitoring datasets from a complete dataset where the "true" maximum concentrations are known. The Agency generated multiple subsamples to see how limited monitoring combined with SEAWAVE-QEX modeling produces variability in the estimated

maximum concentration, and to assess confidence intervals. In fact, for each subsample, not only could the closeness of the maximum concentration estimate to the true maximum be assessed, but also the extent to which estimated confidence intervals include the true maximum. Looking across the multiple subsamples can determine how far off the point estimates are, and how often the estimated confidence bounds do not include the true value. While the Panel appreciated the effort of the Agency to consider sampling variability, the statisticians on the Panel noted how the number of replicates considered by the Agency – five subsamples – is too small to really assess sample-to-sample variability in the estimates. The statisticians on the Panel recommend examining 20-30 subsamples and performing a more rigorous analysis of within and between variability in the maximum concentration values generated.

Actionable item: The Panel recommended that the Agency consider repeating the analysis; increasing the number of subsamples to at least 30 in order to better distinguish between within sample and between sample variability in the estimated maximum concentration.

Secondly, several Panel members noted the lack of criteria to determine if sampling makes a large difference or to determine the best sampling strategy. For example, the magnitude of the underestimation was 50% of the observed estimate for the 365-day average, and as much as 20% of the observed estimate for the one-day estimate. According to Vecchia (2018), this is a high bias. However, in the white paper, there was no definition of an acceptable bias and underestimation rate. Similarly, there was no definition of criteria values (frequency and magnitude) for over-estimation. A Panel member noted that having unrealistically high concentration values may make getting buy-in for the model more difficult.

Actionable item: The Panel recommended that EPA set threshold values for under and overestimation frequency (i.e., how often is the concentration under or overestimated) and bias (i.e., by how much is it under or overestimated) as a function of the use of the model.

One Panel member remarked that while the current approach, which compared the SEAWAVE-QEX generated results using subsamples to the log-integrated full dataset, is a proper model evaluation method, it might be prone to model over-fitting. There are two main reasons for this; (i) SEAWAVE-QEX tries to generate the best-fit model to the particular subset of the data, and (ii) the subsampled datasets are essentially from the same population (the full dataset).

The Panel member believes that this does not allow evaluation of the performance of the model when applied to datasets that have different seasonal wave and trends from the dataset current employed for evaluation.

In general, the Panel had difficulty understanding the take-home message from the assessment of point estimates in the white paper. The summary concentrates on the percent of time the SEAWAVE-QEX estimates from the limited monitoring subsamples are lower than the true measured maximum concentration. The white paper offers that this occurs approximately 15% of the time, which does not sound very bad. However, Figure 9.20 also suggests that when the SEAWAVE-QEX estimates from the limited monitoring exceed the true maximum, it typically is less than twice the true maximum but can be as high as 19 times the true

maximum. No effort seems to have been applied to exploring which site or sample characteristics correlates to better or poorer estimates. Such information might further inform when SEAWAVE-QEX could be used.

The difficulty of applying SEAWAVE-QEX to sparse monitoring data is tied to the general difficulty of properly estimating the seasonal standard deviation (SSD) and associated autocorrelation coefficient (correlation time lag) from such few data. Getting these statistics estimated well is key to producing maxima estimates that have good performance. The results summary suggests that when minimum data requirements are met, the maxima estimates are within acceptable limits of over- and underestimation.

A final comment regards the issue of estimated model parameters. As one looks through the diagnostic plots, the impression is that often, some of the regression parameters are not significant and the estimates are quite small. In some applications of linear regression, non-significant predictors are eliminated from the linear regression model and the model is re-fitted to the data. This does not seem to be considered here. One statistician on the Panel wondered whether this would increase the estimated maximum since the little variability that was being explained by the non-significant component was now mostly incorporated into the SSD term, likely increasing it and hence, increasing the potential of generating a maxima. According to this Panel member, this could have a large effect if the non-significant term is the STFA. Another statistician on the Panel disagreed with this suggestion and advised against the idea of dropping the non-significant terms.

Finally, it was not clear to the Panel whether the requirement for the SEAWAVE-QEX model is of 12 pesticide concentration measurements per year or 12 measurements per seasonal wave. A more explicit statement about this in the white paper is recommended.

Actionable item: The Panel recommended that EPA clarifies explicitly what are the requirements in terms of number of pesticide concentration measurements needed per seasonal wave and the number of pesticide concentration measurements needed per year.

Panel Response 1c:

  c. The Panel agreed that examples provided in the white paper indicated that using SEAWAVE-QEX with precipitation or stage as covariates in cases when flow data may be unavailable were promising in evaluations of water quality monitoring data. The need for covariates was also highlighted by the observation that budget cuts continue to reduce the number of USGS gauging stations thereby limiting access to flow measurements. It was noted that EPA understands limitations in the use of flow surrogates like precipitation and has established a logical hierarchy of data use, i.e., flow→stage→precipitation. The continuing challenge seems to be uncertainty in determining where use of covariates is appropriate and where not. Answers to these questions may require complex and time-consuming analyses to establish confidence in results. Among other considerations this means that the high level of professional judgement required to perform these analyses will limit SEAWAVE-QEX applications using alternate covariates to the most highly experienced model users.

While the Panel agreed that precipitation may be an effective surrogate, one Panel member observed that using precipitation as a covariate would entrain the loss representation of landscape runoff/infiltration processes. It was emphasized that streamflow is an outcome of precipitation, which also causes pesticide transport. For example, precipitation is the same whether soils are wet or dry. Yet runoff will be quite different and pesticide transport, as well. One inch of precipitation on very dry soils may cause no runoff while the same inch may cause a lot of runoff, if the soils are saturated. Another Panel member supported these comments stating that intuitively, it would seem that precipitation would make a good covariate for predicting pesticide concentrations because it is the runoff that will carry pesticide into surface water. However, it is less clear that precipitation would be a good indicator of flow. As precipitation amounts are highly variable and can result in different flow rates depending on the land upon which the precipitation falls, it will not be a consistent measure of flowrate. For example, in highly urbanized land, a high percentage of precipitation will enter the waterbody and therefore, precipitation would likely be a major factor in the flow rate in the near term (temporally). However, in rural areas, precipitation is likely to be entrained and may not enter the waterbody or will enter after some exaggerated lag period (as compared to the urban scenario). Additionally, infiltration may be a more dominant factor for water sources entering the waterbody in rural simulations. It was also noted that although figures in the white paper indicated that using precipitation as the covariate can produce reasonable estimates. However, substantial over- or underestimate of pesticide concentrations may result due to the additional complexities precipitation brings to the analysis (discussed above) as well as confounding factors dependent on application type (i.e., aerosol, liquid spray, ground treatment, etc.), and chemical properties.

Both of the Panel members mentioned above also commented on assessments of precipitation use as a covariate using data form the four NCWQR watersheds located in Ohio. The Panel members agreed that the tests run in the four watersheds showed that precipitation was acceptable as a covariate because it produced similar results when compared to discharge. However, two points were made that limit extension to other watersheds. First, agricultural land in the test sites is tile drained for the most part. This makes the transport of pesticides from fields to streams rapid, which may give an advantage to using precipitation as a covariate. Second, the NCWQR sites used are all in the same area and can be expected to have similar patterns of precipitation. It was recommended that an investigation of mixed use, urban, and other sites not evaluated in the whitepaper be conducted to determine the effect of different land types and land usage on the applicability of the model. It is also recommended to conduct comparisons of SEAWAVE-QEX using flow and precipitation in other regions of the country, which would have different patterns of precipitation.

Regarding precipitation and its functional form in the model, one Panel member recommended that an explanation of the basis for the transformations used would strengthen arguments for use of this covariate. It was also suggested that consultations with the SEAWAVE-QEX model developers may be helpful. In the USGS report describing SEAWAVE-QEX, it is stated that "if daily streamflow are not available, surrogate variables computed using estimated precipitation from the watershed may be considered in place of streamflow" (Vecchia, 2018). Johnson et al. (2011) was also cited. In this paper an early version of the SEAWAVE model was used to calculate pesticide loads using monitoring data collected in California, Oregon, and the State of Washington. In the model, transformed precipitation (A_p) was used in conjunction with, not in place of flow (see equation below).

$\ln(L) = a_0 + a_1 \ln(Q/\sim Q) + a_2 \ln(T/\sim T) + a_3 W + a_4 A_Q + a_5 A_P$

L = load of the constituent

Q = daily stream flow

T = decimal time in years

$\sim$ = indicates use of a centering variable

W = the fitted wave form

A_Q = transformed (unitless) flow term (ratio of recent flow to flow in prior 30 days);

A_p = transformed (unitless) precipitation (ratio of recent rain to rain in prior 30 days) scaled from 0 to 1.

$a_0, a_1, a_2, a_3, a_4, a_5$ = fitted scalar coefficients

The Panel observed that there is a sound technical basis for using a precipitation term in the model that accounts for antecedent rainfall as was done in this version of SEAWAVE. This is because relatively small differences in soil water content can have a large impact on runoff during subsequent storm events (Truman et al., 2008).

Another point of agreement among Panel members was that if precipitation is used as a surrogate covariate in SEAWAVE-QEX, clear guidance is needed on what precipitation data to use, model or measure. It was observed in the white paper that using measured precipitation provided better SEAWAVE-QEX model fits than modeled precipitation. While this seems reasonable, further work appears necessary for verification. In addition, Panel members emphasized that use of measured precipitation can be challenging since it is often highly variable over relatively short distances (miles). Refer to Bosch et al. (1999), as an example. Guidance on how to address variability in measured values is needed. It was suggested that tools such as NEXRAD imagery may be useful in providing estimates of precipitation at larger scales (NOAA, 2019).

The Panel also recommended that irrigation and incidental runoff in arid areas be accounted for in SEAWAVE-QEX where appropriate. Currently, the model does not account for irrigation and other water inputs. One possibility is to add irrigation amounts to daily precipitation. However, this requires deriving irrigation estimates at watershed scales. Data of this type at spatially relevant scales may not be readily obtainable. A related point, made by a Panel member, was that irrigation is an effective tool to reduce pesticide runoff (Potter et al., 2008). Well timed irrigation can incorporate pesticides into soil, therefore, reducing amounts available for runoff. Thus, in addition to the potential to increase runoff, irrigation may reduce pesticide losses. Given this knowledge of when and where irrigation was used is needed to inform risk assessments.

As was the case with precipitation, the Panel agreed that the examples provided in the white paper indicated that stage may be an effective surrogate for flow in SEAWAVE-QEX, but there needs to be more effort in defining when this is the case. One Panel member commented that as a covariate, it is clear that stage data would likely be a good predictor of streamflow. A second Panel member supported this, explaining that the relation between flow and stage can typically be described with a power function which goes well with a log-linear model. Thus, the same Panel member thought that, mathematically, using stage as a surrogate for flow should work. It was observed that the power is usually not the same over the whole range of stages but it should be inconsequential because the focus is on the high concentrations, which are usually associated with high flows.

Another Panel member took a more cautionary approach, when commenting on the use of stage, observing that relationships between stage and discharge are typically non-linear and not easily predicted. This is why rating curves are needed to compute flow from stage data. There are many settings where this is challenging. For example, low gradient streams and rivers that are typically found in the southeastern Coastal Plain often expand well beyond stream channels into riparian forests that dominate lower landscape positions. This makes it difficult to relate to discharge. Stream-discharge relationships may also change over time as stream channels are altered by erosive flow, debris deposition, ice-jams, tidal influence, etc.

The same Panel member agreed that, as noted in the document, it is difficult to use stage for a reservoir because stages may change minimally in response to incoming flows. In addition, water releases from the reservoir control that reservoir's stage.

In regard to "infilling" for missing flow data, one Panel member highlighted the potential for filling missing streamflow data, with the "fillMiss" function in the waterData R package (https://cran.r-project.org/web/packages/waterData/index.html). The function appears to fill missing values using a method close to linear interpolation. The function allows the user to set the maximum gap it will fill, but the default of 30 days is larger than can likely be filled reliably. The Panel member also agreed that overall approaches described in the white paper seemed reasonable, although there is a need to show how data processing steps (i.e., inserting small numbers in place of zeros) may impact results. In some cases, a zero is a zero, such as in cases where there is water in the stream channel for collection and analysis but there is no flow. This often happens in small streams and rivers when evapotranspiration is very high. A related question highlighted by one Panel member is how provisional data may be handled. The USGS states that data are provisional and subject to revision until they have been thoroughly reviewed, have received final approval, and users are cautioned to consider carefully the provisional nature of the information before using it for decisions that concern individuals or the public at large (USGS, 2019). Again, guidance is needed.

Discussions were concluded with one Panel member commenting that the United States covers a continent with wide variation in climate and watershed characteristics; thus there is no perfect model that is going to work in every situation. The same Panel member recommended consultation with and/or hiring a hydrologist to evaluate model assumptions and outputs to ensure they are reasonable. Finally, the Panel were in consensus in their recommendation that for transparency assumptions/validation should be documented whenever SEAWAVE-QEX is used.

Actionable item: The Panel recommended that the EPA should perform its evaluation using precipitation as a covariate in areas without artificial drainage and with high infiltration, where pesticides with high solubility and low adsorption have a chance to reach the aquifer and get to the streams with groundwater.

Actionable item: The Panel recommended EPA to include parameters obtained with precipitation, stage, and flow, in order to assess whether the changes are applicable.

Panel Response 1d:

        d.   A Panel member began discussions by emphasizing that from a risk assessment perspective, priority should be given to evaluating pesticide concentration dynamics in static community water systems (CWS) using SEAWAVE-QEX and other tools that may be available. This was supported by reference to findings of the AMP conducted on CWS by the herbicide's primary registrant. A summary of data showed that the upper centiles of atrazine concentrations in raw water samples trended higher in static versus flowing CWS (Mosquin et al., 2012). This finding coupled with the observation that more than 70% of the CWS studied in the AMP program were static/non-flowing (Hendley et al., 2012) indicated that exposure risks in drinking water were likely much higher for static versus flowing CWS. A related concern is that exposures may have a longer duration with static systems. This was supported by SEAWAVE-QEX applications described in the white paper. Waves that were fit to data had relatively low amplitude and long duration as indicated by their broad widths.

A second Panel member concurred, observing that the Panel recognizes the importance of managed surface impoundments and reservoirs for drinking water supplies. It was noted that although SEAWAVE-QEX has been applied to low-flow and no-flow water bodies in test cases, more analysis of reservoirs and water supply storage systems is needed to guide different sampling schemes, and potentially a different model should be used. The Panel questioned whether the SEAWAVE-QEX functional form was relevant and optimal for such water bodies and indicated that the topic issue of watershed scale has not been adequately addressed in the current evaluations. Fractal scaling behavior of pesticide concentrations in stream water has been demonstrated using monitoring data from the Heidelberg Water Quality Laboratory and the USGS National Water Quality Assessment Program and the National Stream Quality Accounting Network (NASQAN)(Gustafson et al., 2004). Those analyses concluded that maximum daily concentrations generally decrease with watershed scale, despite lower percentile concentrations not scaling similarly, and the concentrations estimated at edge-of-field scales matched results of the Pesticide Root Zone Model (PRZM) (Carsel et al., 1998). Based on this knowledge, peak concentrations will likely be lower in rivers of large watersheds than at upstream monitoring locations. Contributions from large watersheds to drinking water supply storages/reservoirs are expected to behave similarly. Even so, multiple contributing areas and point sources need to be assessed.

Regarding evaluations of SEAWAVE-QEX with low-flow and static water system pesticide monitoring data that were described in the white paper, there was general agreement that they have provided a "proof of concept" that SEAWAVE-QEX could be applied to low-flow and static systems. However, overall model performance was difficult to assess. To quote from the EPA white paper, "SEAWAVE-QEX provided a satisfactory fit for some of the data for both flowing and static CWSs in the AMP, although the model was not appropriate for use with all evaluated systems. Further, for static systems only three of the six analyzed fit WAVE functions." With some systems where fits were observed it was noted that "[i]t is unclear from the data for any of these systems if these can truly be considered peaks or seasons; however, the model was able to fit the data." This is indicative of a problem with the model application. Another concern was the use of precipitation as a covariate for reasons described previously (see responses to charge

question 1.c). Examining use of other covariates such as pesticide use, and timing of application could prove beneficial and was recommended.

This uncertainty led to one Panel member to observe that for a non-flowing system (i.e., static) SEAWAVE-QEX does not appear to be appropriate at this time and additional analyses will be necessary if SEAWAVE-QEX is intended for use on these systems. It was stated (as described above) that, although some analysis is provided in the EPA white paper, the analysis is insufficient to ascertain whether the model is appropriate for non-flowing systems. It was highlighted that from Vecchia (2018), the model is comprised of several factors; 1) the seasonal wave, 2) the STFA, and 3) the MTFA. For a non-flowing system, factor 1 would be expected to be identical to a flowing system as it is a loading factor (or at least akin to one). However, factors 2 and 3 are less clear in their interpretation and it is not clear how these factors would apply to a non-flowing system.

Based on the description of STFA in both the EPA white paper and Vecchia (2018), the STFA value should be positively correlated with concentration when runoff via precipitation is important. From Vecchia (2018), it appears that a negative correlation corresponds to dilution, rather than runoff, being the dominant factor. Thus, although inferences as to the behavior of STFA can be made, it is not clear that this term is applicable to a non-flowing system. Furthermore, the white paper states that MTFA does not have the same implications as STFA, because it is correlated with long-term flow trends. Vecchia (2018) again provided additional detail and stated that MTFA is the seasonal variability associated with changes to the mean annual flow. As with the STFA term, it is not clear how MTFA would apply to a non-flowing system.

Vecchia (2018) concluded that the above descriptions rely on a mechanistic understanding of the anomaly terms. As flow is not likely to be the dominant factor for non-flowing systems these terms are likely analogous to the environmental fate parameters relevant to the chemical in question (i.e., each correspond to an aggregate of loss and/or fate terms such as sorption, degradation, hydrolysis, burial, etc.). Therefore, correlating these processes to flow or precipitation (as you would in determining the STFA and MTFA values for a flowing system) appears to be tenuous at best.

The Panel was less critical regarding applications to low-flow systems. One Panel member expressed that in this case (low-flow) SEAWAVE-QEX appeared to be appropriate for use. However, the Panel member expressed concerns that the resulting data would be prone to larger error than for full-flowing systems because low-flow systems would be more prone to flashing, i.e., spontaneous and rare high flow events. These events would be expected to be the driving factor behind high concentrations. Moreover, it is unlikely that samples would coincide with these events. Therefore, the Panel recommended the EPA to use caution when interpreting results from SEAWAVE-QEX from low-flowing sites. The discussion concluded with the Panel providing several references and recommendations for infilling data for low-flowing systems. Some potential methods for infilling data can be found in Gyau-Boakye and Schultz (1994). Another, more recent approach, although it may be more intensive than those presented in the aforementioned references, is shown in Tencallec et al. (2015).

Recommendation:
- Evaluate the applicability of SEAWAVE-QEX to low-flowing systems by analyzing the STFA, MTFA, and seasonal wave terms and determine if their behavior is realistic and expected for these systems.

Panel Response 1e:

e.   The Panel recommended that if EPA plans on adopting and using SEAWAVE QEX, it should do some outreach to various state and federal programs conducting pesticide monitoring. To aid organizations conducting SEAWAVE-QEX compatible monitoring, EPA should prepare a document stating the data needs for such models (i.e., sampling frequency, duration, pesticide use data at the watershed, drinking water source, proximity to USGS flow/state data). If an agency can accomplish their monitoring goals, and with minor adjustments, get data EPA could use, the Panel felt that most researchers will try to accommodate. The argument for making the data more useful is strong. Lastly, it would benefit the EPA to develop a database to receive vital organizational data. For example, the FDA has the Electronic Laboratory Exchange Network (ELEXNET) for its Food Emergency Response Network (FERN) labs; Asking various stakeholder groups to contribute data will require additional QC efforts, however it would provide the EPA additional monitoring data.

The Panel recommended that sampling, wherever possible, be conducted to SEAWAVE-QEX requirements (i.e., in places where flow is measured) and should be done in flowing systems.

The Panel noted that more samples are generally better than fewer, and that more may be required to ensure the minimum requirements for SEAWAVE-QEX are met, i.e., a sufficient number of detects in each year of sampling. However, beyond the requirements for SEAWAVE-QEX, additional samples may not provide sufficient additional information to justify their expense.

The SEAWAVE-QEX model has three basic minimum requirements relating to the number of years of sampling, number of samples per year collected, and percent allowable censorship of the measurements. Three years is the minimum requirement of SEAWAVE-QEX for a given site. The minimum number of samples that would need to be collected each year for use in SEAWAVE-QEX will depend on the pesticide; for example, additional samples may be necessary for pesticides with low seasonality, intermittent use, and/or low persistence. The Panel agreed with EPA that, even with additional samples collected, SEAWAVE-QEX may not be suitable for some pesticides with sporadic occurrence and low seasonality, as was described by EPA in the white paper for simazine. The Panel also agreed that analytical methods should be sufficiently sensitive to detect and quantify pesticide concentrations in water below biologically significant thresholds.

One Panel member noted that insecticides, which may have toxicity concerns for human health or aquatic life at low concentrations, also may have sporadic occurrence, low detection frequencies, and at least in some regions, low seasonality. The SEAWAVE-QEX model may be less suitable for such insecticides.

Because the evaluation presented in the white paper covered only four chemicals in one geographical area, one Panel member recommended developing sampling programs for other

regions and chemicals with the extensive sampling needed for further evaluation of SEAWAVE-QEX.

The Panel previously discussed that with enough years of sampling, the data could be split into sets of years that are similar to each other, and SEAWAVE-QEX fit separately for each set. Conversely, an anomalous year of weather might make fitting SEAWAVE-QEX more difficult, so sampling for more than the minimum of three years would make it more likely that the program would produce sufficient data for SEAWAVE-QEX.

Regarding non-targeted monitoring programs, one Panel member noted that many such programs use broad spectrum analytical methods for pesticides, so an entire suite of pesticides would be analyzed in every sample; from a practical standpoint, the number of samples would be the same for all pesticides, but could be set based on the most problematic (or most important) pesticide.

Panel Response 1f:

      f.   As an overall comment, the Panel noted that the criteria provided by EPA in the white paper and in the charge question are more stringent than those described in Vecchia (2018). Thus, considering this and the other data requirements provided in the Panel's response to charge question 1.a., the Panel deems the data requirements provided as appropriate.

With respect to sampling frequency, minimal dataset requirement and model evaluation, the Panel had several comments.

As for sampling frequency, the Panel would like to note that simulation results provided in Vecchia (2018) showed that if the modeling assumptions underlying the SEAWAVE-QEX model are satisfied, then bias and 80% percent error bounds are smaller with a censoring rate of 30% compared to 70%. Additionally, the simulation results also showed that a greater sampling frequency is associated with a smaller bias and smaller than 80% error bounds, particularly in relationship to the correlation in the data.

According to one Panel member these findings can be used to exercise some flexibility in terms of overall sampling frequency. Particularly, the finding seems to indicate that if the pesticide concentration data has residual temporal correlation once the effects of streamflow (or precipitation), seasonality and fluctuations in streamflow both at the monthly time scale and at the daily time scale have been taken into account, then it is important that the pesticide concentration is sampled frequently enough to capture this residual temporal correlation. However, if all the autocorrelation in the water pesticide concentration data is already explained by these factors, frequent samples are not necessary.

Subject matter (hydrological, toxicology, etc.) knowledge might be useful in understanding whether it is likely that for a given site that pesticide concentration data would still be correlated, even after already having accounted for seasonality, streamflow and mid- and short-range streamflow anomalies, thus informing a user/assessor whether pesticide data with a higher sampling frequency is needed to reliably generate daily chemographs.

As for sampling frequency within years, a Panel member noted that not much flexibility can be exercised in terms of detection rate, as the 30% detection rate requirement applies to each year of data being modeled. To illustrate this, the Panel member elaborated that under the minimal data requirements for SEAWAVE-QEX, one would expect to see at least 36 monitoring data points over a three-year period with at least 11 data points not below detection (censored). The Panel member added that if one of the three years has no uncensored values, then that year should be dropped from consideration, thus making the site fall below the minimal SEAWAVE-QEX data requirements, regardless of whether the year is the middle of the three-year period or is the first or last year. Ignoring the failure to meet the minimum data requirement and fitting the SEAWAVE-QEX model would lead to plots of the residual by time that show heterogeneity of residuals over time, clearly going against one of the SEAWAVE-QEX model assumptions. The same Panel member also noted that the white paper contains part of this discussion in section 6.2.1 of the SOP.

The Panel also noted that the SEAWAVE-QEX SOP suggests even more stringent requirements as for minimal datasets as indicated in section 6.1, page 41, of the EPA SOP for using SEAWAVE-QEX where advice on handling poorly fitted data are provided along with the following statement:

> "If more than three years of data are being analyzed, do at least the first and/or last year of data individually fulfill the SEAWAVE-QEX requirements of minimum samples and detection rates? If the first and/or last year of data being analyzed does not fulfill the SEAWAVE-QEX requirements, then remove any non-compliant first or last years from analysis. The first and last years of the dataset are anchor years for the model, so it is important for them to meet the minimum requirements, if not have higher detection frequencies, while maintaining a full dataset that meets the minimum requirements of SEAWAVE-QEX."

The Panel also remarked that other key characteristics in the minimal dataset condition have to be considered when exercising flexibility in sampling, specifically, the ability to detect a seasonal pattern and the ability to define a season.

With regard to detection of a seasonal pattern, the Panel noted how in section 6.1, page 41 of the SOP document the question is asked "Do the chemical detections appear to be random with little or no seasonal pattern?" This statement highlighted the importance that timing has to the detections in a dataset timeline for a successful SEAWAVE-QEX model fitting.

As for the definition of season, the Panel noted that the EPA white paper and the SOP both mention altering years of analysis and the benefit of using "in-season" monitoring data for some sites. However, it is not clear how a "partial sampling season" is to be defined. The SOP mentions a partial sampling season as one having ≤ six months and suggests using "best judgement" in the designation of the season. It would seem important to the selection of the best seasonal wave form for there to be non-detects at the beginning and at the end of the sampling season, but it is not clear how many should be there. With 12 data points per season and only three to four detects, if the detects are located in the middle of the season, then there would be four non-detects at the beginning and four non-detects at the end. It is not clear what would happen if there is only one detect at the end of the season and seven to eight non-detects at the beginning. This lack of clarity

highlights how the spread of monitoring data across the use season within a year and across all years as well as the location of detects within and among years is critical to obtaining acceptable results from the SEAWAVE-QEX model.

With respect to model evaluation and data requirements, the Panel observed that as noted in Vecchia (2018), "[…] model verification can be especially difficult when a large percentage of the data (more than 50 percent) are censored, in which case a large percentage of the normalized residuals are generated values from a conditional trace and not observed residuals. The generated residuals, by definition, satisfy the model assumptions." Therefore, the Panel recommended that a bright-line criteria be provided in the framework document to explain what would disqualify a model fit from consideration.

Additionally, the Panel suggested that to incorporate more data, it may be possible to perform fits over alternative year schedules (i.e., starting just prior to the application season) and, therefore, more datasets could meet the minimum requirements.

Action items:
- The Panel recommended that EPA determine a bright-line criteria to explain to users and regulated industry when a model fit should be disqualified from consideration. This criteria should be considered with the difficulties in evaluating highly censored datasets in mind.
- The Panel recommended that EPA evaluate whether alternate "years" could increase the amount of data available (i.e., by using water years, years coinciding with the first application, etc).

The Panel also offered a suggestion on improvements that can be made to diagnostic plots used to determine the adequacy of fit of the SEAWAVE-QEX model.

The SEAWAVE-QEX model assumes that the normalized residuals are approximately normally distributed with mean zero and variance equal to one. To assess adequacy of this assumption, the SOP suggests examining plots of the normalized residuals with time of the year. Within these plots the advice is to look for obvious seasonality remaining in either the mean trend or changes in spread over time. The SOP also suggests that this plot will show obvious non-normality (skewness, outliers, and so forth). The Panel believes that a better diagnostic for examining non-normality would be a normal quantile-quantile plot (Thode, 2002). Normal quantile-quantile plots would also show clearly the percent of censoring in the data and how much of the censored residuals have been imputed. The Panel also suggests that on the recommended plot of normalized residuals versus time, a loess smooth curve be added to the linear trend line to better illustrate trends in the mean over time. It might be possible to add a similar smooth curve to the plot to better illustrate residual spread.

Finally, one Panel member commented and recommended that, for mid- to long-term, the Agency continue the refinements and improvements of the SEAWAVE-QEX DWA approach but consider implementing it within a Bayesian modeling framework using MCMC simulation methods, which allows the incorporation of new information as it becomes available. Specifically, the Panel member noted the SEAWAVE-QEX model formulation contains parameters [i.e., $W(t)$, $A_{MT}(t)$, $A_{ST}(t)$, $(t-t_m)$] informing pesticide concentration [i.e., $C(t)$] that vary over time, and whose

distribution could be easily updated in a Bayesian modeling framework using the monitoring data that EPA has already collected for the four test chemicals.

The Panel member also observed that a Bayesian approach will allow EPA to continually improve the estimation of the SEAWAVE-QEX model parameters as new monitoring data and other relevant information are available. Additionally, it will allow the Agency to use some of the currently "unused" data (a concern indicated in CropLife America's written comment) in the simulation-based model fitting (Bernillon and Bois, 2000; Marino et al., 2006; EPA, 2010). The application of MCMC would have the benefit, among others, of minimizing the effects of variability and uncertainty as well as addressing the issue of codependency and covariation resulting from interactions among parameters (Bernillon and Bois, 2000; Marino et al., 2006; EPA, 2010). The Panel member believes that the recommendation to EPA of considering Bayesian modeling is timely because: (1) of the availability of many years, if not decades, of monitoring data that are ideal for MCMC/simulation based model fitting; (2) some or all of the "unused" datasets may be usable in model fitting; (3) some of the units in EPA such as CPHEA have consolidated experience with Bayesian modeling and MCMC (EPA, 2010); and (4) the application of a Bayesian modeling framework that allows to incorporate prior information would alleviate the concerns of the EPA/OPP staff on the limited amount of sampling data available.

Charge Question 2. – *SAMPLING BIAS FACTORS*

While SEAWAVE-QEX provides a way to estimate daily pesticide concentrations from non-daily surface water monitoring data, for many pesticides there are not sufficient monitoring data to use SEAWAVE-QEX. This is because the data are too highly censored or there are not enough samples per year or across years. Therefore, EPA would like to have an alternative approach to estimate drinking water concentrations from non-daily pesticide surface water monitoring data. Previous SAPs supported the development of SBFs to derive bounds on pesticide concentrations in surface water.

EPA used SEAWAVE-QEX to develop daily pesticide chemographs from infrequent surface water data. These chemographs were used to generate short-term SBFs for acute exposure durations of concern, utilizing the methods supported by past SAPs for sites with varying attributes across the contiguous United States. Building upon that previous work, EPA evaluated different sampling strategies and imputation techniques to develop short-term SBFs and concluded that using a random sampling strategy with a log-linear imputation is suitable for deriving short-term SBFs.

In addition, EPA developed a new method for developing long-term SBFs for chronic or cancer exposure durations of concern. EPA concluded that for pesticides with chronic and/or cancer endpoints, that as few as four samples per year can be used to estimate 365-day average concentrations for use in DWAs.

> **2.a.** Please comment on the use of SEAWAVE-QEX estimated daily pesticide concentrations, including the underlying data requirements, to expand the data available to derive pesticide-specific SBFs. Please comment on EPA's optimization of short-term SBFs, derived using a random sampling strategy with log-linear imputation, to estimate

the range of potential concentrations not measured between sampling events. Please also comment on the long-term SBF approach.

 2.b. Please discuss the strengths and weaknesses of developing site-specific SBFs using a percentile (for example, the median [across realizations] and median [across years]) from SEAWAVE-QEX estimated daily pesticide concentrations. Describe the utility of this approach for use in pesticide DWAs as highlighted in the attached DWA case studies.

 2.c. Please comment on the utility of using the maximum short-term and median long-term SBFs for the four pesticides (i.e., atrazine, carbaryl, chlorpyrifos, and fipronil) to estimate upper-bound concentrations of other pesticides in surface water for which either SBFs cannot be or have not been derived. In addition to using these values as a screen, EPA proposes an option to select SBFs for an individual pesticide (i.e., atrazine, carbaryl, chlorpyrifos, and fipronil) based on other defining attributes, such as environmental fate properties, use profile, flow rate, basin size, waterbody type, and/or land use. Please comment on what factors EPA should consider when selecting an alternative SBF for estimating upper-bound pesticide concentrations on a national or regional scale.

 2.d. In Chapter 4.6, EPA concludes that SBFs are a reasonable tool for increasing the amount of available monitoring data that can be used as a quantitative measure of exposure in pesticide DWAs beyond those meeting SEAWAVE-QEX criteria. Considering the answers to the questions 2a-2c, please comment on this conclusion.

 2.e. Section 6.6 of the white paper describes the utility of SBFs in the context of designing a surface water monitoring program. Please comment on EPA's conclusions regarding how SBFs can be used to further optimize a surface water monitoring program design for the greatest utility of monitoring data in pesticide DWAs.

Panel Response 2:

 a. There was consensus from the Panel that SBFs can be valuable tools for risk assessment and the Panel agreed that use of SEAWAVE-QEX simulations to expand the universe of data available for computing short and long-term SBFs is a significant advancement. Sampling bias factors are multiplicative factors used to calculate an upper bound prediction interval (i.e., 95th percentile) for a given measured concentration value to account for a less-than-daily sampling frequency in which peak concentrations may not have been measured. By multiplying the SBF and a calculated summary statistic from the available monitoring data, it is envisioned that an upper-bound concentration can be derived that will address the uncertainty in the measured pesticide concentrations due to infrequent sampling for the development of short-term SBFs for pesticides with acute toxicity concerns and long-term SBF for chronic and cancer toxicity concerns. Sampling bias factors are envisioned as an alternative approach to estimating drinking water concentrations when available monitoring data are inadequate for use in SEAWAVE-QEX, because measured data are not available or too highly censored.

Use of SEAWAVE-QEX in SBF computation has addressed a lingering problem in SBF use, i.e. that estimated SBFs may be highly uncertain, especially for sites that are not well represented by the small subset of sites with high frequency sampling (Mosquin et al., 2018; Vecchia, 2018).

One weakness of this approach is that here must be sufficient data at some sites to apply SEAWAVE-QEX. This includes at least 12 samples per year for three years with more than 25% of reported values above the censoring limit. In addition, flow covariate data are needed to impute daily concentrations with SEAWAVE-QEX with the highest degree of confidence. Thus, SBF is not a universal tool. It is likely there will be cases when there is insufficient data to apply SEAWAVE-QEX, and as a result use of the SBF approach that was described in the white paper may not be feasible. Another concern is that there appears to be very high variability in computed SBF values between sites and years for some compounds. This was well illustrated by examples in the white paper. High variability is a limiting factor for SBF use in risk assessments since computations may indicate significant potential for exposure where the risk is actually relatively small. This may confound risk assessments. Using the median across years as described in the white paper cannot recognize the loss of information of year to year variability. It is quite possible that some site characteristics will vary from year to year. Environmental changes within the watershed, such as changes in land use or changing climate conditions affecting water balance, soil moisture and vegetation, etc., might add to the year-to-year variation and account for some of the year to year variability in SBF. If these changes are systematic, they have the potential to render a site-specific SBF less and less effective with time.

Regarding EPA's optimization study, it was agreed that the study was performed properly and that conclusions from the statistical analysis of the results were correct, reasonable, technically sound, and a useful path forward for development of both short and long-term SBF. Of note was the use of medians as opposed to averages to reduce bias due to high variability across years and sites and the use of a random sampling strategy with log-linear interpolation to estimate the range of potential concentrations not measured between sampling events. Because of the way SBFs are conceptualized, as a ratio of the true value divided by a statistic derived from subsampled monitoring realizations, any statistic could be used in the denominator. The median is a logical choice for denominator statistics since it has better statistical properties than any other percentile (strength). In the end, the best justification is that it works when tested with the best data available, which is what is presented in the white paper. Use of log-linear interpolations was in agreement with a recommendation by Mosquin et al. (2018). Other features of the data analysis that were highlighted were use of the root mean square error (RMSE) to minimize prediction error and findings which showed that computed SBFs using SEAWAVE-QEX captured the range of SBFs computed directly from datasets where high-intensity sampling was conducted. Further, in the white paper there was little difference in analysis conclusions between the short-term SBFs and long-term SBFs. The Panel agreed with this conclusion and that analysis models and results were reported to be similar to that reported for RMSEs from short-term SBFs. Hence, comments related to the short-term were likely valid for the long-term SBF analysis.

In sum, the Panel agreed that site-specific Sample Bias Factors can be valuable in some cases to inform an existing sampling program at a given site to increase the amount of monitoring data that may be used in risk assessment. This can make sampling at a given site more

effective going forward or can be used in designing a sampling plan at sites that are sufficiently comparable or similar to the ones use in testing and evaluating SBFs.

While the Panel generally supported EPA's study and conclusions derived from it, there were concerns about the model used. One Panel member proposed a different statistical analysis model. The model specifies site and yar within site as random effects that is more in line with the discussion and conclusions in the white paper. A recommendation was also made to revise the white paper to include better and clearer discussions and summarization of analysis results.

In discussion of the alternate model, sections 4.4 and 10.1.1.2 of the EPA white paper were referenced. It was observed that a simulation study was described which examined the impact of four factors −Exposure Duration, Sampling Frequency, Sampling Strategy and Interpolation Method −on the ability of the SEAWAVE-QEX chemographs (time series of daily chemical concentrations) to produce short-term (1-day, 4-day and 21-day rolling average maxima) SBFs with acceptable properties. Acceptable was operationalized by looking for combinations of these factors that minimize the 50th and 90th percentiles of the RMSEs from the simulations. This analysis was performed for each sampling interval and exposure duration separately, so the analysis focus was on estimating differences in Sampling Strategy and Interpolation Method. A Generalized Linear Mixed Model (GLMM) assuming log-normal error was assumed with SAS code for implementation included.

A recommendation was made to add comments to the SAS code to inform readers of the purpose of each block of code in support of the analysis. In addition, a Panel member noted that with 50 realizations per condition, the residual term had very high degrees of freedom resulting in Sampling Strategy and Interpolation Method effects always being significant. As a result, it was suggested that the analysis should discuss the magnitude of differences in RMSEs for different sampling strategies and interpolation methods and the covariance terms related to site and year*site random effects. Further comments were that there are no differences in RMSE for SBFs from random versus stratified sampling protocols, but that log-linear interpolation provided superior results (smaller average RMSEs?) compared to linear interpolation. To this end another Panel member observed that the white paper does not present result summaries in support of these conclusions (an analysis summary for example). EPA was urged to address this and to provide normal quantile plots and a log-normal quantile plot of regression residuals to support the appropriateness of the assumption that a log-normal distribution was appropriate for the distribution of the median RMSE since a more Gaussian distribution would be expected. The log-normal distribution assumption might be more acceptable when applied in the analysis of the 90th percentile of a RMSE statistic.

Some concluding comments were that in the analysis model used, instead of separating out site and year*site random effects, the analysis should consider site and year within site as random effects, and compute and discuss related covariance terms. In this context, the Panel members highlighted the importance of discussing whether site-to-site variability in RMSE is different from year within site variability. It was suggested that the specification for a site random effect should assume independence among sites with common variance (identify structure), and the random effect year within site should have correlation structure that accommodates temporal autocorrelation, similar to an AR(1) [first order autoregressive] or CS (compound symmetry) structure. Additionally, it may be useful to examine whether variation

among years within site is constant across sites or whether year within site variability differs by site. The fact that the EPA white paper recommended that the highest annual SBF will likely be used when there is high variability from year to year in SBFs for reference sites further supports the specification of site and year within site as random effects needing detailed analysis and discussion.

As for the long-term SBF derivation, one Panel member noted that no parametric form was used to infill the bootstrapped data. The same Panel member expressed concern with the use of the formula for the standard error of the mean used to compute the average annual pesticide concentration from the bootstrapped data in the long-term sampling bias calculation. While it is true that the sampled data is very likely to be independent, the true daily pesticide concentration data is correlated. This means that the standard error used to compute the upper limit of the 90% prediction interval for the mean is likely too small, and it does not capture the sampling variability in the estimated annual mean pesticide concentration. The consequence of this correlation is that the upper limit of the 90th prediction intervals are likely to be smaller than they should be making the long-term SBFs larger than they should be. However, the Panel member noted that larger long-term SBFs, as also pointed out in a public comment, will err on the protective side leading to a larger rate of Type 2 error or false positives than by pure chance. It was agreed that EPA may want to address this matter, however there was consensus that since long-term SBFs are used for screening, it is unlikely that the concerns expressed bear significant consequences in SBF use.

A final limitation of computing SBFs using SEAWAVE-QEX simulations that was discussed was that it is time-consuming, complex, and requires high level technical expertise. Thus, use only seems feasible at the highest tiers in risk assessments. Other concerns expressed were the need for datasets that meet minimum SEAWAVE-QEX requirements. This may substantially limit the amount of monitoring data which can be used. In addition, for the reasons described in responses to charge questions 1.c. and 1.d., use of covariates such as precipitation or stage in place of flow when flow data are unavailable and applications to non-flowing systems may be problematic. It is likely that this will place additional demands on risk assessors who must determine if SEAWAVE-QEX applications to static systems and or surrogate covariate use in SEWAVE-QEX effectively reflects available data. To increase confidence in results, one Panel member recommended that SEAWAVE-QEX should only be used for flowing systems and when flow data are available.

Recommendation:

The Panel recommended further work identifying sources of year-to-year and between site variability. As an intermediate term task, Panel members believe that EPA should at least look into the potential for doing this study. If there is little to no information available on year to year variation in site attributes, then this study may not be feasible. Nevertheless, this should at least be mentioned in the white paper.

Panel Response 2b:

      b.   The Panel agreed that use of computed site specific SBFs has the potential to increase the amount of monitoring data that may be used in risk assessments. This is the primary

strength of the approach. A weakness is that there must be sufficient data at some sites to apply SEAWAVE-QEX. This includes at least 12 samples per year for three years with more than 25 % of reported samples above the censoring limit. In addition, flow covariate data are needed to impute daily concentrations with SEAWAVE-QEX with the highest degree of confidence. Thus, SBF is not a universal tool. It is likely there will be cases when there is insufficient data to apply SEAWAVE-QEX and as a result use of the SBF approach that was described in the white paper may not be feasible. Another weakness is that there appears to be very high variability in computed SBF values between sites and years for some compounds. This was well illustrated by examples in the white paper. High variability is a limiting factor for SBF use in risk assessments since computations may indicate significant potential for exposure where the risk is actually relatively small. This may confound risk assessments. Further work identifying sources of year-to-year and between site variability is recommended. One factor to consider at the watershed level is the relative distributions of base versus storm water induced surface flow. Bosch et al. (2017) have provided guidance for hydrograph separation and how base versus surface storm water flow may impact contaminants.

There was also consensus that using the median across years of the medians computed from equally likely conditional traces makes both practical and statistical sense. However, because of the way SBFs are conceptualized, as a ratio of the true value divided by a statistic derived from subsampled monitoring realizations, any statistic could be used in the denominator. The median is a logical choice for denominator statistics since it has better statistical properties than any other percentile (strength).

The Panel also recognized that a weakness with using the median across years comes from the loss of information of year to year variability. It is quite possible that some site characteristics will vary from year to year and this may provide some explanation of year to year variability in SBF. As an intermediate term task, Panel members suggested that the Agency should at least look into the potential for doing this study. If there is little to no information available on year to year variation in site attributes, then this study may not be feasible. Nevertheless, this should at least be mentioned in the white paper. A final comment was that in the end, the best justification is that it works when tested with the best data available, which is what is presented in the white paper.

Panel Response 2c:

       c.   The Panel agreed that use of SEAWAVE-QEX to compute the SBFs for the four pesticides was technically sound, but application of the SBFs for these reference pesticides to estimate upper-bound concentrations of other pesticides was not recommended without further analysis of the magnitude of uncertainties that this approach may bring to risk assessments. One Panel member noted that the maximum SBF at the USGS sites for carbaryl, atrazine, and chlorpyrifos appeared to provide protection for metolachlor when compared to high-frequency SBFs at Heidelberg University's National Center for Water Quality Research sites (Figure 10.8 in the EPA white paper, page 181). However, the Panel agreed that more information is needed on sources of variability in SBFs, which vary among sites, over time, and among chemographs, as well as among pesticides. One Panel member noted that long-term SBFs for atrazine had high variability during some site-years, across years, and across sites (Figure 10.11, on page 187); carbaryl, chlorpyrifos, and fipronil (Figures 10.12-10.14, on pages 188-190) showed less

variability among years, but still showed variability between sites. Overall, Panel members agreed that four reference pesticides were too few to draw conclusions on the role of physical/chemical properties or pesticide class and recommended developing SBFs for additional pesticides to define the magnitude of uncertainties in this approach.

As indicated in the EPA white paper, use of environmental fate properties such as the variability of substance degradation, K_{oc} or sorption, DT_{50} and watershed characteristics did not seem promising as predictors of SBF. Panel members observed that EPA appropriately concluded that "additional research is needed before SBFs can be confidently predicted using watershed characteristics and/or fate and transport properties of the pesticides under review" (EPA white paper, page 80). However, several Panel members recommended further evaluation of SBFs in relation to pesticide fate properties, classes, use characteristics and other factors, which may support future application of the method to estimate SBFs for other pesticides. One Panel member suggested using data for atrazine to estimate SBF for other pre-emergence herbicides, data for carbaryl and chlorpyrifos to estimate SBF for other foliar applied insecticides, and fipronil to estimate SBF for insecticides applied primarily in urban settings. Another Panel member suggested that some basis for application of SBF computed for one of the four compounds to other pesticides may be obtained by comparing SEAWAVE-QEX WAVE shapes. Possible metrics include amplitude, width at half height, or a quantitative assessment of peak tailing. Similarity would indicate similarity in watershed scale dissipation processes. It was also intriguing that a positive although weak correlation was observed with terrestrial field dissipation half-life. Further assessment in this case appears warranted along with an evaluation of whether or not similarity in use profile would provide some basis for extending SBF from one compound to another. For example, this could include compiling data on a group of herbicides that are applied pre-emergence and comparing findings to the computed atrazine SBF.

One Panel member recommended examining the USGS WARP for insights into the basis for extending SBFs between pesticides and watersheds (Stone et al., 2013). In applications of WARP, similarity in model performance between pesticides was linked to similarity in pesticide use as well as application timing and methods. Research evaluating how conservation practices may impact nutrient losses from farm fields has indicated that in the upper Midwest, a high density of subsurface drainage structures (tile drains) could be an important factor for watershed comparison (Garcia et al., 2016). Because this may also apply to selected pesticides, one Panel member recommended that EPA examine this research and determine if this watershed characteristic may assist in extending pesticide SBFs.

One Panel member noted that EPA attempted to incorporate vulnerability of land in the analysis and in the selection of an SBF. The same Panel member proposed attributes of the watershed or pesticide in addition to watershed attributes listed in table 11.2 of the EPA white paper (page 201):
- Dominant hydrologic soil group of land likely to be treated with pesticide of interest.
- Similarities between land cover to which pesticide is applied, including percentage area in the watershed, as well as type and management of the land cover. Try to be more specific than "ag. land". Perhaps land in the crop(s) relevant to the pesticide.
- Slope of land likely to be treated with pesticide of interest.
- Temperature (because it affects product degradation).

- Rainfall amount or intensity? (because it affects runoff/rainfall ratio).
- Shape of the watershed (longest flow path length). Travel time is already calculated.
- Presence of sub-surface artificial drainage.
- Presence of a restrictive layer in the soil profile.
- Fraction of land treated with pesticide in high vulnerability. Several research groups have come up with vulnerability indices:
  - The Conductivity Claypan Index (CCI) identifies areas that are most vulnerable to sediment loss, and nutrient and herbicide transport in a claypan setting.
  - The Soil Vulnerability Index (SVI) is a new index proposed by the USDA National Resources Conservation Service (NRCS) to identify areas vulnerable to environmental loss when cultivated.
  - The Topographic Index identifies areas most likely to be saturated when it rains in a setting characterized by permeable soils overlaying a restrictive layer.
- Bio-chemical properties of the pesticide (such as water solubility, half-life).

The same Panel member also found that the analysis presented in section 11 (Appendix D, page 196) was useful to identify relevant factors, but asked EPA to consider whether the model should be multiplicative.

Panel members noted a few additional concerns and limitations about applying the SBF approach in DWAs. Two Panel members pointed out that the NCWQR sites used to determine the high-frequency SBFs are all located in the Midwestern U.S., where the environmental setting is primarily agricultural land; the principal crops are corn, wheat, and soybeans, and tile drains are common. Because more variability in SBFs may occur across environmental settings, this suggests uncertainty as to whether the maximum SBF values for atrazine, carbaryl, and chlorpyrifos determined from these Midwestern sites would be protective for streams in other environmental settings. One Panel member commented that there was greater variability in SBFs at USGS sites, which are located across the U.S., compared to SBF values within the Midwestern U.S. at NCWQR sites, although some of the additional variability in SBF values at the USGS sites may be due to the sparse datasets on which they were based. One concern was noted that small watersheds with "flashy" hydrology may provide unreasonably high SBFs. Several Panel members raised a concern that it does not appear SEAWAVE-QEX can be used confidently to describe monitoring data from static non-flowing systems. Thus, an alternate means of calculating SBF for these systems may be needed.

Panel Response 2d:

     d. The Panel agreed that the white paper (section 4.61) combined with the two case studies provided a useful background on how SBF may be used in risk assessments. In sum, the quantitative use of SBF in both long and short assessments appears to be reasonable for screening and that upper-bound estimates derived from SBF computations are likely conservative.

There are two primary concerns in SBF application in risk assessments. First, computed SBF exhibit high variability from year to year and across sites. Better understanding and further refinement of factors that contribute to SBF variability would increase their utility in risk assessments. Second computed SBFs are pesticide specific. It was suggested that selected

SBF could be applied to other pesticides with similar use profiles and properties. EPA has described a series of investigations in this regard (EPA white paper section 11) and concluded that comparative studies that demonstrate this did not provide defensible results. One Panel member suggested chemically based comparisons could be enhanced by comparison of WAVE shapes that were fit during SEAWAVE-QEX implementation. Metrics such as the amplitude and peak width at half-height of WAVES could be used as a basis for comparison. Similarly, in-WAVE characteristics could improve confidence in applying chemical specific SBF to other compounds.

Recommendations:

Chemically based comparisons could be enhanced by comparison of WAVE shapes that were fit during SEAWAVE-QEX implementation. Metrics such as the amplitude and peak width at half-height of WAVES could be used as a basis for comparison. Similarly, in-WAVE characteristics could improve confidence in applying chemical specific SBF to other compounds.

Panel Response 2e:

   e. The Panel agreed that SBFs could be of significant value in designing new or improving existing water monitoring programs, primarily in specifying sampling intervals and schedules for flowing water sources. Using SBFs to reduce required sampling for pesticides of concern for long-term exposure and chronic health impacts seems the most appropriate situation for applying SFBs for this purpose.

Sampling Bias Factors that were developed using the four chemicals and USGS sites that represent a range of environmental and use conditions across the landscape, are being proposed for use as a screening-level analysis at the Tier 3 level of a DWA. To obtain chemographs for development of SBF, non-daily monitoring data need to be interpolated with SEAWAVE-QEX; this requires the acquisition of covariate data, running SEAWAVE-QEX, and reviewing the model fits in the diagnostic plots. Sampling Bias Factor Python programs then have to be run to estimate SBFs and summarize the values for use in an assessment. Finally, a determination must be made of whether the land use or land cover characteristics of the watershed where the SBFs were developed are sufficiently similar and appropriate in comparison to the USGS sites, including relevance to DWIs and pesticide use areas.. Chemical-specific SBFs will rely primarily on the use of SEAWAVE-QEX model runs because in most cases daily monitoring data will not be available, so chemical-specific SBFs will be used to adjust concentrations at the Tier 4 level. Panel members also identified several design considerations that are of concern.

1. Choice of covariates: the EPA white paper states that while stage and precipitation may be acceptable, streamflow data have proven to be a more appropriate covariate when using SEAWAVE-QEX (page 44). In exploring alternative covariates for a given pesticide occurrence dataset when streamflow is either not available or not appropriate (i.e., periods of backflow) daily stage data was found to be a useful alternative when available for typical flowing systems. Measured precipitation was also determined to improve model fits over the modeled precipitation (page 152).

52
RX 13 Page 54 of 82

Because stream stage data performed very similarly to stream flow for the WSDA sites, and since flow is typically derived from stage; stage is likely a viable alternative covariate for sites that have typical (positive) flow patterns. If an assessor does not have streamflow data or stage data available, the suitability of using modeled precipitation should be considered in light of the results of this first-pass qualitative analysis (i.e., SEAWAVE-QEX).

2. Sampling duration: a minimum of three years was chosen for use with SEAWAVE-QEX, even considering the advantages of programs with three to five years duration (page 98). Vecchia (2018) indicates that three years are sufficient to produce estimates of the annual maximum daily concentration (AMDC) if the assumptions of the SEWAVE-QEX model are met. However, verification of the assumptions is more difficult in cases with short sampling record, sparse sample frequency, or more highly censored data (page 19). The 3-year time period does not have to be consecutive, but there is no indication as to accepted interval between non-consecutive years, or the effect on concentration estimates.

3. Estimation of pesticide concentration for acute effects: in 2010 a FIFRA SAP concluded "if inference is to be at the daily level, then sampling needs to be at least daily, for four day rolling averages at least two samples are required within that period". This is most critical when EPA considers acute effects resulting from single day exposures, or other exposures of potentially high concentration and short duration. Most CWS monitoring programs are sampled quarterly and are not designed to generate results for short-term concerns, SEAWAVE-QEX can generate daily estimates of these pesticide concentrations. These estimated concentrations compare well with measured pesticide levels when available, although there is a tendency for underestimation.

4. Estimation of pesticide concentration for chronic effects: Chronic and cancer risk assessments consider long-term exposure concentrations, such as a 365-day average or the average concentration over a lifetime of exposure (currently defined as 30 years due to the availability of weather data for aquatic modeling). As there are no datasets that contain 30 years of pesticide monitoring, SBF-adjusted 365-day average concentrations can be developed using as few as four samples per year to estimate upper end long-term average concentrations. SBF-adjusted 365-day average concentrations considered over the number of years available for a site will also provide an upper-bound estimate, that can be used in lieu of a 30-year, 365-day average estimate.

Charge Question 3. – *SPATIAL RELEVANCE OF MONITORING DATA AND SAMPLING BIAS FACTORS WITH WATERSHED AND PESTICIDE CHARACTERISTICS*

To meet EPA's drinking water protection goals, monitoring data used in drinking water assessments should be relevant to drinking water intakes. EPA developed two methods to investigate the spatial relevancy of monitoring data and sampling bias factors. EPA developed sampling bias factors for four pesticides with a range of use profiles, physical-chemical properties, and environmental fate properties, using occurrence data from sampling sites

across the contiguous United States. SEAWAVE-QEX was used in the sampling bias factor development for these sites, which have different watershed or catchment attributes. The sampling bias factors were related to watershed and pesticide attributes to develop regression equations to estimate sampling bias factors for watersheds where there are not enough data available to develop these factors. Additionally, a weight-of-evidence approach was developed to assess how relevant sampling bias factors or monitoring data concentrations are to drinking water intakes.

3.a. Please discuss the suitability of the underlying data, methods, and parameters used to develop watershed regression equations for estimating sampling bias factors for systems with limited data. Please comment on EPA's conclusion that the short-term sampling bias factor regression equations for the four pesticides evaluated, as well as the environmental fate properties regression analysis, provide minimal predictive ability for estimation of sampling bias factors for other sites. Discuss approaches and the value in continuing to investigate quantitative relationships of sampling bias factors with watershed and pesticide characteristics.

3.b. Please comment on the weight-of-evidence approach to determining spatial relevancy of monitoring sites to source drinking water. Are there additional factors EPA should consider in this approach? Please discuss the relative importance of the factors, considering data availability and quality.

Panel Response 3a:

a. Even though from a conceptual point of view, the Panel agreed that it made sense to relate SBFs to watershed attributes, based on the performance of the regression models, the Panel agreed with EPA's conclusion that the short-term SBF regression equation for the four pesticides evaluated provide currently have low level of predictive ability and thus should be not used for drinking water evaluation without further improvement. The Panel recommended alternate datasets and statistical approaches that may produce better predicting equations of SBF variability.

To understand which watershed attributes, influence a site SBF, the Agency used a traditional multiple regression analysis approach with 60 potential predictors initially proposed (table 11.1, EPA white paper). Through some pre-processing including inspection of correlations and collinearity among predictors, contingency tables, principal components analysis, and inspection of the degree of variability in each variable, EPA reduced the number of predictors to 36 variables that were used in the final stepwise-regression analysis. Regression modeling was only done for the one-day SBFs generated under 14-day sampling intervals. The final models for the one-day SBFs generated under 14-day sampling intervals included one to two predictors and yielded R-square values ranging from 0.1 to 0.26.

While stepwise regression is a traditional choice for building a predictive model, the Panel believed that more recent approaches and, in particular, regression tree methods (bagging, boosting, and random forests), could be employed and produce more insightful results. The Panel proposed many different alternative approaches:

Approach A: Regression trees

Regression trees are predictive models that predict the value of the outcome variable by splitting several input variables (Breiman et al., 1984; Gareth et al., 2015). A Panel member proposed that EPA considers regression tree for predicting SBFs rather than the linear regression models presented in the paper. To illustrate the improvement in predictive ability of regression trees, the Panel member presented the changes in R-square that simple regression trees would yield in the case of the four pesticides discussed in the white paper.

More specifically, as reported in the white paper the final regression *for atrazine one-day SBFs with 14-day sampling interval*, is given by;

$$Y = 0.084 + 0.018 \text{ x } pmayjuneCat + 0.064 \text{ x } PctAg2006Cat$$

with an R-square of 0.26.

The Panel member showed that a simple (complexity-pruned) regression tree model for the atrazine data yields the classifier model below with full details of the tree provided in Table 1. The R-square for this simpler model is 0.33. The final model is:

$$Y = \begin{cases} 4.19 & if & AgKffactCat < 0.145 \\ 7.56 & if & AgKffactCat \geq 0.145 \text{ and } PctHay2006Cat \geq 1.163 \\ 17.12 & if & AgKffactCat \geq 0.145 \text{ and } PctHay2006Cat < 1.163 \end{cases}$$

While this is just a simple illustration of a complexity-pruned regression tree, the Panel member noted that it is possible that other regression tree would yield even better results. As the equation of the regression tree model indicates, the regression tree first split the data based on the mean soil erodibility (Kf) factor of soils on agricultural land in the catchment. After that, for atrazine, the next best split occurs for factor PctAg2006Cat < 34.16478, the sum of the percentages of two categories (PctCrop2006 and PctHay2006) for the catchment. Although for this particular regression tree the data is split based on mean soil erodibility and percentage of Hay, other regression trees could consider other splitting factors.

Node	Split	n	Deviance	Mean	Std Err
1)	root	98	3378.943	5.91	5.87
2)	AgKffactCat< 0.1448338*	68	391.067	4.19	2.40
3)	AgKffactCat>=0.1448338	30	2334.802	9.79	8.82
6)	PctHay2006Cat>=1.163472*	23	628.581	7.56	5.23
7)	PctHay2006Cat< 1.163472*	7	1215.891	17.12	13.18

* denotes terminal node

Table 1. Summary of complexity-pruned regression tree for atrazine data attempting to predict the 1-day SBFs with 14-day sampling interval (personal communication, Portier, K., 2019).
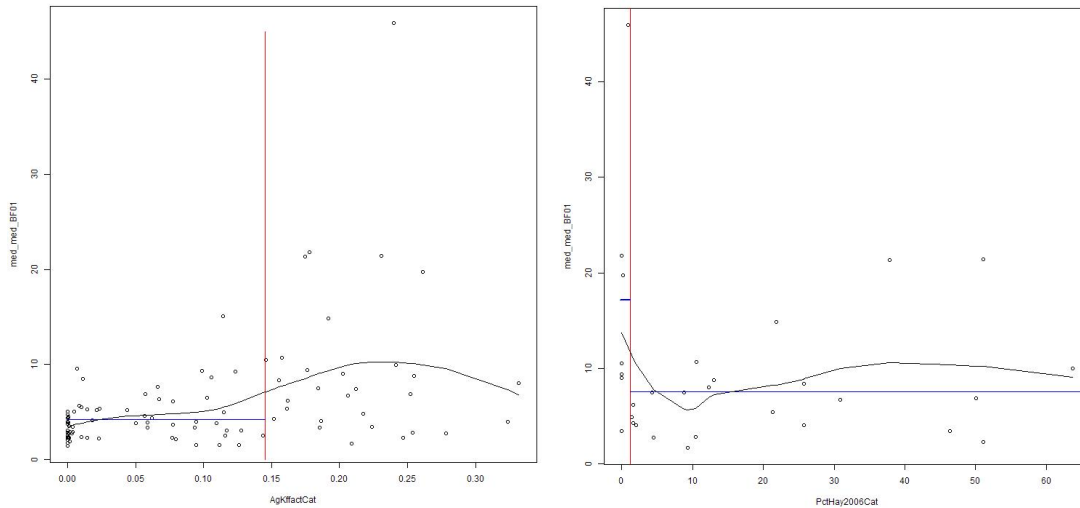
For *carbaryl one-day SBF with 14-day sampling*, the final regression model in the white paper is Y = 69.72 – 0.75 x *curve_numberCat* which yields an R-square of 0.13. Also, the following regression tree model uses *curve_numberCat* as factor but the model states that

$$Y = \begin{cases} 10.98 & if \quad curve\_numberCat < 65.26457 \\ 30.66 & if \quad curve\_numberCat \geq 65.26457 \end{cases}$$

and yields an R-square of 0.20. The regression tree model above fits the data as well as a regression tree model that uses the percent of watershed area classified as woody wetland land cover (NLCD 2006 class 90) with *PctWdWet2006W*s < 0.4427415.

For *chlorpyrifos one-day SBF with 14-day sampling*, the final regression model presented in the white paper is Y = 1.62 + 0.055 x *pmayjuneCat* with an R-square of 0.10. The following regression tree model uses *PctAg2006Ws* (percent crop and hay in watershed) leading

$$Y = \begin{cases} 4.829 & if \quad PctAg2006Ws < 45.5 \\ 18.53 & if \quad PctAg2006Ws \geq 45.5 \end{cases}$$

to an improved R-square of 0.194. The regression tree model above fits the data as well as a regression tree model that uses the percent of watershed area classified as crop land use (NLCD 2006 class 82) with *PctCrop2006Ws* < 30.30783.

For *fipronil one-day SBF with 14-day sampling*, the final regression model presented in the white paper is Y = 4.03 - 0.40 x *satof48Cat* with an R-square of 0.21. The following regression tree model uses *RunoffCat* (mean runoff (in mm) within the catchment).

$$Y = \begin{cases} 4.174 & if \quad RunoffCat < 289.5 \\ 2.212 & if \quad RunoffCat \geq 289.t \end{cases}$$

And provides an improvement in terms of R-square, now equal to 0.41. The regression tree model using *satof48Cat* < 2.139895 provides a much poorer fit to the data.

Approach B. Median regression
The linear regression models proposed in the white paper are all applied to predict the median of the SBFs derived at each site.
As the distribution of the SBFs is typically non-symmetric and right-skewed, a Panel member suggested the use of quantile-regression rather than linear regression. The Panel member motivated the suggestion by pinpointing the fact that linear regression models are used to estimate the mean of a distribution, and in the case of skewed distributions, as is the case of the SBFs distributions, median and means do not coincide.

Approach C. Principal component analysis
A Panel member suggested that to improve the predictive power of the linear regression models for SBFs, the Agency could consider applying principal components analysis (or exploratory factor analysis) on the predictors to generate new factors that could be better predictors than the individual watershed properties.

Approach D: Change the outcome variable
A Panel member noted that given that SBF is a ratio of two quantities, an alternative approach might be focused on attempting to predict the numerator of the SBF ratio with the denominator of the SBF itself a covariate used in the model along with all the other watershed attributes. In other words, the Panel member suggested that the Agency tries to predict the true max as a function of the median-of-the-median value. The Panel member noted that the relationship may not be exactly linear and that plots with a loess smoother superimposed would help to determine what kind of relationship is expected.

Besides comments on alternative statistical modeling approaches that EPA could attempt, the Panel provided several comments regarding the list of predictors that EPA should use in a regression modeling approach. These comments would apply also if the Agency intends to pursue one of the approaches suggested above.

Specifically, Panel members highlighted the following points:

1. Pesticide usage data: Pesticide usage is perhaps the most important variable in determining pesticide concentrations in surface water. Although obtaining pesticide usage data for areas outside of California might be difficult, adding this variable could improve the regression analysis. The Agency could potentially use the USGS estimated county-level usage data. In addition, there is an ongoing effort within EPA to generate nationwide pesticide usage data. The Agency's Environmental Modeling Public Meeting on October 16th, 2019 features the discussion on obtaining pesticide usage data.
2. Precipitation: Precipitation is another important parameter of pesticide concentration as indicated in the white paper. The Panel agreed with EPA that further analysis should use annual precipitation data instead of 30-year average. In addition to precipitation, in arid

and semi-arid regions, irrigation should be included. A potential approach is to add estimated irrigation to precipitation as total water input. Antecedent events like droughts and rainfall frequency could also improve predictability.

3. Climate data: A Panel member suggested that the Agency performs the analysis to predict the median SBF for a given pesticide in any year rather than summarizing the SBFs across time at each site, and multiple Panel members advised against the use of climate variables averaged across a 30-year time period as predictors in a regression modeling framework as these variables are constant across the time period considered.

4. Alignment of timing between independent variables and SBFs: The Panel agreed with EPA that future work should try to align the timing of the pesticide sampling with that of the independent variables.

Various members of the Panel pinpointed how SBFs for a given pesticide in a given year are used to estimate the upper prediction limit for a summary statistics of the pesticide concentration data. Thus, in itself SBFs provide more of a measure of the uncertainty or variability in the pesticide concentration data than a measure of the expected pesticide concentration level. For example, as shown in the white paper, "flushy" or "flashy" systems such as the Honey Creek and Rock Creek tend to have higher SBFs. This is also in line with the white paper's finding that terrestrial field dissipation half-life is the most significant explanatory parameter among all pesticide properties explored. In light of that, some members of the Panel encouraged the Agency to focus more on understanding which characteristics of a given sampling site, such as the timing of use (relatively constant input over the year vs. sporadic use), flow rate, variability in precipitation within a year, pesticide degradation etc., and which characteristics of a particular year might explain the variability observed in the daily pesticide rather than focusing on identifying which factors affect pesticide concentration. Subject matter experts might be helpful in identifying such variables. Exploratory regression analyses to determine what watershed parameters predict maxima or predict width of 80% confidence intervals could be useful in this regard (intermediate priority task).

In 2012, the Panel was presented with an extensive GIS modeling analysis done by Syngenta Corp. on atrazine concentrations across the Midwestern US. A long-term research project might follow this line of analysis and perform an analysis separately for every year, tackling the problem as the geostatistical problem of predicting median SBFs over space for any given pesticide. The underlying motivation for this approach lies in the understanding that the pesticide concentration observed at a monitoring site is likely more highly correlated to areas that are geographically closer than broad watershed characteristics, especially for monitoring site that are located somewhere other than at the exact output of the watershed.

For this type of analysis, spatial statistical modeling approaches (i.e. spatial regressions where the outcome is the SBF for a given year at all sites for which SBF are available and predictors are some of the predictors that are expected to influence the degree of variability in pesticide concentration data) could be employed (Cressie (1993), and Banerjee et al. (2014) for overview of spatial statistical modeling approaches). The covariates to include in the spatial regression could be identified through variable selection or they could be summarized by creating PC factors via PCA. An initial assessment as for whether spatial regression approaches could be useful in this regard could be done by fitting a linear regression model on the SBFs for a given year at all sites with SBFs available and computing an empirical semi-variogram of the residuals. If the semi-

variogram indicates no residual spatial dependence in the SBFs, then a geostatistical approach might not be warranted.

The Panel recognized that this line of research would represent an expensive and time-consuming alternative with no guarantee of satisfactory return.

Finally, one Panel member emphasized that the data considered in the white paper referred only to four pesticides including three insecticides −carbaryl, chlorpyrifos, and fipronil −and an herbicide; atrazine. The Panel noted that the mode of application and use patterns of these products differ widely. The herbicide is typically applied to bare soil prior to planting and/or crop emergence, whereas the two insecticides; carbaryl and chlorpyrifos, are sprayed on crops and applied directly to plant canopies, and the third insecticide, fipronil, is used primarily in urban settings. Use and application patterns may strongly impact potential for entrainment and loss of pesticides in surface runoff from treated areas. Given this, the poor outcome of the regression analysis was not surprising to the Panel member. If data from these pesticides are used to predict SBF for other compounds the Panel member suggested grouping data by use category. For example, the atrazine SBF could be used to predict SBF for other soil applied herbicides, while carbaryl and chlorpyrifos could be used to predict SBF for products sprayed on growing plants and plant canopies. Finally, the fipronil SBF could be used for insecticides used in urban environments. The Panel recommended that for broad application this work be extended to other compounds in these, as well as other use categories. The Panel also highlighted the immediate need for investigations that evaluate the SBF of a fungicide that is applied directly to crops.

In conclusion, regarding the short-term SBF regressions, the Panel agreed with the EPA's conclusion that the short-term SBF regression equations for the four pesticides evaluated, as well as the environmental fate properties regression analysis, provide minimal predictive ability for estimation of SBFs for other sites. Previous studies have indicated the lack of relationship between stream flow, rainfall intensity and volume and concentrations of materials such as pesticides entrained in runoff (Leopold et al., 1964). Small rainfall events (low intensity and volume) may yield no runoff, so mobilization of pesticides by these events does not occur or it is relatively minimal. Large rainfall events usually produce large volumes of runoff for longer durations. These large rainfall events can mobilize particulates as well as dissolved and sorbed pesticides that can be transported to water resources and subsequently to drinking water intakes. Rainfall events producing the highest pesticide concentrations are often intermediate, with a relatively intense initial pulse. These rainfall events producing the highest pesticide concentrations are often preceded by a prolonged dry period with no rainfall during the season when pesticides are applied.

The suitability or adequacy of the underlying data, methods, and parameters used to develop watershed regression equations for estimating SBFs for systems with limited data is an open question. The pesticide monitoring data that could be used to evaluate accuracy or extrapolation performance are limited. Critical correlated factors such as streamflow, rainfall patterns, soil character, and application techniques, etc. are often missing from data-sets or are inaccurately measured. Therefore, SBFs may have considerable inherent uncertainty. The logic supporting this approach needs to be explored (links, cause and effect (i.e. dose-response), fidelity, predictability, what is the question, is this the best or most efficient way to obtain the answers). The ultimate

arbiter for this question is the accuracy in estimation of the exposure (concentration, duration, form, etc.) of a pesticide in a drinking water resource and the amount actually arriving at the source water intake and ultimately the tap and consumer.

If further analysis could result in models/relationships with higher predictive ability, such models would be very valuable for relating SBF from high intensity monitoring sites to low intensity monitoring sites. However, the results presented in the white paper and the details available through perusal of the supplemental materials appear to suggest that further investigation is unlikely to produce better models.

Summary and Recommendations:

1. The Panel concluded that the approach to relating SBFs to related watershed attributes does make sense. Options in performing the regression analysis that may produce fitted models with higher predictability have been suggested and illustrated.
2. The Panel agreed with EPA's conclusions regarding the low level of predictability of fitted regression models. Not much can be expected from predictor models with such low R-square values.
3. The Panel recommended alternate datasets and approaches may produce better predicting equations of SBF variability. The Agency should consider this task as having intermediate priority since high quality predicting models is not guaranteed, although potentially likely.

Panel Response 3b:

b. The Panel agrees that the Agency's proposed use of a weight-of-evidence approach is a creative and potentially useful approach for determining the spatial relevancy of monitoring sites to source drinking water; providing relevant monitoring data or SBFs to areas lacking them, searching for appropriate surrogates, and evaluating the possible utility of nearby monitoring detects at values of concern. Panel members raised a number of important issues that need resolution when implementing this approach.

A Panel member noted that potential downsides to this approach include the complexity and subjectivity of relationships between input metrics and their utility. It was also stated that results may not always be repeatable or transferable. As stated in the white paper section 5.3.4 *Pulling It All Together*, "[…] the assessor will need to use their judgement to weight the factors that give them confidence that the monitoring sites represent concentrations that may be anticipated at the drinking water intake and those factors that make this conclusion less (un)certain." Because of this, criteria and SOPs need to be clarified.

The determination of spatial relevancy of monitoring sites to source drinking water could be operationalized in a number of ways, such as clear guidance through a decision tree, or an (ArcGIS) ModelBuilder model that can be used to compare a given watershed to acceptable SBF watersheds, making this process simpler, standardized and less subject to error from unwise subjective decision making.

In the short-term, EPA needs to define terms so that they can be used consistently and be specific about the requirement for "a site that is spatially relevant" beyond what is currently in the white paper. The reasonableness or utility of this requirement should be demonstrated by analysis, or at least with a clear rationale.

Ranking the criteria used is a wise aspect of such an analysis but doing so is potentially complex. The ranking procedure and resulting use of the ranked watershed characteristics and spatial relationships in the analysis should reflect a logic for determining relative importance that is evidence-based, transparent, and can be consistently applied. The ranking logic should also be explicit in terms of the relative importance among indicators, and properly reflect it in quantitative terms, if possible. One Panel member noted that even the weight-of-evidence approach is qualitative, the criteria for making decisions could be quantified using specific numeric values. For example, in evaluating the distance between sample sites and drinking water intake points relative to half-lives, a ratio could be developed to reflect the probability of whether a pesticide would be detected at the drinking water intake sites.

The term weight-of-evidence itself implies that the lines of evidence would not necessarily be equal in terms of their influence on the analysis, so weighting of the indicators needs to be considered. Not weighting them is, in fact, a decision that the influence of each is equal but, as one Panel member pointed out, the lines of evidence when combined may provide more than additive weight-of-evidence. The white paper states that "The more similar the properties are, the more confidence the assessor has in applying the SBFs" which is why they call it *weight*-of-evidence. This is another way in which ArcGIS tools such as ModelBuilder could make analysis both easier for risk assessors and standardize the process for consistency and as a way to document the analytical process.

Summary and Recommendations:

As an intermediate goal, EPA can ease implementation of this approach by arranging for geospatial data that risk assessors and others can easily obtained. As a GIS, 2QAZ would be used by risk assessors to implement this approach, and EPA should provide both the necessary geospatial tools and access to a variety of data types. These data might be made available as part of the National Atlas (https://catalog.data.gov/dataset/national-atlas-of-the-united-states) or could be placed on an appropriate EPA download site to ensure that the data are available, updated appropriately, and that assessors nationwide use the same data.

In addition to the factors currently under consideration by EPA – pesticide use, proximity of the sampling site to drinking water intake sites, hydrologic connectivity, contributing area characteristics, weather and climate factors, and persistence/mobility of the pesticide chemicals – the Panel has some additional suggestions/comments for the Agency to consider. In addition to the data used in the white paper example (section 5.3) geospatial data suggested by the Panel that would be both useful to this process and are widely available include:

- Digital elevation models, preferably with hydrologic corrections;
- Watershed boundaries and stream reach data that can be queried to determine hydrological connectivity and streamflow distance.
- The National Hydrologic Dataset

- Watershed boundaries may not be accurate in areas of low relief, or where the elevation data has low spatial resolution, or in areas with significant landscape alterations and engineering that may have changed the hydrology of the watershed. These boundaries can be modeled on the area of interest if better elevation data is available either using ArcHydro within ArcGIS, or by online access to the SD Supercomputer Center via opentopography.org).
- Agricultural use and crop data (some states might have their own high-quality data of this type).
- Land Cover:
  - National Land Cover Dataset – the definitive nationally consistent dataset for land cover and land cover change, as it covers 2001-2016 updated every 5 years, providing opportunity to include land cover change into the analysis.
  - The USGS has published the U.S. conterminous wall-to-wall anthropogenic land use trends (NWALT) land use dataset which is based on the National Land Cover Database 2011 (NLCD 2011); this dataset records land use and land cover change over time, so that land use characteristics can be consistently compared from 1974 to 2012.
- Precipitation – as cumulative raster layers.
- Soil layers (such as SSURGO and STATSGO) with appropriate descriptive and physical characteristics.
- Pesticide application (available for California and some other states).
- Census block data to estimate potentially impacted populations.
- Community Water Systems boundaries.
- Locations of groundwater drinking wells – to estimate populations that may not be exposed to surface waters under consideration.
- The locations of all drinking water intakes and contributing upstream watersheds in an accurate and reliably updated geospatial data file is essential. Some currently available government datasets have large uncertainties regarding locations of drinking water systems (DWS). Accuracy and completeness of the geospatial data on drinking water intake sites is the most important element in the weight-of-evidence analysis, so the agency should consider error checking and validation in developing this dataset. As security concerns become more prevalent, accurate coordinates for DWS intakes may become more difficult to acquire.
- In addition to considering pesticide usage in the host watershed, upstream watersheds that might have contributed pesticide from upstream to the target watershed.
- Limnological morphometry such as shoreline development index (ratio of reservoir shoreline length to the diameter of a circle encompassing the same area – a measure of potential interaction and influence of shoreline activities on the reservoir water).
- Crops and crop area in watershed.
- Pesticide application – including number and type (method) of application.
- Pesticide-specific fate and effects properties.
- Soil properties (especially those properties affecting pesticide fate).
- Rainfall events – their frequency, pattern, interval for events, and juxtaposition with pesticide applications.
- Reservoir geometry (dendritic reservoirs and intake location).
- Edge of crop field management practices (i.e. buffer strips, wetlands, etc.).

One Panel member noted that the data used in all assumptions in the risk assessment including derivation of SBFs should be evaluated in a line of evidence and subsequently a weight-of-evidence approach. The source of data used and QA/QC as well as the assumptions in the risk assessment should be clearly stated. It is likely that SBFs are formulated differently for lotic systems and fundamentally differ from those developed for the drinking water resources of interest (i.e. reservoirs). Spatial relevance of monitoring data is likely a parameter currently used in a weight-of-evidence (line of evidence) approach to confirm that the data source has some likelihood of representing the drinking water resource of interest. Initially, a challenge to overcome will be the availability of data of sufficient quality to be useful in this assessment.

A Panel member noted that the United States extends across an entire continent with highly diverse climatic, environmental, and hydrologic conditions. Factors influencing pesticides at a site in the Northeastern US might not apply at all to a location in the Southwest US. It will be difficult for EPA to select factors that can uniformly apply across the US. If nationally consistent data is required, the expectations regarding this endeavor should be managed and realistic. Some of these data vary among states or watersheds in terms of variety, data quality and completeness, spatial resolution and other factors. If possible, EPA should not place national consistency above providing the best data available wherever it is available. This is not the case for EPA's air pollution screening tool for the Environmental Justice Screening and Mapping Tool (EJSCREEN), although the EJSCREEN development team might be a useful partner in this effort.

In addition to the data described above, there are web services currently available that provide high-resolution land use, irrigated area, field-level crop rotation, and associated management operations for user-defined polygons. Land Use and Agricultural Management Practices web-Service (LAMPS; Kipka et al., 2016) uses annual crop type (derived from Landsat satellite images), provided by the USDA-National Agricultural Statistics Service Information (NASS) crop data layer from the USGS irrigation map derived from 250-m resolution MODIS satellite data, and management operations (dates of planting, harvest, and tillage with associated depth and intensity) from the USDA-NRCS Land Management Operations Database (LMOD; David et al., 2014). Land Management Operations Database information is relevant at the field scale, if approximate field boundaries are provided, using representative management for the spatial conservation management zone. So, one can automate quantification in space and time of management, and a new component may relate these data to the space-time distribution of pesticide application for any area of interest. At coarser spatial resolutions, LAMPS may be used to determine annual series of the fractions of lands within each area. For example, at the county level, a recent study answered the question, "Where is the USA Corn Belt, and how is it changing?" (Green et al., 2017).

In the short term, LAMPS can be explored as an option to enhance the weight-of-evidence approach while new analysis tools are being developed. The source code for LAMPS is available in an open-source repository (Kipka et al., 2016), and the developers offer a Jupiter Notebook for LAMPS to demo its deployment and application.

Charge Question 4. – *DRINKING WATER ASSESSMENT CASE STUDIES*

In general, EPA relies primarily on model-estimated pesticide concentrations for drinking water assessments, with limited use of surface water monitoring data for most pesticides. To improve transparency and support communication with stakeholders, OPP developed a DW framework that describes OPP's longstanding peer-reviewed tiered approach to drinking water assessments. The DW framework describes the continuum of approaches from highly-conservative and simple to highly-refined, complex temporal and spatial assessments. EPA applied the DW framework to two cases that represent pesticides with available surface water monitoring data, toxicity endpoints typically evaluated by OPP [i.e., 1-day (acute), 365-day (chronic), and 30 year (cancer)], different use patterns, and different environmental fate and transport properties.

> **4.a.** Please comment on the clarity and organization of the DW Framework.

> **4.b.** In case study 1, EPA demonstrates the implementation of the DW Framework and specifically the use of SEAWAVE-QEX and bias factors to analyze monitoring data for a pesticide with acute risk concerns. Please comment on the SEAWAVE-QEX analysis used to assess sites that had potential drinking water level of comparison exceedances. Please comment on EPA's use of chem1 short-term sampling bias factors to adjust chem1 concentrations for comparison to the potential drinking water level of comparison.

> **4.c.** In case study 2, EPA demonstrates the implementation of the DW Framework and specifically the use of SEAWAVE-QEX and bias factors to analyze monitoring data for a pesticide with chronic and cancer risk concerns. Please comment on the SEAWAVE-QEX analysis used to assess sites that had potential drinking water level of comparison exceedances. Please comment on EPA's use of chem2 long-term sampling bias factors to adjust chem2 concentrations for comparison to the potential drinking water level of comparison.

Panel Response 4a:

a.   Several Panel members described the Framework as generally clear, well organized, and appropriately tiered, with the most conservative assessments conducted at the lowest tiers (1 and 2). A similar approach is used in other well-established systems of pesticide risk assessment. This provides an increased level of confidence in the Framework. One Panel member commented that the introduction to the Framework was strategic and effective. It is a strength of the framework in confirming consistency, transparency, and cross-division coordination as founding principles and is excellent in preparing the reader to place in context, assimilate and understand subsequent sections. The presentation of modeling methods, model parameters, monitoring, exposure characterizations, and next steps, while parsimonious, are clear and fit for purpose.

A Panel member observed that it was important that the Framework addressed the timeline and regulatory time constraints of DWAs, including the number performed per year.

The Panel members agreed that the document should help EPA achieve the stated goals; to forward consistency and transparency in deriving estimated pesticide concentrations in surface water for use in DW assessments, and foster cross-division coordination.

One or more Panel members recommended the following topics for further explanation and/or discussion, to improve the Framework:

- The Agency should clarify the intended target audience of this Framework document, which is not defined and appears not to be fixed. Much communication appears to be directed toward Agency personnel; however, some passages address stakeholders, and others address members of the public. This Framework document should be revised with Agency readers as the sole target audience.
- The Agency should consider developing a separate Drinking Water Framework specifically for the non-Agency target audience. For audiences outside of the Agency, some communication in the current document relies on terms and concepts that require insider knowledge and training and might be termed "Agency jargon." This limits communication effectiveness for non-Agency readers. Two examples include, "ground truth the model refinement" and "risk cup."
- Use of the opening Figure 1.1 should be re-assessed. The strategy of beginning the Framework with this illustration that attempts to convey the multi-tier approach used by the Agency for DW assessments is logical and valuable. However, the design and content of Figure 1.1 are overwhelming, intimidating, and difficult to digest. Thought should be given to simplification of the figure and adding some of the extensive detail from Figure 1.1 to new figures.
- Provide a more clear and detailed presentation of the PWC modeling tool. It is a central method for lower tier assessments. Due to the brevity of the Framework, there was not an opportunity for it to be clearly presented. Because it is not expressly defined and explanatory exemplars presented, there is some concern that it may not be as systematic as desired, and it may not be reproducibly applied across Agency assessments.
- Modeling for tier 3 and 4 assessments use "typical" pesticide application rates, while monitoring, including SEAWAVE-QEX imputations, reflects the pesticide application rates that were used in the monitored watershed. This weakens the Agency's statement that the pesticide is safe "when used as directed," because the assessment considers less pesticide than is allowed. One Panel member recommended that the Framework address this issue and clarify whether the EPA makes any adjustments to account for the difference between what is used for the evaluation and what is allowed by the label. For example, during the input process, adjust the recommended pesticide label rates so that the maximum allowed rates match the typical use rates used for the evaluation.
- Add general guidance on when (in what tier or tiers), and how EPA addresses nonagricultural use for those pesticides that have significant nonagricultural uses. Case study 1 illustrates one approach for addressing the residential use of one such pesticide. Although nonagricultural usage data are generally not available (except in California), urban uses/sources still are (and need to be) considered in DW assessment, and options for doing so are not currently addressed in the Framework.
- The description of modeling includes using single values for environmental fate characteristics pesticides. Pesticide fate characteristics are usually dependent on the

environment and can exhibit large variations from site to site. In addition, because the evaluation of environmental fate studies is done conservatively, the modeling should generally overpredict environmental concentrations. Refining the evaluation of the environmental fate should be explored as a Tier 3 refinement for modeling, although refined fate estimates may require revising the laboratory study designs.

- The current method for estimating pesticide degradation (reference 3 in the Framework) is usually conservative, based on the slowest degradation observed during a study. This was done because most pesticide degradation studies show rates that slow down with time and the current study design does not allow isolating the causes of non-exponential behavior in a study.

- In current laboratory study protocols, the longest intervals between samples are at the end of the study, just the place where the degradation rate is measured, resulting in even more uncertain estimates of degradation.

- There is mention made of adjusting degradation rates for temperature, and there is a reference at the end of the document for the method used, but there is no link between the two. There is also no mention of adjusting hydrolysis for temperature.

- Overall, there are ways that the modeling done at Tiers 2 and 3 could be refined so that it may be less necessary to rely on monitoring data but doing so may require improvements to the standard environmental fate studies. These refinements would apply to the SAM as well as to the PWC.

- In Tier 2, clarify what "alternative model input parameters" EPA considers when multiple residues of concern have been identified (page 23 of the Framework).

- Explain links between durations, exposures, and effects (that is, the toxicological bases) for each of the 1, 4, 21, and 365-day and 30-year durations used in the Tier 2 surface water model (page 26).

- Once decisions are made by EPA on the scope of use for SBF, SEAWAVE-QEX model and weight-of-evidence approach, the Framework will need to be updated to describe the use of these tools, under the appropriate tier.

- The description of SEAWAVE-QEX model should address its limited use in non-flowing waters, which are important as DW in many agricultural watersheds; specify variance in the estimated or extrapolated maximum pesticide concentrations that are compared in deterministic fashion to DWLOCs; and state any assumptions both implicit and explicit.

- The DWLOC is defined as the Drinking Water Level of Comparison in the Framework document (page v). In other EPA documents, DWLOC is defined as the Drinking Water Level of Concern. This discrepancy should be explained and addressed as appropriate. Acknowledging space and conceptual constraints of the Framework document, additional explanation should be added to place DWLOC in the broader context of other parameters employed by the Agency in the space of risk assessment.

- Clarify the conceptual model to show pivotal decision criteria (exposure vs. effects data, etc.) and information required for reliable decisions. Presently, the Framework emphasizes the potential exposure information (i.e., monitoring data and modeling estimates), with relatively little emphasis on the accuracy of the risk comparator or the effects data (i.e. DWLOC, MCL, and HAs). The latter deserves more attention because ultimately, uncertainty in the risk decision hinges on the accuracy and reliability of the components.

- Clarify if, when, and where probabilistic assessments are conducted.

- Address how pesticide use estimates are refined as assessment tiers increase.

- Address where, when and how terrestrial field dissipation data are evaluated.
- Address at what point label restrictions are evaluated (for example, requirements for edge-of-field buffers).
- Add general guidance on when (in what tier or tiers) and how EPA addresses non-agricultural use for those pesticides that have significant non-agricultural uses. Case study 1 illustrates an approach for addressing the residential use of one such pesticide. Although non-agricultural usage data are generally not available (except in California), urban uses/sources still are (and need to be) considered in DW assessment, and options for doing so are not currently addressed in the Framework.
- The basis for applying pesticide-specific SBFs to other pesticides in Tier 3 does not appear to be well established, as EPA white paper findings were not supportive. Consider grouping pesticides on the basis of use characteristics.
- Clarify how the PWC and SBF computed values are balanced in Tier 3, and whether one takes precedence.
- Consider calibration, in particular hydrology, for use in Tier 4 models, such as with SAM.
- Consider use of other exposure models at Tier 3 or Tier 4. For example, APEX is widely used for assessing impact of conservation practices on field-scale pesticide loss. Likewise, SWAT is routinely used at watershed scales to evaluate pesticide losses. Both APEX and SWAT serve as the foundation of the ongoing USDA Conservation Effects Assessment Program (CEAP). Reports that describe the regional application say the HUC-02 level may be of interest (https://www.nrcs.usda.gov/wps/portal/nrcs/detail/national/technical/nra/ceap/pub/?cid=nrcs143_0). A potentially useful feature of the APEX simulations is the "base case" scenarios. These assume that no conservation practices were used, thus providing a worst-case for exposure assessment. This may be an appropriate level of conservatism in some Tiers of EPA pesticide risk assessments.
- In modeling at Tier 4, consider goodness of fit criteria and or other assessments of model performance. Harmel et al. (2012) provide useful guidance.
- There are 18 HUC-02 regions (page 28 of the Framework) but 21 different water resource regions (page 10). These different systems should be explained, and how they play differing roles in DW assessments.
- Address why there are no existing model scenarios available for California and several other regions (i.e., HUC-02 regions 14, 15, and 16; page 30 of the Framework document). It would be helpful to discuss why, and to state whether there are scenarios under development.
- Consider adding a column to Appendix B of the Framework document, to address refinements at each tier.
- Consider including the SOP for SEAWAVE-QEX as an attachment to the Framework.
- Consider attaching to the Framework the SOP documents for DWA that are mentioned on page two of the Framework.

The following minor errors and editing issues should be conveyed to EPA:
- Recurring instances in which human characteristics are imputed to inanimate objects or organizations should be removed. This use of anthropomorphism can be distracting and may even imply to some that the framework is not a fully refined document.
- Correct the Figure numbers; some are incorrect.

- Provide greater explanation of "percentile of vulnerability" in the DW Framework (page 22).
- The term, "the Assessor," should be reconsidered where used in the context of a specific issue or decision. It implies that a single assessor is responsible for issue resolution or final decision. Perhaps use of "the assessment" would avoid this unintended implication.

Panel Response 4b:

b.   Multiple Panel members noted that the case studies were useful for understanding how SEAWAVE-QEX and SBFs fit into the evaluation Framework, as well as the level of complexity of the assessment and the logic and balancing of factors needed for a pesticide-specific document. Case study 1 effectively illustrates how SEAWAVE-QEX and SBFs fit into an acute exposure assessment for drinking water.

One Panel member noted that in selecting chem1 for the DWA case study 1, the Agency appears to have made a carefully considered and strategic choice of pesticide. The same Panel member commented that chem1 is an uncommon data-rich pesticide that satisfies SEAWAVE-QEX model requirements for 31 extensively monitored drinking water sites. Application of the model returned no potential exceedances. With moderate persistence, chem1 shares a common fate profile with a number of other pesticides with moderate terrestrial dissipations rates, particularly near or at neutral pH. Tier 2 and Tier 3 DWAs returned maximum concentrations far above the DWLOC and provide ample justification for a detailed and systematic Tier 4 assessment of chem1. This selection may be able to serve as an effective surrogate for a range of other pesticides that have primarily residential use in areas of low cropland presence and a high percentage of impervious surface. The Panel member recommended that a new short-term case study with this level of sophistication be considered using a pesticide representative of those with long term persistence.

One Panel member commented that pesticide-specific SBF developed for chem1 provided added value to a broader approach of applying the SBFs for the four evaluated pesticides in the white paper (atrazine, carbaryl, chlorpyrifos, and fipronil). Although correlations were not found between chem1 SBFs and site or watershed characteristics, this is a logical approach that should undergo broader investigation, including expanding the number and diversity of pesticides for which SBFs are available, particularly persistent organic pollutants.

This case study illustrates the lack of usable data for SEAWAVE-QEX; out of 19,048 sites with chem1 monitoring data, only 32 sites meet the data requirements and model assumptions for SEAWAVE-QEX. Of those, 6 sites are in region three [one of which is a drain; see page 31 of the EPA white paper)] and six sites are in region 17. The Panel questioned how representative these six sites are of the region. The SBFs were large, showing high uncertainty, so it was helpful to see the Agency's calculations. The diagnostic plots were effective. This case study shows real world problems encountered and the difficulty in making DWAs. One Panel member commented that this case study shows that EPA did the best they could with the data they had and that the SEAWAVE-QEX model has value in these assessments.

One Panel member praised the work by EPA on Figure 1.1 (page 3 of case study 1) as a useful summary of the assessment procedure for chem1, to which one could refer back repeatedly while reading the assessment.

The Panel agreed that the case study 1 document clearly shows the processes used in moving from screening up through Tier 4, with some exceptions noted below. One or more Panel members made the following comments and recommendations for improving the case study 1 document:

- The Panel recommended including more detail about evaluating the SEAWAVE-QEX fits. For chem1, short-term SBFs were calculated for 31 sites, only two of which were discussed briefly. Since evaluation of the results is an important part of SEAWAVE-QEX modeling, and the most difficult part of SEAWAVE-QEX modeling, a little more discussion of some of the weaker fits would help illustrate the process for evaluating SEAWAVE-QEX results. For example, some of the diagnostic plots show what appear to be weak seasonal waves. There should be some discussion of how these weaker waves were evaluated, and where to draw the line between acceptable and unacceptable fits.
- One Panel member commented that the USGS site 14201300 (Zollner Creek, Oregon) is interesting because it was done twice, once for 1993-2003 and once for 2005-2015. The diagnostic plots showed very different shapes for the two sets of years, and the Panel member suggested that adding a discussion of how to interpret this would be useful. The second of these had lower concentrations for the monitoring data and the second diagnostic plot showed a very weak seasonal wave. Also, one of the highest detects was recorded exactly one month before the first date with flow data, and thus was not included in the evaluation.
- A Panel member commented on the following assumption by EPA in this case study: "Residential use patterns were not modeled, as the agricultural modeling is expected to be protective of the residential use patterns based on lower application rates and residential use patterns are expected to be more spread out over time and space (as compared to agricultural use patterns)" (page 7 of case study 1). Because chem1 had higher residential use (3.8 million lbs) than use in agricultural areas (750,000 lbs) (page 4), the Panel member recommended that EPA address whether there may be vulnerable areas with both agricultural and nonagricultural inputs within the watershed, and if so, whether concentrations from urban use would be negligible compared to maximum agricultural concentrations, or whether the results would be additive.
- There may be sites with too few samples to use an SBF, but with data that would exceed the DWLOC if it were used. One Panel member suggested that this case be examined. An example would be a site with only six samples, but one of them showing concentrations high enough that if an SBF were used, it would exceed the DWLOC.
- A Panel member commented that EPA's definition of the "Minimum SBF at DWLOC" (page 20, section 4.3.1 of case study 1) is confusing because the relation to SBF is not clear, and suggested this could be clarified by adding to the text the equation defining SBF as the ratio of the true maximum concentration divided by the maximum measured; then by assuming that the true maximum exceeds the DWLOC, this leads to the equation shown for "Minimum SBF at DWLOC."
- One Panel member commented that Figure 4.3 of case study 1 (page 20), which shows potential DWLOC exceedance for Minimum SBF at DWLOC in relation to SBF, is very difficult to follow. The Panel member noted that the format used for Figure 4.3 is confusing, but it is useful in that it allows comparison of several distributions simultaneously (the distribution for each of the four reference pesticides, plus the environmental data). Two Panel members further commented that the label for the box in

Figure 4.3, "Potential DWLOC Exceedance" is counterintuitive, because an exceedance is expected to be high, whereas here the potential exceedances occur at low SBF values, and both Panel members recommended changing the box label. One Panel member suggested changing the name of "Min SBF to DWLOC," to "DWLOC/Cmax safety margin," where Cmax is the maximum concentration, or a similar name. Because higher risk occurs at lower safety margins (or safety factors), such a name would help explain why exceedance occurs at low SBF values in Figure 4.3, case study 1. The same Panel member also recommended revising the box label in Figure 4.3, case study 1 to clarify what the boxes mean, such as "Potential DWLOC exceedance occurs when the SBF exceeds the DWLOC/Cmax safety margin."

- One Panel member commented that the follow-up explanation of SBF analysis (page 21, section 4.3.1 of case study 1) was well done, in which EPA handled regions with studies having fewer than 13 samples per year or regions that had insufficient monitoring data. This analysis resulted in bringing 3 additional regions into Tier 4.

- A Panel member recommended that the criteria for applying the SEAWAVE-QEX model at water quality portal (WQP) sites (page 29, section 5.2 of case study 1) should include the requirement of flow data (or a surrogate), which is a critical element of the model.

- On page 42 (section 5.3 of case study 1) it is stated that SBFs were developed by running Python code in Spyder. A Panel member suggested that a Jupyter notebook might be a better tool for this, as it would facilitate documentation of the process used in the calculations. This would make it easier to review the process in the future. Similarly, use of SEAWAVE-QEX could be done in either a Jupyter notebook or in an R notebook in RStudio, again with the goal of documenting the process.

- One Panel member commented that the description of the SBF analysis in Tier 4 (page 46, section 5.4 of case study 1) was not as clear as the SEAWAVE-QEX analysis (page 36, section 5.2.3). The same Panel member recommended that the SBF analysis start by specifying the criteria EPA is using, namely, exceedance is defined as having (1) detected concentration above the Method Reporting Level, (2) the SBF-adjusted concentration greater than the DWLOC, and (3) the site location must be relevant to drinking water intake(s). Then as the case study progresses, the logic will be easier to follow.

- For Figures 5.17 and 5.18 (page 52, section 5.4 of case study 1), the same Panel member recommended that EPA color-code those site(s) that met all criteria, that is, were predicted to exceed DWLOC and were relevant to DW intake.

- A Panel member recommended that EPA explicitly address residential areas in the monitoring recommendations in sections 5.6 and 6.0 (pages 56-57) of case study 1. The Agency previously concluded that, "Based on the weight-of-evidence, there are still some areas with potential drinking water concerns: citrus growing areas in Florida, apple growing areas in Washington and Oregon, and urban/residential areas" (page 53). However, monitoring recommendations in section 5.6 mention "soybean, apple, and citrus growing regions of HUC-02 region 03," but not residential areas (page 56). Similarly, section 6.0 concludes that "a more robust surface water monitoring program in Florida, Washington, and Oregon would be beneficial" (page 57), but does not address monitoring needs in residential areas, although Tier 4 found one urban site for which SEAWAVE-QEX estimates predicted a DWLOC exceedance.

Editorial comments:
- There were inconsistencies between the text of case study 1 (page 48) and Figures 5.15 and 5.16 (pages 48-49), as the description in the text of the color-coding of sites does not match either of the figures. Check the figure numbers for errors. The table on page 47 is labeled "Figure 5.14," which may have caused a figure numbering error.

Panel Response 4c:

c. The Panel expressed appreciation to EPA for providing the case study and noted that it effectively demonstrated the risk assessment process in particular how SEAWAVE-QEX and SBFs may be used at higher tiers in assessments. One Panel member went on to observe that SEAWAVE-QEX provides EPA a great new tool to use and that implementation in the near future is recommended.

Specific concerns regarding questions about the applicability of SEAWAVE-QEX to static CWS and whether surrogate covariates can be used effectively when flow data is not available were addressed in responses to other charge questions and were not repeated. Similarly charge question responses that expressed concerns regarding applicability of SBF computed for the four compounds discussed in the white paper to chem2 and other pesticides were referenced and not repeated. The primary concern in this case was SBF use at Tier 3 of the risk assessment. Beyond this Panel member's comments, the following observation were made:

- As noted in Panel's comments for charge question 4b, there should be additional discussion regarding evaluation of SEAWAVE-QEX fits to data.
- In section 5.1 of case study, it is not clear how the Estimated Drinking Water Concentrations (EDWCs) were calculated. In the document is stated that "the maximum PCA within each HUC-02 region was multiplied by every relevant scenario EDWC." It is not clear what a "relevant scenario EDWC" is and how many there were.
- With regard to the potential use of stage or precipitation as surrogates for flow in SEAWAVE-QEX, it would have been informative to compare SEAWAVE-QEX performance using all three parameters, flow, stage, and precipitation, when they were available.
- Some additional discussion of how the ten sites with unacceptable SEAWAVE-QEX fits were rejected would improve understanding of the evaluation process.
- Minor note: in table 4.5, stable is denoted as zero half-life; for the table it would be better to denote it as "stable."
- Figure 1.1 in the case study was helpful however a more detailed graphical representation of the process that showed decision points would have been useful.
- In section 2.1, of case study 2, use characteristics of chem2 are not clear. What fractions were applied pre- and post-emergence? In addition, does the current label permit more than one application per season? If so, how often does this occur?
- In fate summary table 2.1, specify that the half-life units are days, provide a footnote indicating that the "sources" are document numbers, and explain the term $t_{R\ IORE}$.
- Water treatment data were included in the fate summary (section 2.2) but apparently were not used in the risk assessment. Is this correct? If so, this should be stated in the text.

- Monitoring data summarized in table 3.4 characterized results in several ways. Dissolved, total, recoverable. Explain how data were handled for use in SEAWAVE-QEX and SBF analyses.
- In Tier 3 (section 4.4) it appears that PWC and SBF assessments were in agreement. Was this the case? A tabular compilation showing each site and PWC and SBF values would be useful. Was more weight applied to one result over the other?
- In section 5.2, SEAWAVE-QEX appears to have been appropriately applied but some points of clarification are needed regarding the 67 sites that met the sampling criteria. This includes summarizing site characteristics in tabular format showing which had flow data, which did not, the total number of datasets, the number of datasets that met SEAWAVE-QEX criteria, etc. This would reinforce the fact that a limited amount of the monitoring data met SEAWAVE-QEX criteria.
- In addition, among 67 sites which met sampling criteria, only 37 had flow data and after assessing with SEAWAVE-QEX, only 27 sites had acceptable criteria. This represented only 0.51% of the original sites. Comments on how representative this group of sites were to the total would be informative.
- It was noted that in Region 3, "none of the sites were in the citrus region in FL" and in region 18, only 1 site was collocated in a potential use area. Perhaps this was a situation where precipitation could have been used to increase the potential for SEAWAVE-QEX application where high chem2 use was reported. Note this is not a recommendation to go back and do this since it would be very time intensive.
- Of the 27 sites where SEAWAVE-QEX was used, one showed a DWLOC associated with a retention pond. EPA recognized this and removed these results from further analysis. The Panel supported this approach since the site was likely not representative of a drinking water source.
- "Weight-of-evidence factors" were applied with good judgement.
- In spite of questions about of representativeness of the few samples, results of SEAWAVE-QEX appeared reasonable.
- Given that the Agency has 100-150 assessments to do each year, the level of effort and use of SEAWAVE-QEX in the case study appeared appropriate.

LITERATURE CITED

American Society of Agricultural and Biological Engineers (ASABE). Guidelines for Calibrating, Validating, and Evaluating Hydrologic and Water Quality (H/WQ) Models. St. Joseph, Michigan, ASABE. EP621, 15p, DOI : 10.13031/trans.12806.

Baker NT, Stone WW, 2015. Estimated annual agricultural pesticide use for counties of the conterminous United States, 2008–12: U.S. Geological Survey Data Series 907, 9 p., https://dx.doi.org/10.3133/ds907.

Banerjee, Sudipto, Carlin, Bradley P, Gelfand, AE. 2014. Hierarchical Modeling and Analysis for Spatial Data. CRC Press, Boca Raton. ISBN 9781439819173.

Bernillon P, Bois FY. 2000. Statistical issues in toxicokinetic modeling: A Bayesian perspective. Environ. Health Perspective 108 (supple. 5):883-893.

Bosch DD, Sheridan JM, Davis FM. 1999. Rainfall characteristics and spatial correlation for the Georgia Coastal Plain. Transaction of the American. Society of Agriculture Engineers: 42:1637–1644.

Bosch DD, Arnold JG, Allen PG, Lim K-A, Park YS. 2017. Temporal variations in baseflow for the Little River experimental watershed in South Georgia, USA. *Journal of Hydrology: Regional Studies* 10, 110–121.

Box, GEP, and Draper, NR. 1987. Empirical model-building and response surfaces. John Wiley & Sons, New York, NY.

Breiman L, Friedman J, Stone CJ, Olshen RA, 1984. Classification and Regression Trees. Chapman & Hall/CRC Press, Boca Raton. ISBN 9780412048418.

Carsel RF, Imhoff JC, Hummel PR, Cheplick JM, Donigian AS Jr, 1998. PRZM-3, A Model for Predicting Pesticide and Nitrogen Fate in the Crop Root and Unsaturated Soil Zones: User's Manual for Release 3.0; National Exposure Research Laboratory, U.S. Environmental Protection Agency: Washington, DC.

Cressie, NAC. 1993. Statistics for Spatial Data. 2nd edition. Wiley, New York. Print ISBN:9780471002550.

David O, Lloyd W, Carlson J, Yage L, Ascough JC, Green TR,, Kipka H, Rojas K, Geter F2014. The Land Management and Operations Database (LMOD). In: Ames, D.P. et al. (Eds.), Proc. 7th International Congress on Environmental Modelling and Software (iEMSs), San Diego, CA, June 15-19, 2014, https://scholarsarchive.byu.edu/iemssconference/2014/Stream-A/30.

EPA. (2010) Toxicological Review of Dichloromethane (Methylene Chloride). EPA/635/R-10/003A.

EPA. 2019. (EPA White Paper) FIFRA Scientific Advisory Panel Meeting. The Approaches for Quantitative Use of Surface Water Monitoring Data in Pesticide Drinking Water Assessments (Chapter 5). FIFRA Scientific Advisory Panel, Office of Science Coordination and Policy, Office of Prevention, Pesticides and Toxic Substances. U.S. Environmental Protection Agency. Washington, DC. Document available in the public docket: EPA-HQ-OPP-2019-0417.

EPA. 2019. (Framework) Framework for Conducting Pesticide Drinking Water Assessments for Surface Water. FIFRA Scientific Advisory Panel, Office of Science Coordination and Policy, Office of Prevention, Pesticides and Toxic Substances. U.S. Environmental Protection Agency. Washington, DC. Document available in the public docket: EPA-HQ-OPP-2019-0417.

EPA. 2019. (Case Study 1) Drinking Water Assessment Case Study Example for Acute Exposure Concern. FIFRA Scientific Advisory Panel, Office of Science Coordination and Policy, Office of Prevention, Pesticides and Toxic Substances. U.S. Environmental Protection Agency. Washington, DC. Document available in the public docket: EPA-HQ-OPP-2019-0417.

EPA. 2019 (Case Study 2) Drinking Water Assessment Case Study Example for Chronic and Cancer Exposure Concern. FIFRA Scientific Advisory Panel, Office of Science Coordination and Policy, Office of Prevention, Pesticides and Toxic Substances. U.S. Environmental Protection Agency. Washington, DC. Document available in the public docket: EPA-HQ-OPP-2019-0417.

EPA. 2019. (SOP) Standard Operating Procedure for Using SEWAVE-QEX to Estimate Pesticide Concentrations. FIFRA Scientific Advisory Panel, Office of Science Coordination and Policy, Office of Prevention, Pesticides and Toxic Substances. U.S. Environmental Protection Agency. Washington, DC. Document available in the public docket: EPA-HQ-OPP-2019-0417.

FIFRA SAP, 2015. Development of a Spatial Aquatic Model (SAM) for Pesticide Assessments". FIFRA Scientific Advisory Panel. Sept 15-17, 2015. Document available in the public docket: EPA-HQ-OPP-2015-0424.

García AM, Alexander RB, Arnold JG, Norfleet L, White MJ, Robertson DM, Schwarz G. 2016. Regional effects of agricultural conservation practices on nutrient transport in the Upper Mississippi River Basin. Environmental Science and Technology. 50:6991−7000.

Green TR, Kipka H, David O, McMaster GS. 2017. Where is the USA Corn Belt, and how is it changing? Science of the Total Environment. 618: 1613-1618. doi:10.1016/j.scitotenv.2017.09.325.

Gustafson DI, Carr KH, Green TR, Gustin C, Jones RL, Richards RP. 2004. Fractal-based scaling and scale-invariant dispersion of peak concentrations of crop protection chemicals in rivers, Environmental Sci. & Tech. 38(11):2995-3003.

Gyau-Boakye P, and Schultz GA. 1994. Filling gaps in runoff time series in West Africa. Hydrological Sciences Journal. 39(6):621-636.

Harmel RD, Smith PK, Migliaccio KW, Chaubey I, Douglas-Mankin KR, Benham B, Shukla S, Muñoz-Carpena R, Robson BJ. 2014. Evaluating, interpreting, and communicating performance of hydrologic/water quality models considering intended use: A review and recommendations. Environmental Modelling & Software, 57: 40-51.

Hendley P, Chen W, Joseph R, Williams R, Harbourt C, Miller P. 2019. Improved Characterization of the Temporal and Spatial Variability of Potential Surface Water Drinking Water Exposure by Using Environmental and Historic Monitoring Databases. Presentation to the USEPA Environmental Modeling Public Meeting. 2012. Available on-line at https://www.regulations.gov/document?D=EPA-HQ-OPP-2009-0879-0032. [Last accessed December 6, 2019].

Johnson HM, Domagalski JL, Saleh DK. 2011. Trends in pesticide concentrations in streams of the Western United States, 1993-2005. Journal of the American Water Resources Association, 47:265-286.

Jonsson O, Paulsson E, Kreuger J. 2019. TIMFIE Sampler−A New Time-Integrating, Active, Low-Tech Sampling Device for Quantitative Monitoring of Pesticides in Whole Water. Environmental Science and Technology, 53:279−286.

Kalkhoff SJ, Vecchia AV, Capel PD, Meyer MT. 2012. Eleven-Year Trend in Acetanilide Pesticide Degradates in the Iowa River, Iowa. Journal of Environmental Quality, 41:1566-79. doi:10.2134/jeq2011.0426.

Kipka H, Green TR, David O, Garcia LA, Ascough JC, Arabi M. 2016. Development of the Land-use and Agricultural Management Practice web-Service (LAMPS) for generating crop rotations in space and time. Soil and Tillage Research 155: 233-249.

Leopold LB, Wolman MG, and Moller JP. 1964. Fluvial Processes in Geomorphology. W.H. Freeman and Co., San Francisco. 522 pp.

Marino DJ, Clewell HJ, Gentry PR, Covington TR, Hack CE, David RM, Morgott DA. 2006 Revised assessment of cancer risk to dichloromethane: Part 1 Bayesian PBPK and dose-response modeling in mice. Regulatory Toxicology and Pharmacology, 45:44-54.

Mosquin PL, Whitmore RW, Chen W. 2012. Estimation of upper centile concentrations using historical atrazine monitoring data from community water systems. Journal of Environmental Quality, 41(3), 834–844. https://doi.org/10.2134/jeq2011.0209.

Mosquin PL, Aldworth J, Chen W. 2018. Flow-covariate prediction of stream pesticide concentrations. Environmental Toxicology and Chemistry, 37(1), 260–273. https://doi.org/10.1002/etc.3946.

Nakagaki N, Wolock DM. 2005. Estimation of agricultural pesticide use in drainage basins using land cover maps and county pesticide data: U.S. Geological Survey Open-File Report 2005-1188, 46 p.

NOAA. 2019. National Center for Centers for Environmental Information National Oceanic and Atmospheric Administration NEXRAD Products. Available online at https://www.ncdc.noaa.gov/data-access/radar-data/nexrad-products. [Last accessed December 6, 2019].

Potter TL, Bosch DD, Strickland TC. 2014. Comparative assessment of herbicide and fungicide runoff risk: a case study for peanut production in the Southern Atlantic Coastal Plain (USA). Science of the Total Environment: 490, 1-10.

Potter TL, Truman CC, Strickland TC, Bosch DD, Webster T. 2008. Herbicide incorporation by irrigation and tillage impact on runoff loss. *Journal of Environmental Quality,* 37:839-847.

Stone WW, Crawford CG, Gilliom RJ. 2013. Watershed Regressions for Pesticides (WARP) Models for Predicting Stream Concentrations of Multiple Pesticides. Journal of Environmental Quality. 42:1838–1851. doi:10.2134/jeq2013.05.0179.

Tencallec P. Favre A-C, Prieur C, Mathevet T.2015. Reconstruction of missing daily streamflow data using dynamic regression models. Water Resources Research. 51(12):9447-9463.

Thelin GP, Stone WW. 2013. Estimation of annual agricultural pesticide use for counties of the conterminous United States, 1992–2009: U.S. Geological Survey Scientific Investigations Report 2013-5009, 54 p.

Thode HC. 2002. Testing for normality, New York: Marcel Dekker, ISBN 0-8247-9613-6

Truman CC, Potter TL, Nuti RC, Franklin DH, Bosch DD. 2011. Antecedent water content effects on runoff and sediment yields from two Coastal Plain Ultisols. Agricultural Water Management. 98:1189-1196.

USGS. 2019. United States Geological Survey. Provisional Data for the USA: provisional data statement. Available on-line at https://waterdata.usgs.gov/nwis/?provisional. [Last accessed December 7, 2019].

Vecchia AV. 2018. Model methodology for estimating pesticide concentration extremes based on sparse monitoring data: U.S. Geological Survey Scientific Investigations Report 2017–5159, 47 p. https://doi.org/10.3133/sir20175159.

APPENDIX 1: Dr. V. Berrocal's Description of the SEAWAVE-Q and SEAWAVE-QEX Model from Vecchia, 2018.

From the EPA white paper (page 33), "The seasonal wave is a periodic solution to a conceptual storage equation, with a period of one year and up to two seasons per period." The variables driving this model are the number and length(s) of pesticide application season(s), season start time (or phase shift), pesticide input rate and removal (dissipation) rate. The model also considers two variables calculated from daily flow records, the mid-term flow anomaly (MTFA) and the short-term flow anomaly (STFA). The MTFA helps the model account for long-term seasonal variability in flow whereas the STFA helps account for high-frequency variability in flow, such as might occur from a rainfall event. Finally, when fitting the model to available monitoring data, consideration is taken of the span of time the data represents and the potential for high correlations of concentrations from samples taken very close together in time (the built-in data thinning function for 3-day spacing). When the model is fit to available monitoring data, an output file is produced that contains all estimated model parameters.

The SEAWAVE-Q model is expressed as:

$$\text{Log}\{C(t)\} = \beta_0 + \beta_1 W(t) + \beta_2 A_{MT}(t) + \beta_3 A_{ST}(t) + \beta_4(t-t_m) + \varepsilon(t)$$

Where,
Log is the base 10 logarithm,
$C(t)$ is the pesticide concentration reported in micrograms per liter,
β_0, $\beta 1$, β_2, β_3 and β_4 are regression coefficients,
$W(t)$ is the seasonal wave described below,
$A_{MT}(t)$ is the MTFA,
$A_{ST}(t)$ is the STFA,
t is the time reported in decimal years,
t_m is the midpoint of the time interval being analyzed, and
$\varepsilon(t)$ is the model error.

The seasonal wave $W(t)$ models seasonality in log-transformed pesticide concentration is a member of a class of functions indexed by m, the pulse input model number (m=1,2,…,6), h, the decay rate measured in decimal months (h=1,2,3 or 4) and s, the phase shift also measured in decimal months (s=0,.5,1,…,11.5).

The seasonal wave generally is the most important term in the SEAWAVE-Q model because the seasonal wave usually explains the most variability in concentrations compared to other variables in the model.

The two flow-anomaly terms in equation 1 ($A_{MT}(t)$ and $A_{ST}(t)$) are dimensionless variables calculated from a daily streamflow record assumed to be available from a streamflow-gaging station at or near the site being modeled. The mid-term flow anomaly is computed using log-transformed daily streamflow for 30 days up to and including the current time. The short-term flow anomaly (STFA) is computed by subtracting the mean

and the MTFA from log-transformed flow. In a sense, observed streamflow is decomposed as the period of record mean plus the MFTA plus the STFA.

For analyzing trends with sparse monitoring data, the error term in equation 1 generally can be assumed to consist of independent normal random variables with constant variance.

The SEAWAVE–QEX model consists of the SEAWAVE–Q model (eq. 1) modified to include seasonal variance and serial correlation in the model errors.

$$\text{Log}\{C(t)\} = \beta_0 + \beta_1 W(t) + \beta_2 A_{MT}(t) + \beta_3 A_{ST}(t) + \beta_4(t-t_m) + SSD(t)Z(t)$$

$SSD(t)$ is a function of the seasonal standard deviation $s>0$ and a regression on the seasonal wave. SSD is modeled as

$$SSD(t) = \sigma[1+\alpha W(t)]^{1/2}$$

$Z(t)$ is a normalized and temporally autocorrelated model error assumed to have a normal distribution with mean zero, variance one and serial correlation function

$$Corr(k) = e^{-|k|/CTS}$$

Where e is Euler's constant, k is the time lag, in days, between observations and CTS is the correlation time scale measured in days.

Fitting the model is a 5-step process involving 17 tasks that are implemented as 15 functions in R.

Step 1 Data preparation and screening.
1. Download daily discharge data and compute flow anomalies
2. Merge concentration data and flow anomalies
3. Determine appropriate period of record
4. Determine if data meet minimum screening requirements

Step 2 Regression analysis and selection of the best-fit seasonal wave.
5. Ignore serial correlation and assume constant seasonal standard deviation ($\alpha=0$).
6. Estimate regression coefficients for seasonal wave, flow anomalies, and trend, selecting the best fit seasonal wave from the 1152 possible seasonal wave forms.
7. Compute fitted values from regression equation.
8. Compute regression residuals (includes censored values).

Step 3 Estimation of seasonal standard deviation and serial correlation parameters.
9. Using regression residuals, compute maximum pseudo-likelihood estimates of seasonal standard deviation parameters (σ and α) and correlation time scale.

10. Divide regression residuals by estimated seasonal standard deviation to obtain normalized residuals (includes censored values).

Step 4 Generation of conditional traces of daily concentration.
11. Generate values for censored normalized residuals
12. Generate normalized residuals for days with no observations
13. Multiply normalized residuals by seasonal standard deviation and add fitted values
14. Reverse log transformation
15. Repeat to obtain specified number of equally likely conditional traces

Step 5 Estimation of pesticide concentration extremes using generated conditional traces.
16. Compute simulated values of concentration extremes (for example, annual maximum daily concentrations or bias factors) from the simulated traces of daily concentration
17. Use simulated values to obtain point estimates (mean of simulated values) or confidence intervals (quantiles of simulated values)

Fitting the SEAWAVE-QEX model is a 5-step process involving 17 tasks that are implemented as 15 functions in R (Vecchia, 2018, USDOI, USGS Scientific Investigations Report 2017-5159, Model Methodology for Estimating Pesticide Concentration Extremes Based on Sparse Monitoring Data).

While the white paper does not go into detail, the full SEAWAVE-QEX tool includes the 8 diagnostic plots that together allow users to get a very good understanding of the model performance and the quality of the "fit" to a particular set of data. These include:

1. **Simulation Summary Output Graph**: A time series plot of the first conditional trace overlayered with points indicating the synthetic monitoring data, the highest censoring threshold, the estimated annual maximum with an 80% error bound for each year.
2. **Seasonal Wave Summary Output Graph**: A plot of the fitted seasonal wave (timeseries) curve overlayered with points indicating the synthetic monitoring data and curves constructed above and below the seasonal wave curve at 2 times the seasonal standard deviation (SSD).
3. **Adjusted Concentration Versus Mid-Term Flow Anomaly**: A scatter plot of MTFA (x-axis) versus Adjusted concentration (μg/L) overlayered with a fitted linear regression line. Provides insight into seasonal variability in flow. A negative regression coefficient indicates that higher flow leads to lower concentrations across the whole dataset.
4. **Adjusted Concentration Versus Short-Term Flow Anomaly**: A scatter plot of STFA (x-axis) versus Adjusted concentration (μg/L) overlayered with a fitted linear regression line. Provides insight into flashiness and contribution from runoff events. In general, the regression coefficient should be non-negative, since when STFA is elevated (i.e. large rain event) pesticide concentrations should increase, that is, more washoff of pesticide without dilution.

5. **Trend in Measured Concentration over the Entire Simulation**: This is a scatterplot of adjusted concentrations (concentration in μg/L less seasonality and flow-related variability i.e. MTFA and STFA) versus time across all years overlayered with a linear trend line. Used to identify years with unusually high or low concentrations not accounted for in the model.
6. **Normalized Residuals Viewed by Month**: Scatterplot of normalized residuals divided by seasonal standard deviation (y-axis ±3) by month (Jan-Dec) with all years overlayered on one plot.
7. **Normalized Residuals Viewed by Year**: Scatterplot of normalized residuals divided by seasonal standard deviation (y-axis ±3) by year with digital time within year.
8. **Correlation Function of the Normalized Residuals (Correlogram)**: Depicts the fitted exponential correlation function curve of normalized residuals (y-axis) versus "time between observations in days" (x-axis) overlayered with empirical autocorrelations with 95% confidence intervals. A vertical line indicates the estimated CTS which estimated the average time between successive observations. For an acceptable fit, the 95% CI for empirical autocorrelations with time steps lower than the CTS should overlap the fitted correlation function curve.