



June 10, 2016

Mr. Michael Goodis
Acting Director, Pesticide Re-Evaluation Division
Office of Pesticides Programs
United States Environmental Protection Agency
1200 Pennsylvania Avenue NW
Washington, DC 20460-0001

Submitted via Regulations.gov

Re: EPA OPP Draft Biological Evaluations of Chlorpyrifos, Diazinon, and Malathion, Docket identification numbers EPA-HQ-OPP-2008-0850, EPA-HQ-OPP-2008-0351, EPA-HQ-OPP-2009-0317 and EPA-HQ-OPP-2016-0167, 81 Fed. Reg. 21341 (April 11, 2016).

Dear Mr. Goodis:

CropLife America (CLA) appreciates the opportunity to comment on three draft Biological Evaluations (draft BEs) referenced above and to address the “Interim Approaches for National-Level Pesticide Endangered Species Act Assessments” (Interim Approaches), released by the Environmental Protection Agency (EPA or the Agency) and the U.S. Fish & Wildlife Service and the National Marine Fisheries Service (the latter two collectively, the Services) on November 13, 2013. Established in 1933, CLA represents the developers, manufacturers, formulators and distributors of plant science solutions for agriculture and pest management in the United States. CLA’s member companies produce, sell and distribute virtually all of the vital and necessary crop protection and biotechnology products used by farmers, ranchers and landowners. CLA is committed to working with EPA to encourage practical, science-based regulation of its members’ products.

While CLA recognizes the time and effort that went into the development of these draft BEs and the Interim Approaches, the draft BEs reflect seriously erroneous conclusions. These errors arise from flaws in EPA’s current methods used to conduct an Endangered Species Act (ESA) assessment and demonstrate that the Interim Approaches are unsustainable and do not meet the goals that the National Academy of Sciences (NAS) proposed for a streamlined Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA)/ESA consultation process. The Interim Approaches, as reflected in the draft BEs and related documents, is EPA’s and the Services’ first multi-agency attempt at an evaluation process responsive to the 2013 NAS Report.¹ While we recognize that the Agency attempted to post some documents in advance of the comment period opening, we note that these draft BEs provide the first opportunity for stakeholders to have access to the complete files and formally comment on the Interim Approaches. Given the gaps in the record² and EPA’s refusal to extend the comment period, these comments only highlight the most striking scientific and procedural

¹ Nat’l Academy of Sciences, *Assessing Risks to Endangered and Threatened Species from Pesticides* (2013).

² Docket EPA-HQ-OPP-2016-0167, 81 Fed. Reg. 21341 (April 11, 2016) was published for public comment containing many instances of erroneous, incomplete and incorrect data and information.

shortcomings of the draft BEs, Interim Approaches and their related documents.³ Nevertheless, CLA's necessarily limited review reveals that the draft BEs have not established appropriate data quality standards or meaningful tiered evaluations. In describing assumed chemical/species interactions, the draft BEs are filled with serious data errors, omissions and technical problems brought about by approaches that did not successfully address the NAS Panel's successive, tiered risk assessment goals.

THE NAS PANEL'S INTENT OF A TIERED PROCESS FOR RISK ASSESSMENT IS NOT ACHIEVED BY THE INTERIM APPROACHES

One common purpose of both FIFRA and the ESA is to assess foreseeable adverse impacts of regulatory actions. As such, both statutes are grounded in measuring ecological risk, including species and habitat. The Interim Approaches, as demonstrated here, have struggled with but failed to meet basic scientific standards for a tiered risk assessment. Due to unrealistic scientific assumptions, flawed models and missing or inappropriate applications of data and analyses, crucial steps in the assessment process that were envisioned by the Interim Approaches to screen out species do not do so.

In envisioning and now initially implementing the Interim Approaches, EPA and the Services seem to have missed a key point made by the NAS panel:

The committee views coordination among EPA and the Services as a collegial exchange of technical and scientific information for the purpose of producing a more complete and representative assessment of risk, including the types and depths of analyses to be conducted at each step in the process. Such coordination would allow EPA's expertise in pesticides to be effectively combined with the Services' expertise in life histories of listed species and in abiotic and biotic stressors of the species. . . . *The agencies can use Steps 1-3 as a framework for such discussions but need not be constrained by them...*⁴

Contrary to this advice, the draft BEs use an unnecessarily rigid, overly expanded and complex approach to the reevaluation of the effects of long-registered pesticide products on threatened and endangered species and their critical habitat. Although the Interim Approaches have adopted elements of the procedures and models from the Agency's FIFRA screening-level risk assessment process, EPA's assembly and application of those pieces is not transparent. The manner in which EPA has interpreted and implemented processes and models under the Interim Approaches disregards accepted and proven risk assessment processes and has lessened the certainty and reliability of results. In many cases, the Interim Approaches add unnecessarily conservative assumptions to already unrealistically conservative screening-level risk assessments, making them even less realistic, and thereby less relevant to meaningful risk assessment and screening. Additionally, the draft BEs are based on modeling that does not simulate pesticide exposures that are likely to occur in real-world environments, and use pathways or exposure scenarios that are

³ More detailed and technical CLA comments are presented as an attachment to this letter and respond to EPA's April 5, 2016 instructions for commenting on the draft BEs (Instruction Letter). See, e.g., EPA-HQ-OPP-2008-0351-0032, EPA Office of Chemical Safety and Pollution Prevention, "Instructions for Commenting on the Draft Biological Evaluations for Chlorpyrifos, Diazinon, and Malathion," (April 5, 2016).

⁴ NAS Report at 35 (emphasis added).

either speculative or physically impossible, or both. As a result of these unrealistic assumptions,⁵ the Interim Approaches have failed to provide evidence supporting the NAS Panel's goal of screening chemical and species interactions to successively eliminate risk presumptions and avoid wasting resources on multiple rounds of scrutiny of "what if" scenarios that are unlikely to exist.

Determining the impact of well-established farming practices on endangered species and their critical habitats, in accordance with the mandates EPA and the Services have been given by Congress, requires a much more reasonable focus that should include at least:

- Substantial synthesis of existing data on land use, land cover and species attributes, which is not currently undertaken in the Interim Approaches;
- Identification, qualification and use of data that are relevant, robust, reliable, and qualified for their fitness in risk assessment;
- Improved modeling and model parameters that are realistic, validated and representative of probable field conditions and interactions;
- Reasonable and representative scenarios; and
- To the extent the existing Services' "no effect/may affect" approach is retained, findings based upon factors more refined than simply co-occurrence.

THE BEs LARGELY IGNORE ESTABLISHED PRINCIPLES OF RISK ASSESSMENT

As the Agency is well aware, its regulatory actions must comply with its statutory obligations, which these draft BEs do not. The ESA requires that ESA assessments be based on the "best available scientific and commercial data," and neither FIFRA nor the ESA authorize EPA to use an overly precautionary approach.⁶ The draft BEs clearly demonstrate that EPA not only did not use the best available science, but also improperly applied unrealistic and overly precautionary assumptions and principles. Indeed, each of the draft BE terrestrial and aquatic determinations and accompanying appendices, attachments and modeling tools suffers from a number of serious shortcomings that result in unreasonable overestimates of exposure and risk to endangered species from the OPs at issue. The shortcomings include:

- Use of unrealistic assumptions;
- Incorrect or irrelevant endpoints;
- Modeling design and errors;
- Failing to take advantage of best available biological information that would produce more realistic exposure estimates that are relevant to the species being assessed;
- Numerous examples of insufficient information; and
- Lack of support to enable reviewers to understand how model input values were selected.

⁵ For example, the draft BEs assume the organophosphates at issue affect listed species if there is any overlap between pesticide use and a species' geographic range. The draft BEs ignore available temporal, agronomic, and geographical circumstances that eliminate the possibility of exposure.

⁶ In addition to the ESA's mandate that EPA make decisions based on "available" information – not speculation – it also mandates avoidance of "needless economic dislocation produced by agency officials zealously but unintelligently pursuing their environmental objectives." *Bennett v. Spear*, 520 U.S. 154, 176-177 (1997.)

In short, the draft BEs are deficient in data quality and use of accepted risk assessment practices, and demonstrate serious technical flaws and unrealistic assumptions and judgments, resulting in an overly precautionary approach to risk assessment. Instead of moving forward with what is an unworkable process, the Agencies should agree that the draft BEs should be used as a learning exercise and, further, that future time and resources be spent on more effective and productive means of pesticide regulation coupled with species protection.

THE INTERIM APPROACHES WILL NOT ALLOW THE AGENCIES TO MEET STATUTORY TIMELINES

Both the ESA and FIFRA include specific mandates for timely action. FIFRA requires EPA to review every registered pesticide every fifteen years. Currently, EPA intends to complete ESA assessments and consultations as part of registration review. At the current rate that EPA and the Services are working, as evidenced by the BEs, there is simply no way that EPA can complete registration review within statutory timelines. With regard to the statutory timelines set out in the ESA, these draft BEs make it abundantly clear that EPA and the Services cannot conclude any formal consultation within 150 days from initiation of formal consultation, as is required by the ESA.⁷

EPA and the Services should consider procedural steps towards simplifying and streamlining the consultation process, including immediately implementing the Department of Interior's Counterpart Regulations for optional formal consultation procedures, as well as exploring the potential to address groupings or categories of species or habitats.⁸ Consultation procedures that streamline the process will be far more protective of listed species and habitat and will fully comply with the Agencies' statutory FIFRA and ESA obligations. The Agencies should also explore opportunities for creative approaches to mitigation, including through conservation actions.

* * *

In sum, the draft BEs demonstrate that the Interim Approaches have failed as a predictive screening tool. Their use has led to "Likely to Adversely Affect" decisions for assumed interactions of species with chemicals that have been used safely across the country for 40 plus years. For example, the draft BEs show that the OPs are likely adversely affecting the Kirtland Warbler, a species well-documented as on the road to recovery. The Interim Approaches have therefore failed to provide a meaningful path towards a tiered risk assessment that will lead to an efficient, effective means to fulfill EPA's FIFRA and ESA obligations. Instead, they have created a process that does not screen and does not meaningfully assess risk, but instead threatens to remove from use valuable tools needed for production agriculture and public health, all while diverting resources away from more meaningful efforts towards species protection. CLA and its member companies continue to advocate for an ESA review process that works towards protecting species from potential adverse effects of agricultural operations.

⁷ 16 U.S.C. § 1536(b)(1)(B).

⁸ See 50 CFR §§ 402.46, 402.47.

CLA promotes reasonably characterizing use and exposure in a way that encourages more efficient and sustainable use of government, registrant and grower resources to continue to support species protection and, where possible, species recovery.

Sincerely,

A handwritten signature in blue ink that reads "Imad Saab". The signature is written in a cursive style with a period at the end.

Imad Saab, Ph.D.

Senior Director, Environmental Policy

Attachment: COMMENTS BY CROPLIFE AMERICA ON EPA OPP DRAFT BIOLOGICAL EVALUATIONS OF CHLORPYRIFOS, DIAZINON, AND MALATHION



**COMMENTS BY
CROPLIFE AMERICA**

ON

**EPA OPP DRAFT BIOLOGICAL EVALUATIONS OF
CHLORPYRIFOS, DIAZINON, AND MALATHION**

**DOCKET IDENTIFICATION NUMBER EPA-HQ-OPP-2016-0167, 81 FED.
REG. 21341 (APRIL 11, 2016)**

June 10, 2016

CropLife America
1156 15th Street NW
Washington, DC 20005
www.CropLifeAmerica.org

128 Pages

Representing the Crop Protection Industry



TABLE OF CONTENTS

Contents

1.0 INTRODUCTION.....	6
1.1 Participants in this Commenting Effort	6
1.2 Summary of Key Points and Recommendations	7
2.0 ENVIRONMENTAL MODELS USED IN THE DRAFT BEs – TERRESTRIAL EXPOSURE.....	12
2.1 Overview	12
2.2 Missing Information.....	13
2.2.1 Information Contained in the Biological Evaluation Dockets	13
2.2.2 Use of Best Available Data.....	14
2.3 Corrections to Errors in EFED Models	15
2.4 EFED was Unreasonably Conservative in Selecting Exposure Scenarios and Input Values.....	16
2.5 Necessary Terrestrial Modeling Improvements.....	19
3.0 COMMENTS ADDRESSING AQUATIC MODELS.....	20
3.1 Model Inputs.....	20
3.1.1 Pesticide Application Methods:.....	20
3.1.2 Environmental Fate and Transport:	21
3.1.3 Drift Estimates:.....	21
3.2 Aquatic Models.....	21
3.2.1 Scenario Development for the Agricultural Environment.....	25
3.2.2 Aquatic Exposure Modeling Tools	27
3.2.3 Parameterization of the Various Aquatic Bins.....	28
3.3 Spray Drift Parameterization	30
3.3.1 Spray Drift Exposure Modeling.....	30
3.3.2 Appropriateness of Model Selection.....	31
3.3.3 Current State of the Art for Ground Modeling - RegDisp.....	32
3.4 Spray Drift and Mosquitocide Use Assessment	32
4.0 COMMENTS ON THE APPROACH USED FOR STEP 1 [MAKING MAY AFFECT [MA] OR NO EFFECT [NE] DETERMINATIONS]	37



4.1 General Comments.....	37
4.2 Management of Data	39
4.3 Effects Determinations in Step 1 (Aquatic Assessment).....	40
4.4 Effects Characterization for Terrestrial Plants	41
5.0 COMMENTS ON THE APPROACH USED FOR STEP 2	43
5.1 Risk Characterization	43
5.2 Decision Criteria (i.e. NLAA, EECs, RQs)	44
5.3 Deficiencies in the “Step 2” Approach.....	45
5.4 Modeling	47
5.4.1 Aquatic	47
5.4.2 Terrestrial.....	48
5.5 Monitoring Data.....	50
6.0 CONSIDERATION OF FURTHER REFINEMENT	51
6.1 Aquatic.....	51
6.1.1 Aquatic Modeling Parameterization and Species Relevance.....	51
6.1.2 Other Observations	52
6.2 Endpoints and Thresholds	53
7.0 COMMENTS ON THE USE/USAGE OF CHLORPYRIFOS, DIAZINON, OR MALATHION.....	54
8.0 NON-TECHNICAL COMMENTS OF A BROADER REGULATORY NATURE CONCERNING THE REGISTRATION REVIEW PROGRAM OR THE ENDANGERED SPECIES PROGRAM	58
9.0 DOCUMENT ERRORS AND TECHNICAL CORRECTIONS	59
9.1 Chapter 1 and its Attachments and Appendices.....	60
9.1.1 Chapter 1 Body	60
9.1.2 Chapter 1 Attachments.....	64
9.2 Chapter 2 and its Attachments and Appendices.....	65
9.3 Chapter 3 and its Appendices.....	74
9.4 Chapter 4 and its Appendices.....	79
10.0 TOPICS FOR CONSIDERATION AND COMMENT	88
10.1 Identification of "Best Available" Spatial Data to Represent Potential Pesticide Use Sites and Species Locations (Attachments 1-2 and 1-3)	88



10.2 Methods Used to Identify Potential Overlaps (and Extent) of Species Locations and Potential Use Sites and their Applications in Effects Determinations Made in Steps 1 and 2 (Attachment 1-6)..... 90

10.3 Estimation of Exposure in Various Aquatic Environments (Bins) that have been Regionally Delineated and the Parameterization of Bins and their Relevance across the Landscape 95

 10.3.1 Static Waterbody Parameterization and Methodologies..... 95

 10.3.2 Regional Aquatic Modeling Parameterization and Species Relevance 96

10.4 Evaluations of Exposures in Flowing Water Bodies and in Non-Freshwater Habitats 96

 10.4.1 Flowing Water Systems 96

 10.4.2 Appropriateness of Marine Bin Surrogates..... 102

10.5 Evaluation of Exposure to Terrestrial Organisms, including Dietary and Non-Dietary Routes of Exposure..... 103

10.6 Use of Species Sensitivity Distributions to Evaluate Effects (Attachment 1-5)..... 103

10.7 Characterization of Toxicity Data from Registrant Submitted Toxicity Data and Literature, and Utility of Sublethal Effects Data 107

 10.7.1 Submitted and Published Literature Data 108

 10.7.2 Use of Incident Data 111

 10.7.3 Sublethal Effects Data - Animals 111

 10.7.4 Sublethal Effects Data – Plants 112

 10.7.5 Interpretation of “Sublethal” in Context of the Assessment..... 112

10.8 Use of Mortality Effects Thresholds Based on Chance Effects 113

 10.8.1 General Issues..... 113

 10.8.2 Summary Data Arrays for Fish and Aquatic-Phase Amphibians 113

10.9 Weight of Evidence Approach Used, including the High, Medium, and Low Weighting Assignments to the Various Lines of Evidence to Evaluate Risk and Make Effects Determinations 114

 10.9.1 A Solid Approach to Considering the Weight of the Evidence Begins with a Well Designed Problem Formulation..... 115

 10.9.2 EPA’s Current Method on Weight of the Evidence Ignores Scientifically Established Approaches to this Exercise 116

 10.9.3 Improvements Needed 122

 10.9.4 High, Medium and Low Weighting Assignments 124

 10.9.5 Other Observations 125



10.10 Qualitative Marine / Cave Species Assessments..... 127
10.11 Other Types of Assessments..... 128



1.0 INTRODUCTION

CropLife America (CLA) has called upon its member companies' scientific and regulatory professionals, as well as consultants to them, to compile as comprehensive a response as possible to the topics that EPA has listed in its April 5, 2016 instructions for commenting on the draft Biological Evaluations ("Instruction Letter"),¹ given the time available. CLA does not endorse the interim process nor the methods being, by EPA's own admission at various public meetings, "invented along the way" as a FIFRA/ESA process is experimentally applied, automated beyond the ability to produce meaningful results, and put forth as "best available science," something that the contents of the dockets definitely do not support or embrace. However, CropLife understands the monumental task before the Agencies, given the manner in which they have decided to address it. We therefore seek to provide meaningful, constructive comments on the draft Biological Evaluations ("BEs") and their supporting documents and have made every effort to organize this response in the manner requested by the Instruction Letter.

The Instruction Letter identified twelve topics for consideration (under the Instruction Letter's section entitled "Topics for Consideration and Comment" and comment and 7 grouped items to be addressed in a preferred outline (in the Instruction Letter as "How to Group Comments")). However, not all of the topics for consideration are included in the outline for grouping comments. Therefore, we have used the outline where it covers the topics EPA identified, and added a section that completes the coverage of the topics EPA identified as of interest. In this way, CLA has given a comprehensive response organized in the manner EPA requested.

1.1 Participants in this Commenting Effort

Participants in the production of these comments include those individuals listed in Figure 1.

¹ USEPA Office of Chemical Safety and Pollution Prevention letter from Neil Anderson, Kelly Sherman, and Kevin Costello to the Chlorpyrifos, Diazinon and Malathion Registration Review Dockets (EPA-HQ-OPP-2008-0850, EPA-HQ-OPP-2008-0351 and EPA-HQ-OPP-2009-0317, respectively) dated April 5, 2016 and titled "Instructions for Commenting on the Draft Biological Evaluations for Chlorpyrifos, Diazinon, and Malathion"



Figure 1. Individuals Contributing to these Comments

From ADAMA:

Laura Phelps

From BASF:

Dan Edwards
Scott Jackson
Jiafan Wang

From Bayer CropScience:

Michael Dobbs
Brad Glenn
Tilghman Hall
Sean McGee
Matt McCool
Jane Tang

From Compliance Services (CSI):

Katherine Carr
Ashlea Frank
Jim Gagne
Jeff Giddings
Chad Jones
Bernalyn McGaughey
Robert Taylor

From CropLife America:

Janet Collins
Imad Saab

From Dow AgroSciences:

Pat Havens
George Oliver

From DuPont Agricultural:

Aldos Barefoot
Alan Samel

From Exponent:

Anne Fairbrother
Jane Staveley

From FMC:

Paul Whatling

Independent Consultant:

Bruce Hope

From Intrinsic:

Roger Breton
Dwayne Moore
Sara Rodney
Scott Teed

From Monsanto:

Joy Honegger

From NovaSource/Tessenderlo

Kerley, Inc.:

Kevin Henry

From Risk Analysis Solutions:

Nick Poletika

From Stone Environmental:

Michael Winchell
Lauren Padilla
Naresh Pai
Tammara Estes
Katie Budreski

From Syngenta:

Jonathan Maul
Natalia Peranginangin

From Waterborne:

Matt Kern
Gerco Hoogeweg
Nathan Snyder
Kendall Jones
Raghu Vamshi

1.2 Summary of Key Points and Recommendations

Key observations and recommendations in the sections of this comment document include but are not limited to:



- The hurried reaction pace driven by litigation deadlines has caused a leap into the use of the NAS Panel document as a policy guidance source for the procedural steps in a risk assessment. That transfer of authority is incorrect and confuses the exercise of FIFRA pesticide environmental risk assessment with the legal requirements for making no effect, may affect (NLAA/LAA) and jeopardy decisions under ESA. Incorrectly, flawed risk assessment procedures are being forced into the three decisional procedures of ESA, and in doing so, established FIFRA policy and environmental risk assessment policy have been ignored.
- The risk hypotheses as currently proposed are not set up to address the key factor/question in determining the risk to listed species. As written, there is an inherent assumption that exposure will occur, when one of the first questions that should be asked is whether or not there is a likelihood of exposure. It is only after it is determined that there is reasonable likelihood of exposure that the magnitude of exposure and potential effects should be addressed.
- The terrestrial and aquatic effects characterizations and the exposure assessments and their accompanying appendices, attachments and modeling tools suffer from a number of serious shortcomings that led to unreasonable overestimates of exposure and risk to endangered species for the 3 OPs. The shortcomings included use of unrealistic assumptions, incorrect or non-relevant endpoints, modeling errors, and failures to take advantage of best available biological information that would produce more realistic exposure estimates that are relevant to the species being assessed.
- The aquatic modeling was focused on defined aquatic Bins with a fixed set of dimensions (size and flowrate). The land area contributing to the fixed dimension Bin was modified by region of the country (defined by 2-digit hydrologic unit) following a complex, but not species specific analysis. The resulting aquatic exposure modeling presented in the BEs included some astronomically high EECs. The Agency has, in several public meetings, directly asked for help on aquatic modeling and has acknowledged that the methods envisioned for the Interim Approach are not providing output that is reasonable with respect to predicted concentrations, particularly in flowing water. CLA recommends that the Agencies reflect on the improvements we suggest here and rethink their approaches to modeling and the underlying aggregation methods (HUC units and bins).
 - The largest static bin, “Bin 7,” was described as a “High-volume static,” but is the standard 1 ha pond used in prior FIFRA assessments, not a new category representative of larger waterbodies as the name would indicate. Consequently, all static bins of smaller size are modeled using assumptions that go unrealistically and unjustifiably beyond the already conservative farm pond model. The biological evaluations fall short of using best data available. For example, the assessments include descriptions of aquatic species habitat and in many cases, actual waterbody names, but there is no use or synthesis of such data.
- For mosquitocide and other non-crop uses, there are ways to more accurately determine spatial distribution even in the absence of land cover and land use datasets. We have discussed an example of this in our discussion of ULV treatments of mosquitocides. We also observe that the



method the Agency used to portray nursery uses severely overstates use for many reasons, one of which is it combines pass-through “nurseries” with production and rearing locations.

- Little additional refinement was done in the iteration of Step 2, again resulting in missed opportunities to determine NE or NLAA outcomes that would simplify the overall process and conserve precious resources, particularly in the Services. Another difficulty with Step 2 is improper application of the weight of evidence process, so that any one of the lines of evidence can result in a LAA determination regardless of the weights of the other lines. Such improper application results in species not being screened out that are not at risk and that should be removed from further evaluation.
- Any additional non-fundamental attribute change such as altered behavior should be considered only when it can be linked to the three fundamental ones (survival, growth and reproduction) with exposure-response evidence that would reasonably be expected to occur and measurable in the environment. EPA has relied on the fundamental endpoints for decades when conducting appropriate risk assessments and this follows the recommendations in the 1998 Agency-wide (not just OPP) Risk Assessment Guidelines.
- EECs are estimated by predictive models and applied to the overstated spatial extent. The predictive models have both extreme conservativeness in model and scenario assumptions and large uncertainties in model accuracy and scenario relevance for listed species and their habitat requirements. Finally, none of this construct has undergone peer review or a validation process to insure the predicted exposures are fit for purpose, as described in the 2009 Guidance on the Development, Evaluation, and Application of Environmental Models, EPA/100/K-09/003 [2009 Guidance]).
- Rather than use grossly overstated predicted exposure values from faulty models as a line of evidence in the risk characterization, this portion of the analysis should instead rely on environmental monitoring data, when available, for chemicals like these that have been in use for many decades. Bringing in model predictions as a line of evidence should be deferred to the future when adequate models are available. EPA acknowledged that monitoring data is available and provides information that could be included, qualitatively, in the weight-of-evidence approach to support potential exposure pathways and exposure concentrations. However, monitoring data were not included as a line of evidence in the assessments.
- EPA’s assessments use generalized opinions of “possible” (but unlikely) exposures exceeding “possible” (but not probable) effects thresholds expressed by using terms such as “could,” “might,” or “is possibly.” These loose definitions all then combine with consistent EPA conclusions: that any level of exposure exceeding a threshold results in ecological adversity, and risk is claimed even when a particular assessment has great acknowledged uncertainty.
- Ranges of values are used instead of distributions and associated probabilities. The conclusions are not predictive, and are only precautionary, with no probability framework, as recommended in the 1998 Guidelines and the 2013 NAS report. The most useful risk hypotheses are those expressed in the form: “an exposure (x) to stressor (y) could, with probability (p), cause an



adverse response (z) of magnitude (m) in receptor (q)". Unfortunately, the BEs do not contain such useful risk hypotheses using quantitative risk characterization described in this manner.

- To establish relationships that are relevant to a valid risk assessment, the Agency should
 - (1) use taxonomic association between the listed species and laboratory species for appropriate effects thresholds,
 - (2) organize listed species by habitat type and apply appropriate exposure analysis based on these habitats,
 - (3) apply probabilities to risk analysis to get a better view of risk,
 - (4) consider effects endpoints against the frequency, magnitude and duration of exposure potential.

This range of information provides a good relative range of risk. Otherwise all combinations are lumped together and a poor risk analysis is the result.

- The biological evaluation should exclude inflammatory and irrelevant discussions in incident reports regarding illegal (or unknown) product applications. Further, only the association between current use patterns and confirmed incident reports should be considered.
- The Interim Approach document, published in advance of Agency BE discussions, and CLA concur on sublethal effects: they are relevant only when associated with growth, reproduction and survival and reproduction. The BE spends a considerable amount of time collecting sublethal data, arraying the range of information, and then assuming that the lowest effects values for these measurements translate into whole organism and species effects, with little explanation or justification of how this is related to growth and survival of the species.
- The Agency's "lines of evidence for risk hypotheses" are actually assessment endpoints, not lines of evidence, and this is more than a matter of terminology. "Lines of evidence" and assessment endpoints are two related but very different things. The Agency gives no indication of criteria for weighing lines of evidence, or criteria for making decisions based on overall weight of evidence.
- There is just one line of evidence (for toxicity) being applied to several different attributes for a given assessment entity. "Chemical stressors" is a variation on toxicity involving environmental mixtures or the parent a.i. alone. "Abiotic stressors" are measures of characteristics (such as temperature, bacterial/viral prevalence, or environmental pH) that could enhance an assessment entity's susceptibility to the toxicity of the parent compound. Thus this WOE analysis comes down to just the basic exposure vs. toxicity comparison approach expanded to include a larger number of assessment entities and their attributes. While there may be useful information about toxicity to be gleaned from such an expansion, it does not constitute a WOE analysis that weighs a number of truly different types of evidence (i.e., that beyond just exposure/toxicity) to support conclusions about whether a pesticide may or may not pose a threat to a given assessment entity.



- Use of open literature data is appropriate only if properly screened for scientific method and completeness of documentation. Studies that fail distinctive, clear objective evaluation criteria should be given little or no weight in the weight-of-evidence process.
- Since many open literature studies use non-standard test designs, relevance is also an issue for consideration before incorporating data into the risk assessment. Therefore greater transparency is needed with regard to data reliability and relevance of the data used in the assessment.
- Much data selected for use, often including those data values identified as defining the most sensitive species, are of questionable value and acceptability. In addition, since many of the endpoints are not from standardized studies, the hazard values are not consistent and are generally not correlated to duration of environmental exposures. We recommend that Agencies use standardized GLP studies in the identification of the most sensitive species. For other consolidation purposes (e.g. SSDs) in higher tiers of risk assessment, additional studies of confirmed scientific rigor may be helpful.
- The lowest effects values were used across studies to calculate thresholds and construct SSDs, without respect to the relevance and reliability of the endpoint, the underlying quality of the study, or the relevance of the study to the species of concern. A functional weight of evidence approach and adequate effects characterization of species sensitivity is lost with this approach.
- The Agencies address using separate thresholds from studies conducted with the technical grade and formulated product when available. Toxicity values are used from the lowest toxicity threshold if they are from studies with the formulated product. What is missing, however, is a discussion of the relevance of the formulated products to the uses they are labeled for and their behavior from an environmental fate and relevance perspective.
- The main deficiency in the lines of evidence used in the BEs is the reality that, in effect, only a single line of evidence is brought into the analysis: toxicity. As pointed out above, this is made clear on page A9 (PF)-9, where risk is determined only by comparison of a point estimate of exposure to a variety of toxicity thresholds.
- There is no justification for the conservatism in the threshold for indirect effects based on mortality. The current approach sets the threshold at the concentrations/dose estimated to cause 10% effect on to the most sensitive species, or 5th percentile most sensitive depending on data availability, of the taxa that are relevant to the specific species being assessed. This value implies that the assessed species fitness depends on 90% survival of the most sensitive relevant species, which infers an obligate relationship between the assessed species and the relevant species which is not the purpose of this threshold. Further, this approach neither considers the duration of effects and recovery potential of the relevant species nor the diversity of the diet or habitat of the assessed species.
- Temporal scale is particularly important for interpreting ecological adverse outcome, and since the OP insecticides assessed in the BEs have been used for more than 50 years, the entire approach of the predictive analysis (that assumes this Agency action involves release of a new toxic chemical into the environment) is conceptually unsound. If one considers a listed species



to be challenged by multiple stressors limiting recovery, and a specific chemical stressor (for example, chlorpyrifos) is evaluated in a predictive assessment for its marginal contribution to possible extirpation and found to greatly impact the fitness of individuals of populations for 99.9% of the species, then it is reasonable to ask the question why this predicted mass extirpation has not occurred after decades of potential exposure. This question is not found in the BE analysis but should be considered.

- Regarding indirect effects thresholds for listed species that consume organisms that are dietary items, there is no practical relationship stated or justified between the thresholds used and their actual relationship to indirect effects. Indirect effects thresholds should be stated, or dealt with qualitatively as the Panel suggested.

2.0 ENVIRONMENTAL MODELS USED IN THE DRAFT BEs – TERRESTRIAL EXPOSURE

2.1 Overview

For terrestrial modeling, EFED briefly described the tools used to estimate exposure of birds, mammals, herptiles, terrestrial plants and terrestrial invertebrates to the 3 organophosphate compounds. For Tier I, EFED primarily relied on the TED tool which integrates T-REX for birds and mammals, T-HERPS for herptiles, the earthworm fugacity model, and TERRPLANT for terrestrial plants. EFED estimated exposure to terrestrial invertebrates via several methods (e.g., diet, dose, soil) and KABAM or empirical BCFs to estimate exposure to aquatic prey consumed by wildlife. AgDRIFT was used to estimate the distance from treated areas at which exposure due to spray drift is no longer a concern for receptors. EFED conducted higher tier probabilistic risk assessments for 13 listed bird species using TIM to estimate acute risk and MCnest to estimate chronic risk.

The terrestrial and aquatic effects characterization portion of Chapter 2 and the exposure portion of chapter 3 and their accompanying appendices, attachments and modeling tools suffer from a number of serious shortcomings that led to unreasonable overestimates of exposure and risk to endangered species for the 3 OPs. The shortcomings included use of hyper-conservative assumptions, incorrect or non-relevant endpoints, modeling errors, and failures to take advantage of best available biological information that would produce more realistic exposure estimates that are relevant to the species being assessed. In addition, there are numerous examples of EFED not providing sufficient information and support to enable reviewers to understand how model input values were selected. We discuss some of the major issues below. Given the short timeframe for review, however, there are almost certainly more issues that will be found in the future.



2.2 Missing Information

2.2.1 Information Contained in the Biological Evaluation Dockets

Insufficient information was provided in the BE dockets for reviewers to critically evaluate or replicate terrestrial exposure analyses. There are numerous examples of this issue for both the screening-level and refined terrestrial exposure analyses. Because of space considerations, we provide only two examples here.

Example 1: EFED considered potential indirect effects on terrestrial organisms by examining, among other things, effects on their food sources. Aquatic prey organisms, if affected by the OPs, were considered to contribute to an indirect effect on the listed species consuming these dietary items. EFED stated that concentrations in aquatic prey were estimated using empirical Bioconcentration Factors ("BCFs"), the Pesticide Root Zone Model (PRZM5), and the Variable Volume Water Model (VVWM) (Section 6 of Attachment 1-7). The Agency does not describe, however, how the water concentration estimates were derived using PRZM5/VVWM.² For example, they do not specify what assumptions were made regarding model scenarios (e.g., which aquatic bin or bins?) and parameters, nor do they specify an averaging period, or whether a mean or selected percentile was used to derive the exposure concentrations for risk assessment.

Example 2: The method for estimating EECs for listed terrestrial invertebrate species is not described in the BEs, nor are EECs presented for listed terrestrial invertebrate species. This comment was also made in the letter from the OP registrants requesting extension of the comment period.³ EPA responded to this comment stating, *"Three types of EECs are calculated for terrestrial invertebrates: dietary based (mg a.i./kg-food), dose based (mg a.i./kg-bw) and in soil (mg a.i./kg-soil). EECs for different food items consumed by terrestrial invertebrates (e.g., broad leaves) are calculated in the same manner as described in Attachment 1-7 (see section 4). Dose-based EECs for terrestrial invertebrates that are above ground or soil-dwelling are also described in Attachment 1-7 (see section 4). The method for calculating exposures in soil is described in section 5 of Attachment 1-7. EECs are provided in the TED tools parameterized for each chemical (see response to previous comment for information on accessing this file) in the worksheets titled 'Min rate concentrations' and 'Max rate concentrations.'"* However, there are no dose-based EECs in the TED tool for terrestrial invertebrates. Moreover, to estimate dose, species body weight and food intake rate are required. No such information was provided in Chapter 3, Attachment 1-7 or in the TED tool.

² For more discussion of PRZM/VVWM, see Section 10.4 Evaluation of Exposure in Flowing Water Bodies and Non-Freshwater Habitats

³ Weinberg (Weiley Rein) and Menotti (Crowell & Moring) to Goodis, April 29, 2016.



2.2.2 Use of Best Available Data

Repeatedly, EFED failed to use the best available data throughout the assessments. In many cases, EFED defaulted to out-of-date and now invalid input values, and thus did not apply the best available information in parameterizing their terrestrial exposure models. Several examples are presented below.

Example 1: Numerous field studies have been conducted with flowable chlorpyrifos (as well as the other OPs in the assessments) and these should be used to determine chemical-specific residue unit doses (RUDs). Instead of using these relevant and best available data, however, EFED relied on the generic RUDs that were calculated by Fletcher et al.⁴ for all pesticides. The differences between chemical-specific and generic RUDs can be profound, because the properties of the product, the manner and setting in which it is used, and the application rates – as well as various other environmental factors – vary. For example, the generic mean and upper bound RUDs for above ground terrestrial invertebrates in T-REX and the TED tool are 65 and 94 mg ai/kg ww/lb ai/A (Table 3-17 in chlorpyrifos Biological Evaluation although units not specified there). Based on a comprehensive review of the available literature, however, Moore et al. (Section 1.2 of Appendix SI-3)⁵ determined mean RUDs of 1.58 and 14.6 mg ai/kg ww/lb ai/A for soil surface invertebrates and foliage-dwelling invertebrates, respectively, in orchards. The corresponding values in field and row crops are 4.16 and 13.7 mg ai/kg ww/lb ai/A. These best available data should be applied to refine the screening-level generic assumptions. Use of the chemical-specific RUDs would lower exposure estimates to realistic levels for listed insectivorous bird, mammal, and herptile species as well as many listed terrestrial invertebrate species.

Example 2: In Section 3.2 of the chlorpyrifos Biological Evaluation, EFED stated that a foliar dissipation half-life of 4 days was used “based on data reported by Willis and McDowell (1987)”⁶ to account for dissipation of pesticide residues on food items. This approach ignores more recently published information on foliar dissipation that would have enabled calculation of a dissipation half-life for each food item. As described in detail in Section 1.2 of Appendix SI-3 of Moore et al.,⁷ several high quality field dissipation studies are available that can be used to estimate half-lives for dietary items: below-ground invertebrates, foliage-dwelling invertebrates, and grass and broadleaf forage. The resulting mean half-lives (DT50s) of residue estimates for each food item are 3.12, 2.35, 2.92 and 2.09 days, respectively, which collectively indicate that EFED was overly conservative in assuming a foliar dissipation half-life of 4 days for every item of food.

⁴ Fletcher, J.S., J.E. Nellesson and T. G. Pfleeger. 1994. Literature review and evaluation of the EPA food-chain (Kenaga) nomogram, an instrument for estimating pesticide residues on plants. *Environmental Toxicology and Chemistry* 13:1383-1391.

⁵ Moore, D.R.J., R.S. Teed, C.D. Greer, K.R. Solomon and J.P. Giesy. 2014. Refined avian risk assessment for chlorpyrifos. *Reviews of Environmental Contamination and Toxicology* 231:163-217.

⁶ Willis G.H. and L.L. McDowell. 1987. Pesticide persistence on foliage. *Reviews of Environmental Contamination and Toxicology* 100:23-73.

⁷ Moore, D.R.J., R.S. Teed, C.D. Greer, K.R. Solomon and J.P. Giesy. 2014. Refined avian risk assessment for chlorpyrifos. *Reviews of Environmental Contamination and Toxicology* 231:163-217.



Example 3: The food ingestion rate models used by EFED in the TED tool and TIM are based on a data review conducted by Nagy in 1987,⁸ which is now considerably out of date. Food ingestion rate models have been updated and revised at least twice since the original publication (see Nagy et al., 1999;⁹ Nagy, 2005¹⁰). The first and most extensive update (Nagy et al., 1999) involved the addition of data for 12 eutherian mammal species (increasing the eutherian mammal dataset by 26%), 45 bird species (nearly doubling the avian dataset), and 20 reptile species (which increased the reptilian dataset by almost 60%). The updates are in readily available peer-reviewed publications and thus there is no reason for EFED to have not identified and used these newer and best available data.

2.3 Corrections to Errors in EFED Models

For the Biological Evaluations, EFED developed several new risk assessment tools, applied untested automation routines, and extensively modified some existing models. These tools contain thousands of calculation cells and lines of code and it was thus impossible to carefully and methodically run a quality assurance (QA) evaluation of each new or modified tool under the time constraint for comment response – an exercise which should be completed thoroughly when any model is developed, merged with another, or modified. In a random exercise conducted for testing the validity of the models, several lines of code were pulled from the methods and spot-checked with respect to some of the calculations they produce. In making this incomplete but telling random check, closer examination uncovered troubling errors. A few examples are described below, and it is certain that more will be found with the complete QA evaluation that is required to ensure that these tools are reliable, relevant and correct.

Example 1: In their worksheets “min and max rate – dietary concentration results” of the TED tool, EFED estimates the “Number of exceedances of thresholds and endpoints for upper bound and mean EECs.” These estimates are counts of the number of days that dietary EECs exceed threshold values per year. However, for terrestrial invertebrates consuming arthropods, the dietary EECs (i.e., mg a.i./kg diet) were compared to the dose-based thresholds (i.e., mg a.i./kg bw). This is an incorrect approach because dietary-based EECs, derived from dietary toxicity studies, cannot be compared to dose-based effects metrics, which are derived from single oral –gavage doses.

Example 2: In the TED tool, the Agency erroneously calculated the dermal route equivalency factor (F_{red}) as the oral LD50 divided by the log(dermal LD50) in column O of the “Min rate doses” and “Max rate dose” worksheets. Specifically, EFED neglected to raise the estimated log(dermal dose) to the power of

⁸ Nagy, K.A. 1987. Field metabolic rate and food requirement scaling in mammals and birds. Ecological Monographs 57:11-128.

⁹ Nagy, K.A., I.A. Girard and T.K. Brown. 1999. Energetics of free-ranging mammals, reptiles, and birds. Annual Review of Nutrition 19:247-277.

¹⁰ Nagy, K.A. 2005. Field metabolic rate and body size. Journal of Experimental Biology 208:1621-1625.



ten to obtain the correct estimate for F_{red} (i.e., oral LD50 divided by dermal LD50). An example used in the spot-checking QA exercise conducted in response to the BEs demonstrates the implication for the avian exposure estimates and illustrates that this error in the TED tool *overestimates the dermal contact dose by 200 fold*.¹¹

Example 3: In estimating exposure to listed wildlife predator species in the TED tool, EFED erroneously assumed that when a predator ingests vertebrate prey, it consumes their total daily intake, rather than their body burden. Prey body burden can be quite different from total daily intake for a variety of reasons but primarily because rates of uptake across the gut lining differ from rates of metabolism and excretion.

Example 4: The food ingestion rate parameters identified as applying to “mammals not in the Rodentia order” in the TED tool are incorrect and, in fact, apply to “all eutherians” assessed in Nagy (1987). The analysis that produced these parameter estimates was actually based primarily on rodent data (33 of 46 data points). Parameters for eutherians not in the Rodentia order were not reported for food intake rate by Nagy when he did his work in 1987.¹² Even in applying the older data to the model, it was done so incorrectly.

Example 5: In the TED tool, upper bound dietary EECs are incorrect for any listed species having large mammals as a dietary item (e.g., Florida panther with white-tailed deer and feral hog as primary prey). In the “min rate doses” and “max rate doses” worksheets, the cells in the “Upper bound Conc. in diet (mg/kg)” column for large mammal dietary items incorrectly reference the upper bound broadleaf plant EECs and not the correct “Large Mammals (1000g, short grass diet)” EECs. In short, EFED referenced the wrong cells, which results in the calculations that produce an incorrect value for all listed species having diets that contain large mammals.

2.4 EFED was Unreasonably Conservative in Selecting Exposure Scenarios and Input Values

Although it is appropriate to be conservative in the application of Tier 1 screening models, there are numerous instances of EFED being far too conservative, particularly when the best available data

¹¹ Consider the dermal contact exposure for the northern aplomado falcon (*Falco femoralis septentrionalis*) following a single application rate of 5.1 lb a.i./A of malathion. EPA estimated a dermal contact dose of 14,858 mg ai/kg bw. The upper bound dietary dose for a diet of arthropods was 110.3 mg a.i./kg bw. The estimated body weight for the falcon is 325 g. The surface area based on the equation provided in the TED tool is 473.6 cm². According to the Agency, F_{red} is 62.9, based on an estimated dermal LD50 of 2.16 mg ai/kg bw, which supposedly corresponded to an oral LD50 of 136 mg ai/kg bw. However, using the correct equation (i.e., log(dermal dose) raised to the power of ten), the dermal LD50 estimate should have been 145 mg ai/kg bw and resulting in a F_{red} of 0.94. When the correct parameter values are used, the dermal contact dose for the northern aplomado falcon is 74.4 mg ai/kg bw, which is 200 fold lower than the EFED estimated dose of 14,858 mg ai/kg bw.

¹² Ibid

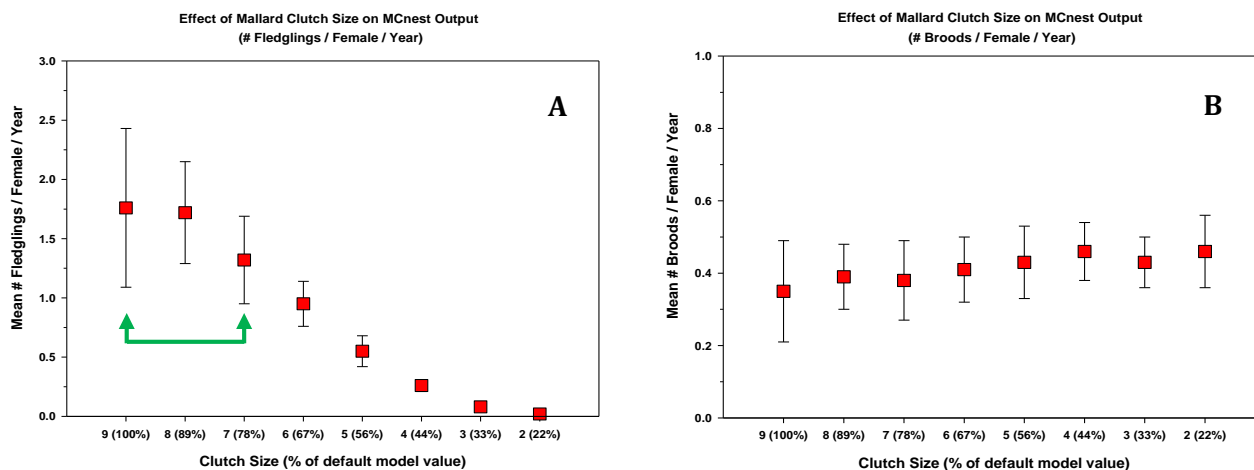


supported less conservative values. By introducing such unnecessary conservatism, the results of the Tier 1 screening process have little value. Further, the refined avian models, TIM and MCnest, are probabilistic models and therefore should have explicitly incorporated variability and uncertainty parameters rather than conservatively biased assumptions. In many aspects of these models, the recommendation of the National Academy of Sciences Panel (“Panel”) to use uncertainty analysis and probabilistic methods was ignored. The following examples illustrate the hyper-conservatism in EFED’s terrestrial exposure modeling.

Example 1: Exceedance of the reproductive toxicity threshold triggers a complete nest failure in MCnest. This assumption is overly conservative. For malathion, the reproductive toxicity threshold used in MCnest simulations was the NOEC for number of eggs laid, meaning that in laboratory studies, this is the level at which there were no observed effects on the number of eggs laid. To assess the biological and ecological significance of using percent reduction in egg production instead of assuming total nest failure, MCnest simulations were run using EFED’s assumptions and results from control or untreated groups of mallards. Using the model in this fashion, adjusting the data from untreated (unexposed) animals, we demonstrate the flaw in nest failure predictions over time. For a specific example, baseline or control conditions for mallard (i.e., no exposure) with default life-history parameters from the EPA’s mallard life-history library were used for subsequent model simulations, and were run reducing clutch size in a step-wise fashion. Fledglings per successful nest were reduced proportional to the clutch size reduction.

All other life history parameters remained constant at default values (i.e., baseline). MCnest output was

Figure 2. Effect of mallard clutch size on MCnest output under control (i.e., no exposure) conditions



compared across clutch size by visually inspecting 95% confidence intervals for Fledglings/Female/Year and Broods/Female/Year (see Figure 2).



Clutch size reduction to 78% of baseline resulted in a mean # Fledglings/Female/Year with substantial overlap of 95% confidence intervals with the baseline 95% confidence intervals (green arrow in Figure 2 above). Figure 2 (A) above demonstrates that if the percent reduction in eggs observed from an avian reproduction study is less than 22%, modeled fecundity responses from MCnest would be very similar to the baseline (i.e., clutch size = 9). However, in the refined avian analyses described in the OP Biological Evaluations, effects to 1 or 2 of 9 mallard eggs resulted in total reproductive failure. For the other OPs, the toxicity thresholds used in MCnest were different. For example, NOECs from avian reproduction studies based on a reduction in 14-d old nestling weight, an endpoint used in the other assessments, is highly unlikely to lead to the entire nest failure that EFED predicted. The assumption in MCnest that modeled doses above a selected reproductive toxicity threshold result in complete nest failure is overly conservative for a refined effects assessment because it is without scientific support and the degree of relevance of the endpoint to complete nest failure is not addressed or established in any way.

Example 2: To estimate inhalation dose to wildlife in treated areas, EFED assumed that all volatilized pesticide will be trapped beneath the crop canopy (see Section 7d in Attachment 1-7). This is obviously unrealistic and counterintuitive assumption. If a pesticide is sprayed downward, as in ground and aerial applications to field and row crops, interception will be greatest at the top of the canopy, which is exposed to the atmosphere. Thus, when intercepted pesticide volatilizes, much of it will diffuse into the open atmosphere, as opposed to an illusory closed system beneath the canopy. In scenarios in which the crop does not cover the ground entirely (e.g., orchards, early season field and row crops), there is little constraint on airflow below the canopy, particularly at the time scales of hours and days. Thus, the Agency's approach is overly conservative and not responsive to the portrayal of relevant physical processes. To make matters worse, the air concentration and respective volume of air respired were multiplied by 24 by EFED to estimate dose over a 24 hour period. This assumption is hyper-conservative because it does not account for any degradation or dissipation of the pesticide that would occur over a full day.

Example 3: In TIM and the combined TIM/MCnest, frequency on field (FOF) is the amount of time in a simulation that a bird spends on the treated field. TIM requires input values for mean, minimum and maximum FOF values to generate a beta-pert distribution of FOF values for the simulated species of birds. For agricultural fields, listed species that build their nests on the ground in grassland areas are defined as field residents, whereas other species are edge residents. For orchards and vineyards, field residents are those that build their nests on the ground in grasslands and woodlands as well as those that build their nests in the mid-story and canopy, whereas other species are edge residents. Mean FOF values for field residents and edge species are conservatively defined by EFED to be 97% and 69%, respectively, in agricultural fields and 87% for both field residents and edge species in orchards and vineyards. This generic approach to estimating mean FOF ignores a wealth of information available for nearly all listed bird species that should be used to derive much more scientifically defensible FOF distributions relevant to listed species. For example, the Kirtland's warbler exclusively forages in 5-23



year old jack pine forests at least 30 acres in size during the breeding season.¹³ They have never been observed foraging in pastures or other habitats resembling field and row crops, yet EFED assumed that they were an edge species and thus obtained 69% of their diet from treated fields. Even during migration, over 80% of documented migrant warblers (n=184) were found in shrub/scrub habitat, which is characterized as habitat dominated by woody plants ≤20 feet in height and similar in structure to their breeding habitat (Petrucha et al., 2013).¹⁴ An additional 10% of individuals were found in residential areas dominated by private yards.¹⁵ Less than 4% of migrating individuals were found in each of parks, woodlands (closed canopy forest), and orchards (fruit tree plantations; spring only). Thus, EFED was hyper-conservative in characterizing the proportion of the diet that Kirtland's warbler obtains from treated pastures during the breeding season.

As an aside, the fact that (1) populations of this bird are rising, (2) one of its main stressors is cow bird nest parasitism, and (3) its recovery is being used as an example of landscape scale, ecosystem-services-based management should be taken into account at some point. Significantly, most of its breeding habitat is protected and managed using clear-cut logging and controlled burns and/or herbicides conditions that give the warbler the nesting conditions it requires and also allow sustainable forest product harvest.

Example 4: EFED used the TED tool to estimate exposure for insect pollinators rather than using the Bee-REX model. This is inconsistent with the Agency's policy to use exposure estimates from Bee-REX to assess the risk of pesticides to all pollinator species, a fact has been vetted by their Scientific Advisory Panel. Predicted contact exposure using T-REX (TED tool) is approximately 50 fold higher than the corresponding estimates from Bee-REX, thus resulting in highly exaggerated exposure and risk estimates for listed insect pollinator species and listed species that prey upon them or listed plant species that are reliant on them for pollination.

2.5 Necessary Terrestrial Modeling Improvements

The shortcomings in the BE terrestrial modeling must be improved before such exercises provide relevant, reliable and reproducible results. The Agency is not using best available data and realistic assumptions in the exposure and risk characterizations. They are also inappropriately using exposure tools in risk assessment before they have been fully evaluated, quality tested and peer reviewed. At this point in the interim process, EPA has not provided the technical details needed for a critical evaluation of the terrestrial exposure assessment methodology.

¹³ Donner, D.M., J.R. Probst and C.A. Ribic. 2008. Influence of habitat amount, arrangement, and use on population trend estimates of male Kirtland's warblers. *Landscape Ecology* 23:467-480.

¹⁴ Petrucha, M.E., P.W. Sykes, Jr., P.W. Huber and W.W. Duncan. 2013. Spring and fall migrations of Kirtland's warblers (*Setophaga kirtlandii*). *North American Birds* 66:382-427.

¹⁵ Ibid



3.0 COMMENTS ADDRESSING AQUATIC MODELS

The Agency has, in several public meetings, directly asked for help on aquatic modeling and has acknowledged that the methods envisioned for the Interim Approach are not providing output that is reasonable with respect to predicted concentrations, particularly in flowing water. In this section we address aquatic model inputs, models and their parameterization and spray drift, and in particular spray drift and mosquitocide use assessment. The appropriateness of model selection and the importance of model parameterization in various settings are also discussed.

3.1 Model Inputs

Some of the exposure model inputs are derived from interpretation of the (1) pesticide label's Direction for Use; (2) properties of the compound that affect its environmental fate and transport; and (3) drift estimates from drift modeling. Appropriate model input is critical to accurate estimates of exposure. However, the best science and relevant data have not been used to establish these values. Examples of shortcomings regarding EPA's selection are discussed below.

3.1.1 Pesticide Application Methods:

The BEs for chlorpyrifos and diazinon describe that selecting application dates, "considers a number of factors including label directions, timing of pest pressure, meteorological conditions, and pre-harvest restriction intervals." However, this has not been done. In the case of both of these pesticides, application timing appears to have only been based on the selection of the wettest month. The MAL assessment appears to be the only one that incorporated pest pressure information into application date selection. Although the pest pressure information for MAL was provided, it was not clear how and for which scenarios this additional information influenced the final selection of application dates.

EPA acknowledges that "efforts may be made to avoid pesticide application right before precipitation events," and that "usage data are available showing that pesticide applications commonly occur a few days before the day of precipitation events." However, it does not appear that use patterns and geographical variations are considered in any way, and that is extremely important because species do occur with regional and seasonal specificity. Considering only a worst case to represent all uses and all locations deprives the assessment of the opportunity to eliminate species from concern in cases where application timing is different from seasonal species presence. There is no reference provided for the usage data EPA mentions, which prevents determining whether the usage data is representative of all use patterns and all geographical areas where the pesticide is applied. Additionally, if usage information are not based on long term trends, then the selected application date likely will not be a "reasonable estimate."



3.1.2 Environmental Fate and Transport:

The interpretation of the environmental half-life of a product is critical to its proper selection as an input to modeling. However, the BEs do not include any details about kinetic modeling to determine DT50s (Disappearance Time 50, the time required for the concentration of the test substance to decline by 50% of the initial value). The Agency states in the BEs that the DT50s are based on revised kinetic analyses using recent NAFTA guidance. However, detailed inputs and outputs from the revised analyses are not included, thus it is not possible to determine if the environmental half-lives are correctly derived and reproducible.

Additional environmental fate parameters are listed as ranges and do not detail individual values. Sources for the data are not always included. The BEs state that literature values were included in the environmental fate dataset; however, there is no mention of a process in which the literature data was deemed acceptable. Without confidence in the appropriateness of these parameters, their use as inputs to models cannot be validated.

3.1.3 Drift Estimates:

The drift methods followed standard practice for Tier2 FIFRA risk assessment and thus greatly overstate potential exposure. Each assessment assumed that the wind was blowing in the direction of the receiving water body at ~10mph for every application of the 30-year simulation and that drift reduction technology, which may be specified on the label or in best management practices, was not taken into account. The probability of applications actually occurring when the wind is blowing toward the water body should be stated to put data into context. Including probability in the assessment process was an explicit recommendation of the NAS Panel. Reflecting the label language and best management practices is also in concurrence with the Panel's recommendation to use best available data and science.

The drift values used are upper end estimates and do not reflect of realistic environmental exposure conditions. Adjusting the prediction by including variable wind speed, equipment factors, and wind direction over the course of a simulation would produce more realistic loadings.

3.2 Aquatic Models

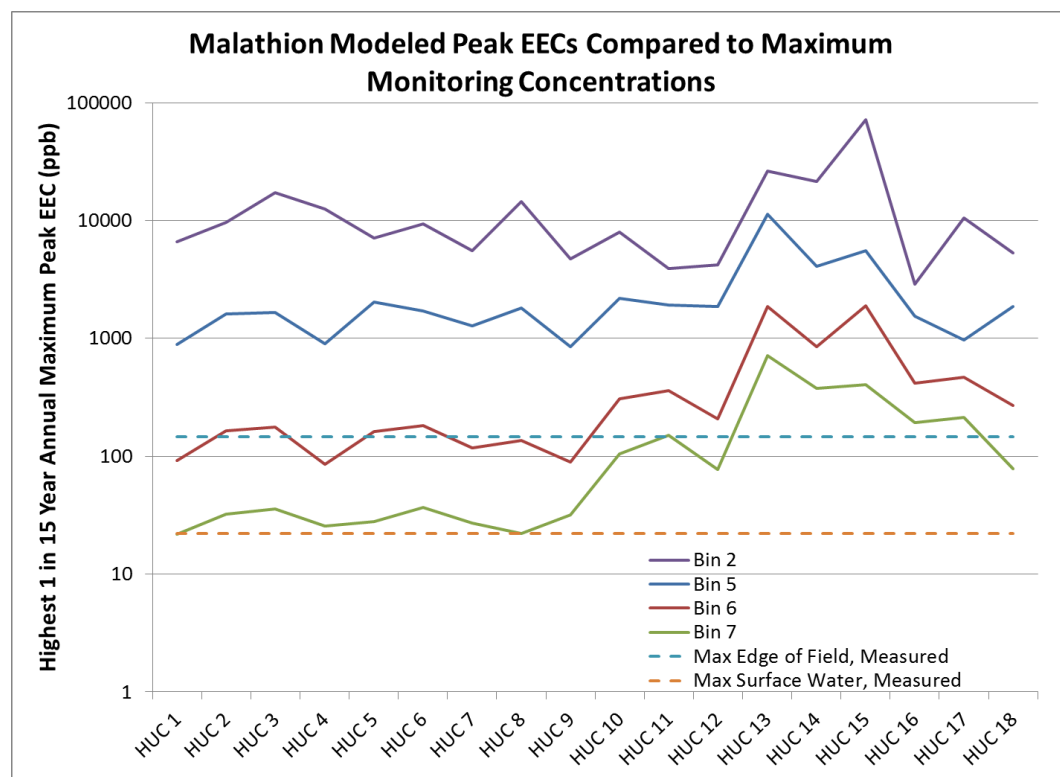
3.2 Aquatic Exposure Modeling

The aquatic modeling was focused on defined aquatic Bins with a fixed set of dimensions (size and flowrate). The land area contributing to the fixed dimension Bin was modified by region of the country (defined by 2-digit hydrologic unit) following a complex, but not species specific analysis. The resulting aquatic exposure modeling presented in the BEs included some astronomically high EECs. The modeled peak EECs reported in the BEs includes values as high as 71,800 ppb for malathion, 65,500 ppb for



diazinon, and 69,900 for chlorpyrifos (final EECs reported for chlorpyrifos were capped at its assumed solubility of 2,000 ppb). These EECs do not include the modeled values for Bin 3 and Bin 4, which were reported in the BEs to be as high as 393,000 for malathion, 2,730,000 ppb for diazinon, and 121,000,000 ppb for chlorpyrifos! The BEs acknowledged problems with this original set of modeled EECs for Bin 3 and Bin 4, and presented an evaluation of potential refinement approaches to improve the modeling results. Because the refinement approaches explored were not successful for all 3 OPs, the final approach used to arrive at EECs for Bin 3 and Bin 4 was to scale them from Bin 2 EECs by factors of 5x and 10x respectively. This resulted in predicted concentrations as high as 6,550 - 7,180 ppb in flowing water habitats representing major watersheds of main-stem rivers with average flow rates of 100 m³/s or greater. The highest concentrations ever measured for these three OPs in any surface water body are in the range of 15 ppb – 62 ppb, more than two orders of magnitude lower than some of the predicted concentrations in major rivers. The peak EECs predicted for in the BEs were up to several orders of magnitude higher than maximum measured concentrations across all the aquatic habitat bins, flowing and static. Furthermore, these peak receiving water EECs were several orders of magnitude higher than peak measured edge of field concentrations. Figure 3 below shows this comparison for Bins 2, 5, 6, 7 and 7 for malathion. These predicted exposure concentrations are not conservative or protective, rather they are implausible and uninformative for use in the endangered species risk assessment process.

Figure 3. Malathion Modeled Peak EECs Compared to Maximum Monitoring Concentrations



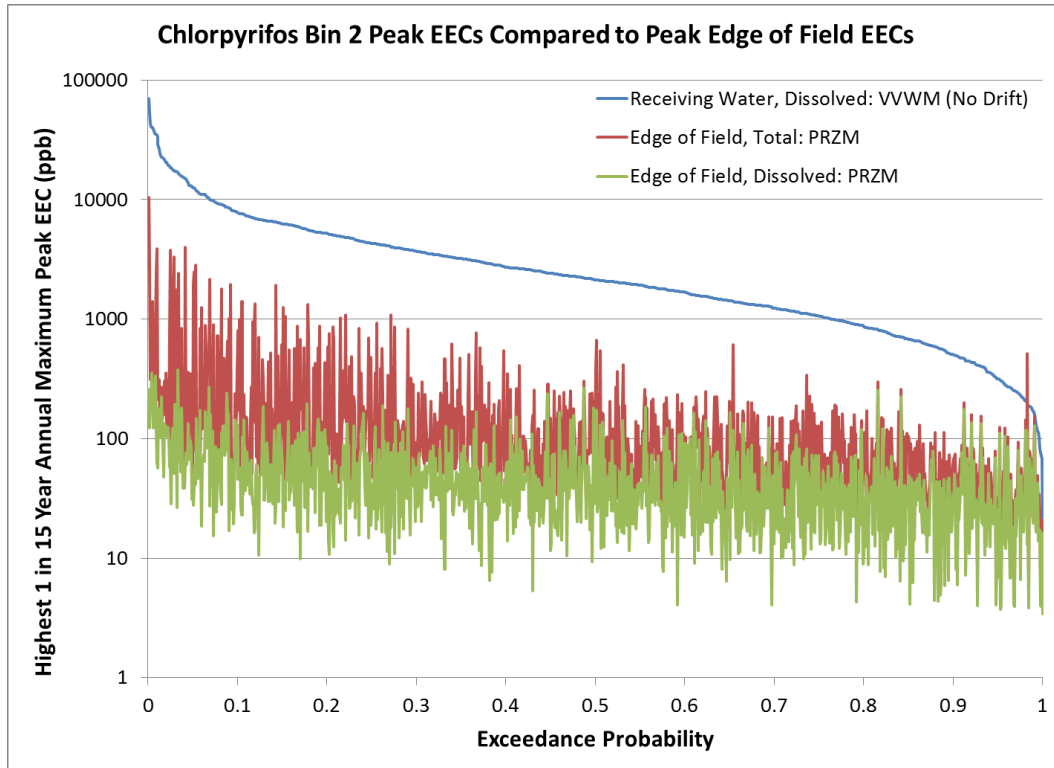


Further evidence of the errors in the predicted peak EECs in reported in the BEs can be shown by comparing the modeled edge of field concentrations to the peak EECs in the receiving water bodies. This analysis was conducted by re-running the BE scenarios with drift set down to a negligible value (0.001) so that a direct comparison could be made between runoff concentrations and the resulting receiving water concentrations. Figure 4 below is an example for chlorpyrifos and Bin 2 EECs. Both dissolved and total (dissolved plus sorbed) concentrations are shown. The exposure results shown are from 1,247 agricultural and broad area use non-seed scenarios that were run for chlorpyrifos BE. The dissolved edge of field concentrations are 1 to 2 orders of magnitude lower than the receiving water EECs across all of the scenarios. The total edge of field concentrations, much higher than we would expect to be relative to dissolved water body concentration, are still often an order of magnitude lower than the receiving water concentrations. This is physically impossible to achieve in reality and is evidence that the methods and assumptions that went into generating these exposure predictions contain fundamental errors.

The anomaly of receiving water EECs being higher than edge of field EECs is not limited to Bin 2. It occurs for some of the ESA scenarios across all of the four habitat bins modeled. Figure 5 shows the ratio of 1 in 15 year annual maximum peak receiving water EEC to the 1 in 15 year annual maximum peak edge of field EEC. Values of greater than 1.0 indicate receiving water concentrations greater than edge of field concentration. While in reality, we would expect receiving waters to provide, sometimes substantial, diluting effects, the predicted receiving water peak EECs often show the opposite.



Figure 4. Chlorpyrifos Bin 2 Peak EECs Compared to Peak Edge of Field EECs

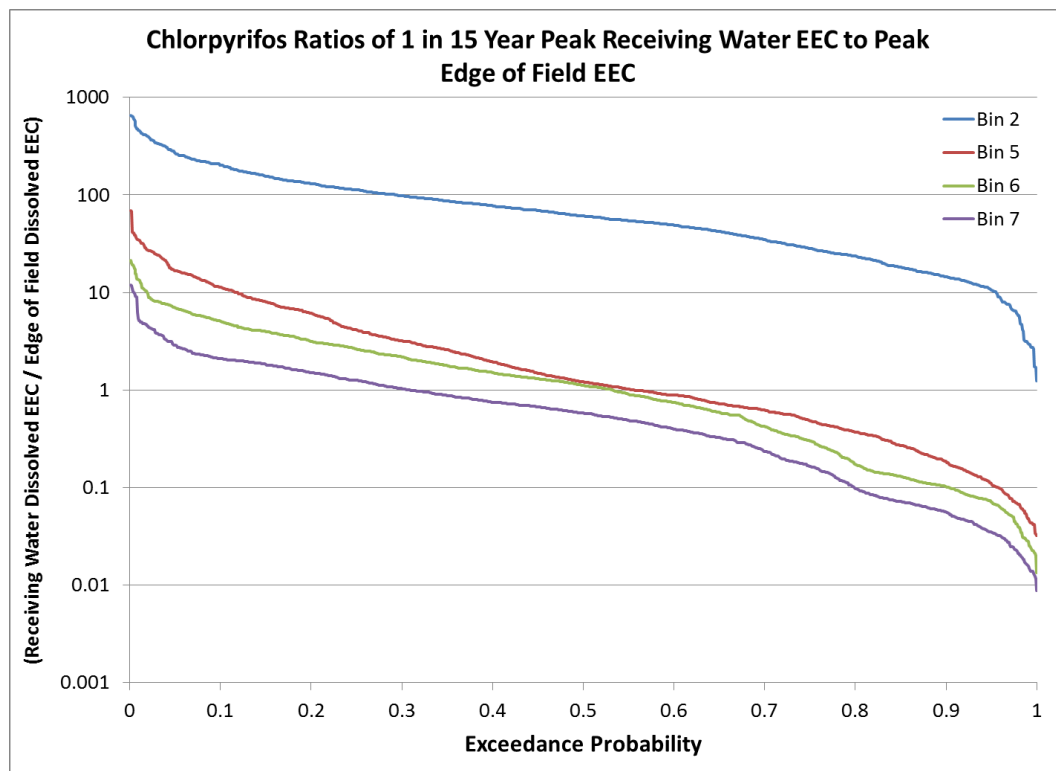


Representing the Crop Protection Industry

1156 15th St. N.W., Suite 400 Washington, D.C. 20005 • 202.296.1585 phone 202.463.0474 fax www.croplifeamerica.org



Figure 5. Chlorpyrifos Ratios of 1 in 15 Year Peak Receiving Water EEC to Peak Edge of Field EEC



CLA has provided comments on many aspects of the BE aquatic exposure modeling throughout this document. Many of these comments are directed towards very specific aspects of scenario selection, model selection, and model parametrization. While the hope is that these specific comments will lead to an improved overall modeling approach, they should not be received without also addressing the most significant issue that must be dealt with is the erroneously high EECs predicted for many of the scenarios resulting from large relative watershed sizes and the implausible assumption of instantaneous pesticide loading to a receiving water with no dilution from runoff water. Additionally, the biological information included in the BEs were not utilized to create species relevant exposure values or even provide at a minimum a confirmation of the weather, soil, or receiving waterbody appropriateness to specific species.

3.2.1 Scenario Development for the Agricultural Environment

The aquatic exposure methods relied heavily on models and tools available and validated for FIFRA assessments, including a suite of scenarios that have been developed under FIFRA and through peer



review. Those models over-predict actual environmental concentrations of pesticide residues in water.¹⁶ However, in the BEs, the existing models were modified and catchment descriptions altered in a way that introduced even more unrealistic characteristics and over-prediction in the model outputs. The existing scenarios for crops and soils were modified selectively, without a logical and defensible derivation of a new “scenario.” Instead, the crop-representative scenario was moved to a watershed/waterbody “scenario” by introducing selective changes that have not been validated for relevancy or realism. For example, modifications to account for a watershed-based approach were made by arbitrary changes to the assigned weather station(s) that would have been used in the existing crop-based scenarios. In existing scenarios, a fixed 10 ha treated area to 1 ha standard pond has been used to represent the “worst case” runoff situation. In the modified scenarios, the standard pond in the middle of the field was replaced with multiple fixed receiving water bodies and an elaborate method of changing the contributing area size, without justification of how these changes were relevant to the actual dynamics of an agricultural watershed area. Furthermore, suites of actual conditions surrounding listed species were not developed and evaluated, and new and relevant scenarios were not created. The scale of the watershed scenario is so large that the localized conditions that are specifically important to characterize species exposure are lost. For example, the watershed selected to represent the Pacific Northwest covers 177,523,042 acres and encompasses everything from temperate rain forest to arid desert conditions. Although the changes EPA has introduced appear to lead to a complex screening-level tier, they actually remove refinements and understandings that have been built up over the many years of FIFRA risk assessment and eliminate many species-relevant exposure values.

The arbitrary, broad brush approach described above is unresponsive to the Panel’s recommendations to use best available data, including data and expertise available from the USDA. Beginning in 2007, USDA’s National Agricultural Statistics Service (NASS) published census data by watershed.¹⁷ In contrast to EPA’s use of only 19 regions to represent the entire US, data from the Census of Agriculture are available for 38 individual land characteristics at the 6-digit Hydrologic Unit Code (HUC) level. To explain this further, EPA has represented their scenarios at the 2-digit HUC level, which for the Pacific Northwest spans 177,523,042 acres – only 27% or less of which is in farms. To better account for species and farm distribution, EPA should subdivide these regions using more refined USGS subregions. The HUC 2 region is divided by USGS into 221 subregions (4-digit HUC level) based on water flow patterns from the major rivers within the region. The subregions are further divided into 376 basins (6-digit HUC level). Even at this level of refinement, there remain an average of over 472,000 acres in a 6-digit HUC unit, much larger than the actual critical habitat and the contributing areas to many species found in very limited and defined locations.

¹⁶ Multiple presentations on this topic from task forces (FESTF, PWG for example) and registrants are available from EPA Public Workshop presentations, SETAC, ACS and CLA conferences and we will gladly provide specific copies to further support the details discussed here

¹⁷ USDA Census of Agriculture 2012 Census by Watershed.

https://www.agcensus.usda.gov/Publications/2012/Online_Resources/Watersheds/



3.2.2 Aquatic Exposure Modeling Tools

The Pesticide Water Calculator (PWC, running PRZM5 and VVWM) model is adequately documented and is a continuation of established exposure methodologies used in EPA FIFRA tiered assessment approaches. The assumptions and limitations are well known. The simple application of this tool with generic scenarios may be appropriate as a screening-level assessment, but these screening-level assessments do not provide species-specific results. This tool is inadequate for representing more complex landscape elements that are necessary to put species-specific exposure in context.

The model PFAM is used to represent exposure resulting from applications to cranberries, a “worst case” setting for aquatic exposure. The cranberry aquatic exposure setting is used to represent all crops and species aquatic exposure assumptions. Depending upon the BE, PFAM version 1.0 or 1.104 was used. Version 1.104 is not available on the EPA Pesticide models webpage, only version 1.0 is available. Furthermore, documentation, descriptions of crop scenarios used, and model output details are lacking and the results reported by EPA cannot be verified or duplicated.

Although PFAM can be used to simulate concentrations in a receiving water body, a conceptual model has not been developed and was not included in the ESA modeling effort... This method assumes species are exposed to in-paddy/bog concentrations because concentrations in flood water releases are used without considering the nature of the receiving area. This is inappropriate for organisms like fish, living in existing receiving waters in which dilution occurs, or for flood releases or overflows spilling into a dry environment (no dilution) where fish would not be present in what was a previously dry environment.

The specific details of the cranberry simulations were not provided, but in rice scenarios that have been provided by EPA, water levels and weir heights were set to the same height. In practice there would be room for paddies/bogs to have some rain without causing overflows/runoff. However, all rice scenarios are set up such that any rain amount greater than the evaporation on that day will cause overflow. Also, since the water level is maintained daily at that level in the scenarios, daily irrigation is triggered and the water levels never go down. These two factors are significant because exposure is based on overflow concentrations. Due to this set-up, overflows will happen much more frequently than in reality and will not allow degradation in water/sediment phases and photolysis to occur before overflows, as it would in a reasonable portrayal of a field setting.

As presented, the EPA Downstream Dilution tool (Appendix 3-5) has documentation of limited detail related to the actual methods, assumptions, inputs, and history of regulatory use. It is important to understand how the tool estimates a “distance” downstream that in turn predicts which concentrations are high enough to cause potential effects, or low enough to be below a level of concern. Explanation on these aspects is completely missing from the documents. Before using a new tool, EPA is obligated to provide source information, detailed methodology, QA/QC procedures and adequate documentation so that the public can verify their assumptions and repeat the analyses.



3.2.3 Parameterization of the Various Aquatic Bins

EPA's aquatic bins are defined based on a fixed, limited set of dimensions applied uniformly across the country. Although this was an attempt to move away from use of the single "farm pond" model, the bin dimensions are not based on best available science and the bins overlook species- and geographically-specific factors. The assumption that all aquatic endangered species can be represented appropriately by waterbodies with dimensions fixed to 9 options is flawed and a severe limitation of the assessment. For example, the biological evaluations include descriptions of widely diverse environments that are represented by a single bin, such as the description for "Bin 5, low volume" which is intended to represent "vernal pools, small ponds, floodplain habitats that are cut off from main channel flows, underground pools, and seasonal wetlands." Furthermore, the largest static bin, "Bin 7," was described as a "High-volume static," but is the standard 1 ha pond used in prior FIFRA assessments, not a new category representative of larger waterbodies as the name would indicate. Consequently, all static bins of smaller size are modeled using assumptions that go unrealistically and unjustifiably beyond the already conservative farm pond model. The biological evaluations fall short of using best data available. For example, the assessments include descriptions of aquatic species habitat and in many cases, actual waterbody names, but there is no use or synthesis of such data. These factual details need to be more than just mentioned in passing; the data can and should be applied to (1) customize model environments, (2) make landscape inputs representative of the known areas of species occurrence, (3) adjust bin standardization for important geographic specifics, and (4) determine if the assigned bin is representative of the described water body that is inhabited by listed species of concern. The current oversimplification of the tiered screening process results in exposure values that are not relevant or meaningful to species-specific conclusions...

3.2.3 Urban Use Scenarios

A new conceptual model for exposure in urban environments was presented. The new conceptual model was not backed up with data to show that they are realistic and representative of actual urban environments. Some specific aspects of the urban conceptual model scenarios that could use some further justification or help in understanding the assumptions made include the following:

- The neighborhood consists of ¼ acre lots with 1000 ft² building footprints. How was this lot size and house size arrived at and how does it compare to neighborhoods throughout the US?
- It appears that the neighborhood only consists of ¼ acre house lots, and does not include additional roadway areas as previous EPA residential conceptual models have (EPA, 2010¹⁸). The residential conceptual model description should include whether other impervious and pervious areas, not specifically part of house lots, make up a portion of a residential watershed.

¹⁸EPA (United States Environmental Protection Agency). 2010. Environmental Fate and Ecological Risk Assessment for the Registration of Indaziflam. Office of Pesticide Programs, Environmental Fate and Effects Division. June 1, 2010.



- The fence use site leads to a very large application area (especially for CPY). In addition, it is unrealistic to have these fences surrounding every lot in a neighborhood.
- Patios and other treated areas are not to scale on map (MAL scenario). 1,000 ft² would be the same size as the house footprint and nearly as large as the garden area. Given that this is a significant area, it should be better justified and represented to actual scale on the diagram to provide a better sense of the proportion of the lot these use sites occupy.
- The presence of a utility easement use site (CPY) across the back of every house lot in a neighborhood does not seem realistic. Are there examples of data to support this? This could be refined by an adjustment for probability or a percentage for expected incidence.

New conceptual models for urban exposure modeling were introduced for the CPY and MAL assessments. It is necessary that a conceptual model for urban pesticide use assessment should be consistently applied to all pesticides, even if some uses do not apply for some pesticides. In the current assessments, there are important differences in the conceptual models used for each compound, including:

- Presence of a utility easement impacting a conceptual house lot for CPY but not MAL
- Presence of a large trash storage area for CPY but not for MAL
- The size and extent of a fence assumed in each conceptual model
- Garden area and patio/garbage cans/shrubs/firewood/ornamentals were present in MAL but not CPY

The conceptual model for urban pesticide use should be consistent across all pesticides, even if some uses do not apply for some pesticides.

Many aspects of the conceptual urban model and its associated supporting assumptions have led to unrealistically high screening level EECs, generated without benefit of model validation. For example:

- Much of the pervious (porous) area use sites used the "Right of Way" scenario, which has a very high curve number (92). This curve number approaches the runoff rate of an impervious surface. The "Right of Way" scenario should be replaced with a more appropriately conservative pervious scenario, such as the "Residential" scenario. Use rates for some uses were outside the realm of reasonable worst case scenarios, and Directions for Use, such as the fence treatment for CPY that resulted in 15.5 lb/ac over the entire pervious portion of the lot and 1.2 lb/ac over the entire impervious portion.
- Assumption of a large impervious trash storage area use site (MAL scenario) equal to the size of the driveway represents an unreasonable large, high runoff use site area for a ¼ acre lot.
- The percent treated areas for the CPY and MAL scenarios were unrealistically high. For CPY, 100% of the residential neighborhood was effectively treated, while for MAL, 39% was treated. Both of these values are erroneously high, especially considering that neither of the two AIs have lawns on the label as a potential use site.



- The residential modeling approach does not account for run-on from rights of way into lawn, where infiltration would occur and reduce exposure potential. Models that are designed to simulate hydrologic transport in residential and urban environments, such as SWMM, account for this process and would be appropriate. Work conducted by the Pyrethroid Working Group (PWG) has demonstrated the use of SWMM in a refined residential pesticide exposure modeling assessment (Winchell et al., 2014¹⁹).
- Extrapolating the single lot conceptual model to watersheds of the size associated with Bin 3, Bin 4, and many Bin 2, Bin 6, and bin 7 scenarios greatly exaggerates the density and percent treated area. This is because the approach does not account for the vast areas of the urban landscape that are not house lots (parks, shopping centers, major roadways, industrial parks) and that medium to larger watersheds will have many different types of land uses throughout them.

3.3 Spray Drift Parameterization

3.3.1 Spray Drift Exposure Modeling

For ESA/FIFRA regulation, two models are currently used to predict deposition, AgDRIFT[®],²⁰ based on data generated by the Spray Drift Task Force, and AGDISP.²¹ Spray drift exposure estimates based on ground applications were derived from models of empirical data which were obtained from field studies designed to determine deposition of drift under circumstances prevailing at the time of the study. Additional empirical data may be found in the model used by the Pesticide Management Regulatory Agency (PMRA) that is responsible for pesticide evaluations in Canada. In Canada, PMRA uses the Agricultural Buffer Zone Workbook²² for estimating drift deposition based on data generated by Agriculture and Agri-Food Canada (AAFC).²³ Additionally, the RegDISP model uses the empirical deposition curves from the Spray Drift Task Force, and the Canadian Agricultural and Agri Food Canada studies (2000, 2004, and 2011) and represents the most complete collection of North American drift data for ground applications. The parameters can be varied to estimate drift deposition using available ground data that in turn depend on the variables monitored and controlled for in the underlying data.

¹⁹ Winchell M, Padilla L, Jackson S. 2014. Development and application of a higher tier urban modeling approach for estimating pyrethroid concentrations in static water bodies using SWMM and AGRO-2014. PWG Report - PWG-ERA-11. MRID 49327502. Stone Environmental Inc. Montpelier, VT

²⁰ AgDRIFT[®] Spray Drift Taskforce software, Version 2.0.

http://www.agdrift.com/AgDRIFT2/DownloadAgDrift2_0.htm

²¹ US Department of Agriculture Forest Service. 2006. AGDISP US Forest Service spray modeling software.

<http://www.continuum-dynamics.com/pr-agdisp.html>

²² Health Canada. 2005. Agricultural Buffer Zone Workbook, <http://www.hc-sc.gc.ca/cps-spc/pest/part/consultations/pro2005-06/index-eng.php>

²³ Wolf, T, and B.C. Cladwell. 2001. Development of a Canadian spray drift model for the determination of buffer zone distances. In Expert Committee on Weeds, Proceedings of the 2001 National Meeting, Quebec City, Sainte Anne de Bellvue, Quebec: ECW-CEM. D. Bernier, DRA Campbell, D. Cloutier, Eds



Typically, droplet spectra (VMD₅₀), release height, and wind speed can be changed. The model for aerial applications in both AgDRIFT and AGDISP is a mechanistic model that has been verified through comparison to field trials.²⁴

Aerial models have many other factors than can be adjusted to impact predicted spray drift. AgDrift aerial modeling is appropriate in tier III risk assessment for determination of spray buffers by air, although the default assumptions of constant meteorology, minimal surface roughness and absence of a crop canopy are likely to over-predict drift deposition at far distances downwind. For this and other reasons, AgDrift aerial modeling is not appropriate for aerial ULV modeling (see discussions for alternative approaches in the discussion of mosquitocides, below

3.3.2 Appropriateness of Model Selection

The outputs of the ground models used by North American regulatory Agencies, i.e., AgDRIFT, AGDISP, and the PMRA tool, are compared in Figure 6 below, which illustrates the magnitude of over-prediction of the two former models compared to actual field results. Both AgDRIFT and AGDISP deposition curves are for the same spray quality or VMD₅₀ (or nozzle). Both the AgDRIFT and AGDISP deposition curves should be close to the AAFC (Agriculture and Agri-Food Canada) flat fan data, which is consistent with the Spray Drift Task Force Data^{25,26} that the AgDrift model is based on. However, at 400 feet both models greatly over predict deposition. The PMRA model provides a good approximation of the AAFC field data assuming use of newer air induction nozzle technology. We can see from this evaluation that both AgDRIFT and AGDISP are not representative of currently used nozzles and drift reduction best management practices, while the PMRA tool provides a much better fit.

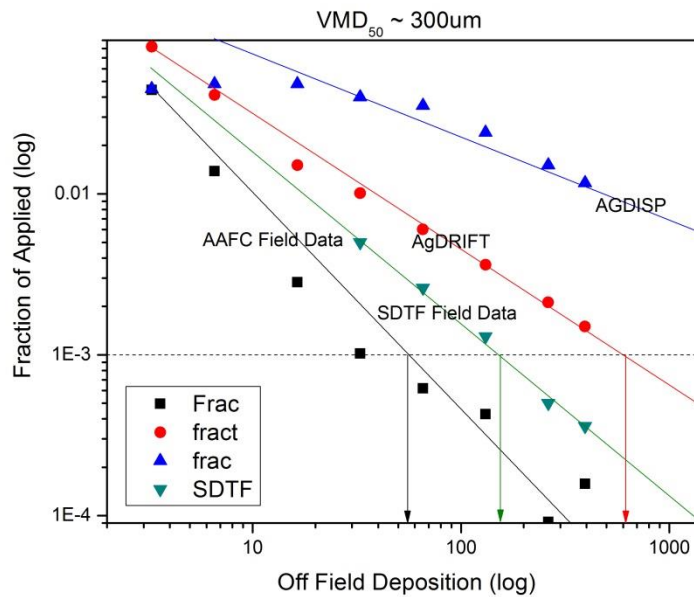
²⁴ Bird, S, S.G, Perry, S. L. Ray, and M. E. Teske. 2002. Environmental Toxicology and Chemistry, 21, pp.672

²⁵Johnson, David. 1992. Ground Field Study in Texas. Spray Drift Task Force, Sponsor study number F92-009, MRID No. 43493801 McKenna & Cuneo, 1900 K Street, N.W. Washington, D.C. 20006, Subdivision R- USEPA Pesticide Assessment Guidelines Pesticide Spray Drift Evaluation Guideline 202- 1

²⁶ Johnson, David. 1994. Ground Field Study in Texas Spray Drift Task Force, Sponsor study number F93-016 McKenna & Cuneo, 1900 K Street, N.W. Washington, D.C. 20006, Subdivision R- USEPA Pesticide Assessment Guidelines Pesticide Spray Drift Evaluation Guideline 202- 1



Figure 6. Deposition Curves from Models and Field Data



3.3.3 Current State of the Art for Ground Modeling - RegDisp

RegDisp predicts spray drift deposition from ground application of pesticides based upon regression equations constructed from data reported by Wolf and Caldwell (2001 and 2004), Wolf (2011), and the Spray Drift Task Force (1992 and 1993). The user can choose from these established deposition datasets by selecting study year, nozzle, boom height, sprayer speed, and wind speed. RegDisp also includes functionality that allows users to input equations constructed from their own data sets. The user chooses the form of equation and the required parameters and then saves their equation with a name to the RegDisp database. Once an established or custom deposition equation is selected, RegDisp estimates deposition as the fraction applied. Existing tools developed as part of the AGDISP interface allow the user to run the deposition and stream assessment tools. RegDisp is a modification of AGDISP 8.26, a tool developed by the USDA Forest Service to predict spray drift from aerial applications of pesticides. RegDisp has been integrated with AGDISP so that the user can choose whether to run AGDISP or RegDisp, and would be the more current and appropriate model to use in the assessment process.

3.4 Spray Drift and Mosquitocide Use Assessment

General Comments on ULV Applications

As noted above, the ULV mosquito adulticide use of malathion is unique compared to the other agricultural and non-agricultural uses, even other ULV agricultural and non-agricultural uses. For typical



agricultural or non-agricultural uses, the aim is to achieve deposition on the plant surface to be protected. As such, spray droplet distributions for these applications aim to maximize deposition and coverage, and applications are made in conditions that do not favor drift. In contrast, ULV mosquito adulticide applications aim to produce a cloud of droplets that remains suspended and moves through a zone where adult mosquitoes are flying. Thus spray droplet distributions have much smaller Volume Median Diameters (VMD or $Dv_{0.5}$ are much smaller), and ideal meteorological conditions for applications promote drift. Due to the diurnal activity of flying adult mosquitoes, applications are typically made in the morning and evening crepuscules.

The second major difference between ULV mosquito adulticide treatments and other agricultural and non-agricultural applications is that ULV mosquito adulticide applications are made by Mosquito Abatement Districts or other governmental departments of public health. EPA (September 23, 2009 – Draft (2009b)) estimated that there are 1200 such entities: 400 Mosquito Abatement Districts and 800 other public Agencies. ULV mosquito adulticide applications require specialized equipment to generate the aerosols for optimal mosquito control. Thus, applications are made by highly trained applicators whose aim is to maximize the concentration in the air and to minimize deposition (Mark Latham, Manatee County Mosquito Control District, personal communication). These state and local professionals are necessarily very familiar with listed species in their districts. They are also very likely to follow the ULV adulticide label as written. In the following sections, we explain an approach to mosquitocide evaluation that reflects both spatial distribution of applications and the directions for use that are on mosquito adulticide labeling.

Modeling Approach

For drift estimates, EPA has used the AGDISP model. However, there are few data available for reliable parameterization of this model of malathion ULV adulticide applications. One excellent study by Mickle et al. (2005) is available, we recommend that these empirical data be used instead of a model run. Drift predictions using AgDrift do not produce valid data for ULV applications.

Scope of Usage for Adult Mosquito Control

The footprint for mosquito control derived by EPA was based on the somewhat trivial observation that mosquitoes are widely distributed across the coterminous US, so the footprint is all of the US. This selection ignores the well-established fact that mosquito adulticides are applied by Mosquito Abatement Districts. A detailed analysis, based on actual use information, was performed as follows.

Assembling ULV Adulticide Usage Information

For the purposes of this assessment, usage information is defined as the application rate, timing, interval, and frequency for a specific geographic location. For mosquitocide and other non-crop uses, there are ways to determine spatial distribution even in the absence of land cover and land use



datasets. The approach that is described here is relevant to other uses that are not associated with a geographic land or use attribute.

Usage information for some agricultural uses is tracked by the GfK Kynetech – AgroTrak® database. This information can be matched with the geographic footprint for a specific crop from USDA, to evaluate the extent, timing, and intensity of pesticide use on that crop. For non-agricultural uses, including mosquito adulticide uses, Kline Inc. does make some market surveys, but their typical surveys provide limited information about timing, application rate, frequency, or intervals. Furthermore, because mosquito adulticides are applied to a wide variety of locations, such as urban or suburban residential areas, their environs, rangeland, pastures, or other uncultivated non-agricultural areas, such as wastelands and roadsides, there is no simple USDA location classification that can be used to show the geographic footprint of ULV mosquito adulticide use.

However, ULV mosquito adulticide applications are very specialized in that they must be made by certified applicators, using very specialized equipment. Indeed, ULV mosquito adulticide applications are typically made within specific boundaries by dedicated Mosquito Abatement Districts and other state or local departments of public health. Applications are based on monitoring of mosquito populations, complaint calls, or may arise from an emergency situation (e.g., hurricane), or public health threat (e.g. Zika virus).

To define the geographic scope of ULV mosquito adulticide use, 4 sources of information were summarized. The first was a survey conducted by Angela Beehler of the Benton County WA Mosquito Control District, for the American Mosquito Control Association (AMCA). The second was a survey of sales and distributors conducted by Cheminova/FMC Inc. The third was from publicly available usage information from Florida, and the fourth was from publicly available usage information from California. The methods used to develop these 4 datasets are described in turn.

The usage information from the American Mosquito Control Association is based on a series of surveys that covered the United States (Angela Beehler personal communication 2015). First, all members of the American Mosquito Control Association were emailed and asked if they use malathion, and if so, how many applications they make per year. This information was combined with the list of all AMCA members to cast a broad net as to the universe of potential users. This list of potential users includes organized mosquito control districts that did not answer the email survey, but that would be capable of making ULV mosquito adulticide applications. To supplement this list, entities that reported use in the California and Florida pesticide databases were added. Then, based on information from the basic producer of malathion, a list of entities that purchased malathion in the last 3 years was obtained. Finally, regional and state vector control associations were asked to provide their member lists and malathion users. We believe that these various “slices” of the universe of entities that use ULV malathion as a mosquito adulticide provide comprehensive coverage of the entities that have recently applied malathion as a ULV mosquito adulticide in the conterminous 48 states. The level of resolution of



the results is application at the county level. Information on application rates, timing, and intervals is not available from this survey. These data were previously provided to the EPA.

To obtain malathion ULV sales data, Cheminova/FMC tabulated all sales locations of its ULV mosquito adulticide products, for Cheminova/FMC and its major distributor, for the years 2011 to 2014. These sales locations were either assigned to states and counties directly, or assigned to cities or ULV applicators. The level of resolution of the results is application at the county level. Information on application rates, timing, and intervals is not available from this survey. This sales information is claimed as confidential business information.

In addition to these broad county level tabulations of potential ULV mosquito adulticide use sites, data were also obtained for the last 10 years from 2 states, California and Florida. The purpose of this exercise was to attempt to gauge the use rates, numbers of applications, and application intervals of ULV mosquito adulticide applications for these states, if possible. These factors were not covered by the AMCA or Cheminova surveys.

The Florida data were obtained from the website: <http://www.freshfromflorida.com/Divisions-Offices/Agricultural-Environmental-Services/Consumer-Resources/Mosquito-Control/Reports>

The dataset covers the years 2003 to 2014. Florida mosquito control data are summarized by active ingredient and county. Gallons of product, pounds of active ingredient, and acres treated are generally available for each county-year combination. Application methods are broken out by aerial and ground. Unfortunately, this Florida database does not provide the timing of applications, the number of applications, or the interval between applications. To estimate the intensity of usage in a particular county, the total county acreages were obtained, and then the total number of acres treated by air or ground were divided by the county acreage. This measure provides a crude estimate of the intensity of air or ground applications in each county.

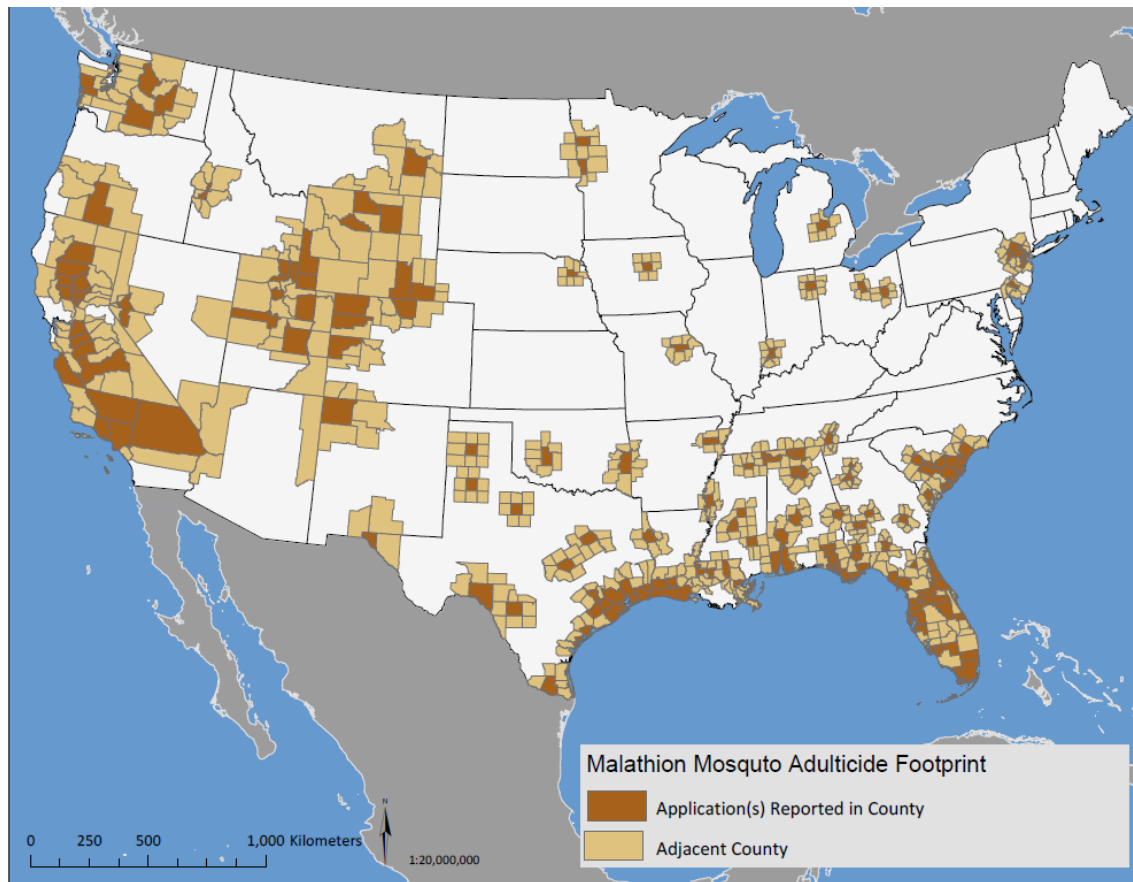
The California data cover the years 2004 to 2013, and were obtained from the website: <http://calpip.cdpr.ca.gov/main.cfm>

Only data for non-agricultural applications were summarized. There is a category called Public Health Applications that relates most closely to the ULV mosquito adulticide uses. Data were obtained by county for each year. Timing of applications was included in the database. However, unfortunately, there were several pieces of missing data, including whether the application was made by air or ground and the application rate per acre. As had been done for the Florida data, county acreages were obtained. To estimate the number of acres that could be treated, the total pounds of malathion applied was divided by the maximum ground application rate. This estimated number of acres treated was then divided by the total acreage in the county to arrive at an index of use intensity for that county.



The 4 datasets were combined to generate a list of counties where ULV malathion mosquito adulticide applications have been made. To conservatively ensure that the potential for drift was taken into account, all counties adjacent to the counties where malathion ULV applications were reported to have been made were also included. The results of this tabulation are shown in Figure 7. ULV malathion applications have been made in 159 counties, and there are 514 adjacent counties for a total of 573 counties with potential for exposure, as opposed to EFED’s assumption that use is in all US counties.

Figure 7. Counties where ULV Malathion Mosquito Adulticide Applications have been made, and Adjacent Counties



The business confidential nature of the usage information assembled by Cheminova/FMC precludes its detailed discussion here. However, two key points can still be made without compromising data confidentiality. These points are: (a) Scope of use -- although pest mosquitoes are widely distributed throughout the conterminous 48 states, actual ULV applications for adult mosquito control have been made in about half of the states and 159 counties; (b) Volume of use – Malathion usage has dropped considerably since 2004.



The relatively restricted geographic extent of use is a direct result of the fact that most ultra low volume applications to control adult mosquitoes are made by Mosquito Abatement Districts or other local, state, or federal entities in response to a direct threat to human health. The reduction in ULV malathion adulticide use is due to increasing preference for other organophosphates or for pyrethroids, but with the reservation that prevention of resistance build-up by chemical class rotation is very important to continuing to be able to control outbreaks of disease-carrying mosquitoes.

Finally, it is important to note that the Consultation Handbook specifically notes “Under no circumstances should a Services representative obstruct an emergency response decision made by the action Agency where human life is at stake.”

4.0 COMMENTS ON THE APPROACH USED FOR STEP 1 [MAKING MAY AFFECT [MA] OR NO EFFECT [NE] DETERMINATIONS]

4.1 General Comments

The Step 1 framework used in the draft BEs is unnecessarily conservative in that actual exposure potential is ignored. Instead, the spatial extent of use sites for both crop and non-crop uses is greatly overstated so that the co-occurrence (overlap) analysis for a potential MA determination is applied in a precautionary, not predictive, manner. Further, the exposure assessment tools used to estimate EECs have both extreme conservativeness in model and scenario assumptions and large uncertainties in model accuracy and scenario relevance for listed species and their habitat requirements. Finally, none of this construct has undergone peer review or a validation process to ensure the decision criteria are fit for purpose. In particular, there are serious errors with the approaches used to define spatial extent of use sites, model aquatic and terrestrial exposure, establish relevance of aquatic habitat bins, and relate effects thresholds to assessment endpoints. A true screening step could have been included in the analysis plan to eliminate species from further consideration. The use and validity of this approach has been demonstrated in three prior EPA endangered species case studies submitted to EPA by registrants (on atrazine²⁷) or FESTF (on 2,4-D²⁸ and clomazone²⁹).

²⁷ MRID 46078601. Overview; MRID 46078602. Summary of EPA-EFED Ecological Risk Assessment; MRID 46078603. National Risk Assessment for Aquatic Organisms and Critical Habitat; MRID 46078604. National Risk Analysis for Terrestrial Organisms and Critical Habitat; MRID 46078605. National Risk Analysis Summary of Results; MRID 46078606. Spatial Scoping Analysis for All Uses; MRID 46078607. Agricultural Usage Data Analysis; MRID 46078608. Site Specific Analysis; MRID 46078609. Taxonomic Specific Analysis.

²⁸ MRID 46535401. Endangered Species Assessment on Non-Target Plants Potentially at Risk from Use of 2,4-Dichlorophenoxyacetic Acid in Almonds, Rice, Strawberries, and Wheat.

²⁹ MRID 47647401. Frank, A.R. Clomazone: Analysis of Potential Direct and Indirect Effects on Threatened and Endangered Species. January 13, 2009. FIFRA Endangered Species Task Force (FESTF), Washington, DC and FMC Corporation, Princeton, NJ.



The hurried reaction pace driven by litigation deadlines has caused a leap into the use of the NAS Panel document as a policy guidance source for the procedural steps in a risk assessment. That transfer of authority is incorrect and confuses the exercise of FIFRA pesticide environmental risk assessment with the legal requirements for making no effect, may affect (NLAA/LAA) and jeopardy decisions under ESA. Incorrectly, flawed risk assessment procedures are being forced into the three decisional procedures of ESA, and in doing so, established FIFRA policy and environmental risk assessment policy have been ignored.

In 2003, OPP and the Services jointly published an advanced notice of public rulemaking on joint counterpart ESA Section 7 regulations (“Counterpart Regulations”).³⁰ These changes were proposed to better integrate the consultation process between the Agencies and increase its efficiency and effectiveness. While a portion of these regulations has been stricken, language survives that allows EPA to determine if an action is likely or not likely to cause an adverse effect (“LAA” or “NLAA” decisions). Given an agreed alternative consultation approach (an Alternative Consultation Agreement between OPP and the Services, which was adopted) and a scientific review of the adequacy of OPP’s program by the Services, now embodied in a letter from the Services to EPA, the Counterpart Regulations provide endorsement that OPP’s pesticide review and evaluation process can stand in for the Services except when a jeopardy determination is required. The important documents that brought the Counterpart Regulations into effect are:

- The proposed and final joint Counterpart Regulations^{31,32}
- The draft and final Alternative Consultation Agreement^{33,34}
- OPP’s “Overview Document”³⁵
- The Services’ evaluation of the Overview Document³⁶

It is important to note that the Overview Document was very much an integral part of the overall response to improving FIFRA/ESA consultation, and it documents OPP’s framework for risk assessment, which the Services’ evaluation endorsed. Following the Counterpart Regulations, in 2005, OPP published its final revised ESPP field procedures.³⁷ Notably, this document envisions a very interactive state involvement and a heavy dependence on monitoring data from terrestrial field studies and aquatic

³⁰ FR 68:3786 (January 24, 2003)

³¹FR 69:4465 (January 30, 2004)

³² FR 69:47732 (August 5, 2004)

³³ Draft Alternative Consultation Agreement for Implementation of Optional Alternative Consultation Procedures. EPA, FWS and NMFS, January 2004

³⁴ Final Alternative Consultation Agreement for Implementation of Optional Alternative Consultation Procedures. EPA, FWS and NMFS, August 26, 2004

³⁵ Overview of the Ecological Risk Assessment Process in the Office of Pesticide Programs: Endangered and Threatened Species Effects Determinations. USEPA January 23, 2004.

³⁶ Williams and Hogarth to Hazen: Evaluation of EPA Overview Document. FWS/NMFS January 26, 2004

³⁷ Federal Register 67 (231): 71549. Notice: Endangered Species Protection Program Field Implementation. November 2, 2005



systems, as well as a monitoring program to determine whether county bulletins are effective in protecting listed species and critical habitat. Services have agreed not only in the Counterpart Regulation process, but also historically, in the establishment of environmental risk assessment policy, that the criteria set by EPA for making risk determinations by the risk evaluation procedures they have established through policy are adequate to protect listed species.³⁸

4.2 Management of Data

In the draft BEs, EPA has substituted their reliable and proven risk assessment program with a series of automated, worst-case compilations that produce results that are impossible to follow or reproduce. For example, rather than work with a screening endpoint like a risk quotient, to establish toxicity distributions and compile available data, EPA uses a data array with all endpoints and ranges of values without apparent consideration of relevance or appropriateness in selecting relevant taxa and toxicity endpoints (the failure of this method is discussed in several sections of this comment document). With this method, species diversity and sensitivities are illustrated but not effectively used in the risk analysis. There is no anchoring policy or guidance to bring consistency to the risk assessment process, or to meaningfully put screening level data into context. If the data array and other newly introduced methods are utilized in the assessment their utility is limited to a first cut no effect/may effect determination. Subsequent validation, interpretation and relevancy of data needs to be applied, along with further quantification of other assessment parameters.

EPA states that it has followed its guidance on the use of open literature studies,³⁹ with modifications as described in Chapter 1, Attachment A8. EPA's guidance on the web site (hereinafter referred to as Open Literature Guidance) states that a study may be considered appropriate for "quantitative" use if it meets following:

- a) is reported in, or can be converted to, environmentally-relevant exposure units;*
- b) the endpoint is consistent with endpoints currently used to derive risk quotients (e.g. LC50 for acute exposure in fish;*
- c) "sufficient information must (emphasis added) be provided in the open literature to substantiate and/or independently evaluate whether the study conclusions (i.e. dose-response) and endpoints are accurate", and*

³⁸ Interagency Agreement between EPA/OPP, U.S. Department of Interior (USDI), Office of Protected Species (OES), and U.S. Department of Commerce (USDC), National Marine Fisheries Service (NMFS), 1980

³⁹ Available at

http://www.epa.gov/pesticides/science/efed/policy_guidance/team_authors/endangered_species_reregistration_workgroup/esa_evaluation_open_literature.htm.



d) the endpoint from the open literature study must be lower (i.e., more sensitive) than the lowest endpoint from a comparable registrant-submitted study for all taxon with the exception of aquatic-phase and terrestrial-phase amphibians, freshwater mollusks, reptiles, and for derivations of SSDs for other groups.

Chapter 1, Attachment A8 states that EPA used ECOTOX as a first screen. Given that the criteria for inclusion in ECOTOX are very minimal, this is appropriate for casting a wide net and as an initial screen. The second level of screening was OPP's, as described above. However, the only criterion that seemed to be retained was "a", regarding the units. Data with environmentally-relevant exposure units were used to build the arrays; the other criteria were not applied. This means studies with very low reliability and relevance can become part of the arrays, essentially rendering the arrays useless. Uncertainty is supposed to be included in the effects arrays (Chapter 1, page-51 in the chlorpyrifos BE) – but how uncertainty is used is not discussed. While the concept of arrays, to allow presentation of a great deal of data visually, is appealing, by including essentially all open literature without critical evaluation, EPA has presented a meaningless compilation of data.

It is stated that the open literature studies used for development of thresholds were "intensively reviewed" and only the study providing the lowest effect information was reviewed in detail. It is further stated in Chapter 1, Attachment A8 that the studies used for thresholds had to meet the definition of "quantitative" as described in the Open Literature Guidance. This means they would have had to meet the criteria given above in "b", "c" and "d." As an example for plants, if data were insufficient to construct an SSD (as was the case for the OPs, and as mentioned previously SSDs were not even considered for plants), only studies providing lower effect levels than registrant-submitted guideline studies were even considered (criterion "d"). This ignores potentially high quality plant effects data that could constructively inform the assessment, because at best, they would only be shown in the arrays (along with data of questionable quality). Moreover, EPA does not seem to have followed criteria "b" and "c" as illustrated by the examples below. It should be mandatory that, if an open literature study is to supplant a GLP-compliant guideline study for threshold derivation, it be of high quality.

4.3 Effects Determinations in Step 1 (Aquatic Assessment)

The step 1 aquatic exposure discussion in section 1.4.1.1.c.1 indicates several overly conservative assumptions that led to an unrealistic and inaccurate portrayal of the potential overlap of pesticide exposure and species location. The smallest representative water body and the highest representative use rate are paired whether this pairing initially is accurate or not, let alone realistic for different settings. To do this, EPA uses the concentrations generated for Bin 2 from the highest use rate in a "downstream dilution tool" that, through a spatial analysis algorithm, flags "Yes" and "No" pixels for inclusion or exclusion. Because the assumed concentrations arise from an extremely small flowing system combined with the highest use pattern, the starting value is an upper end estimate. This value is then applied to all land use pixels in the landscape, incorrectly omitting conditions for different use



patterns for different land uses. Layered on top of that incorrect procedure is the assumption that every pixel of land use is adjacent to a small flowing system and contributing runoff, erosion, and drift. All of these collective assumptions led to a footprint downstream that is not representative of a given use pattern and pesticide exposure potential. To correct this:

- Do not limit values to the concentrations generated from Bin 2 (low flow habitats, flow rate = 0.001 m³/s). The choice of bins should depend on the existence and habitat preferences of the endangered species being studied.
- Demonstrate or statistically evaluate the assumption that a 1-in-15 year exceedance of the EEC has proven ecological significance. This analysis would conform to NAS Panel recommendations to use best available science and methods as well as a probabilistic approach.
- Develop a methodology that more realistically represents entry and dilution in a flowing system. In the pilot assessments, EPA determines percent use area based on (1) an exaggerated portrayal of the chemical's use footprint and (2) the 'yes' and 'no' pixels from the binary land cover. Summing all the 'yes' and 'no' pixels within a watershed and within all such upstream watersheds grossly overstates the chemical use and potential exposure footprint.
- Modify the unrealistic assignment of EEC exceedance to achieve a more realistic portrayal of potential exposure. Assigning the 1-in-15 exceedance assumption to each of the 'yes' pixels is unrealistically conservative.

4.4 Effects Characterization for Terrestrial Plants

In the nontarget plant effects characterization, it is important to array the information appropriately for clarity on the risk characterization. All data should be used, including in field observations, etc. Integral to these assessments are the guideline required pre-emergent (seedling emergence) and post-emergent (vegetative vigor) studies that provide toxicity data for 10 species. It is well known that the results of these studies are conservative in that they are conducted in greenhouse conditions where the plants not field hardy, the soil has low organic matter, and the applications are made directly to the plants (e.g., no not mimic drift). This should be noted in the effects characterization.

Additional information is used in the effects characterization from the literature. Given that some of these studies (such as a chlopyrifos study noting reduced photosynthetic rate and stomatal conductance) are used in the effects determination, it is not clear they have undergone a thorough data reliability and relevance evaluation. For example, how does the endpoint noted above relate to the other data in the array.

In the current assessments, the nontarget plant data (Chapter 2) is presented in many ways, and hence can be a bit confusing. For example, sublethal effects are discussed for pre-emergence and post-emergence studies first; subsequently, these effects are discussed again but considered first for monocots and then for dicots. It would also make more sense to organize the table presenting the



thresholds by first presenting the data used for Step 1 (e.g., NOAEC values) and then the data used for Step 2 (LOAEC and EC₂₅ values). Another example, is in Figure 10-1 Summary of data array for chlorpyrifos. It is not clear what data are included in the array (NOEC, LOEC, EC₂₅ etc...), and if it makes sense to combine them, and what the red dot indicates (mean or median, but should include both). It should be made clear why the data are presented in different ways and how it will be used in the risk conclusions.

Generally, there is no presentation of the effects data from Chapter 2 in the WoE tables presented in Chapter 4. It is not clear by how much does the 'exposure' exceed the NOAEC or LOAEC. The weight of evidence sections in the effect determination should refine the analyses by noting the levels of conservatism in each step of the process, and beginning to integrate species biology into the conclusions. Using the *Pritchardia remota* example, the "high risk" direct effect to growth is determined based on a threshold value NOEC being exceeded for the minimum and maximum application rates at the site of a pre-emergent application, however, for post-emergence application, the LOEC value is not exceeded for the upper bound rate (4 lb a.i./A). Based on *Pritchardia remota* habitat being located in scattered, small groves in two valleys with elevations of 50 to 800 feet (information from WoE Appendix 4-3aa), actual in field or edge of field exposure would be rare at most. It is not evident that this WoE document included overlap in critical habitat from the CDLs, or the TED tool analysis for distance that risk extends from edge of field.

SSDs are proposed to be constructed "for direct effects to animals with robust data sets." What is the definition of sufficient data for an SSD? Why are SSDs precluded from being constructed for plants (assuming there are sufficient data)? For example, for chlorpyrifos there are data for 14 species of monocots and 31 species of dicots, yet the risk characterization only uses data for lowest endpoints. It is stated that "because of the variability in study designs and endpoints, it was not possible to derive an SSD with the available plant data" (page 215, Chapter 2 for chlorpyrifos – *check to see for other compounds*). It would seem that, at least considering the typical Tier II registrant-submitted studies, the data would be suitable, e.g., have similar study designs and endpoints, for at least 10 species.

2). For Step 1, the text on page A4(PF)-2 states the following:

- For direct effects to all taxa - sublethal endpoints (NOAEC, NOAEL, or EC_x):

For plants, the level corresponding to a reproduction/growth no observed adverse effect concentration or level (i.e., NOAEC or NOAEL) for the most sensitive species is used.

Then in Table A 1-4.1, it is stated to use the concentration equal to the lowest value among the available NOAEC and EC₀₅ values, implying an equivalence between the NOAEC and the EC₀₅. There is no published justification that the EC₀₅ represents a meaningful endpoint for a toxicity test with either



aquatic or terrestrial plants, from either a statistical or biological basis. [To the contrary, recent investigations have determined that a 5% effect is within the expected control variance for both aquatic and terrestrial plant standard toxicity tests, and cannot be reliably determined statistically (Staveley et al., 2014; Staveley et al., 2015). In addition, the biological significance of a 5% effect on plant growth is questionable. The reference to EC₀₅ should be deleted from Table A 1-4.1 in reflection of these considerations and to be consistent with the text.

3). In Step 2, for direct and indirect effect thresholds based on sublethal endpoints: Sublethal effects to plants are to be based on “the level that corresponds to the reproduction or growth NOAEC, NOAEL, or EC₀₅ for the most appropriate surrogate species” (Chapter 1, Attachment 1-4, page A4(PF)-4) – how is an evaluation of the “most appropriate surrogate species” made? Characteristics such as “Monocot, dicot, fern and allies, conifer, lichen” are included for each endangered plant species in the “Weight of Evidence” matrices associated with Chapter 4. Is the “appropriate surrogate species” based on this information, or merely based on aquatic vs. terrestrial? The text says that the lowest LOAEC or EC₂₅ is used from available terrestrial plant studies to assess indirect effects related to terrestrial and wetland plants; Table A 1-4.2 states this as the lowest LOAEC and EC₂₅ value from the available seedling emergence and vegetative vigor studies. The latter is much more specific and appears to preclude data from studies not conducted according to the standard seedling emergence and vegetative vigor guidelines. Table A 1-4.2 further states that for indirect effects related to *aquatic plants*, the lowest available LOAEC and EC₂₅ for aquatic plants is used. It is noted that standard aquatic plant toxicity tests typically report the EC₅₀, not the EC₂₅. Will registrant-submitted studies need to be re-evaluated to determine the EC₂₅ or does this imply that registrant studies will not be preferentially used when evaluating indirect effects related to aquatic plants?

5.0 COMMENTS ON THE APPROACH USED FOR STEP 2

5.1 Risk Characterization

The Step 1 procedure implemented in the BEs confuses the identity of the action area with the conduct of a step 1 screening level assessment to determine whether there is an effect or no effect as a consequence of the registration action. Currently, Step 1 is limited to just that: defining the action area. A missing step, whether called “Step 1” or “Step 2,” based on co-occurrence analysis could have screened out more species if individual uses were considered rather than combining all label uses allowed in a specific location. In addition, the tiered framework recommendation in the 1998 Agency-wide EPA Guidelines for Risk Assessment⁴⁰ suggests that a more extensive “Step 1b” could have involved a risk-based screening step to reduce the extent of the risk assessment. This was not done, and there are important consequences related to programmatic inefficiency and inability of the Services to apply their limited resources to complete high quality biological opinions of unnecessarily great scope.

⁴⁰ FR 63(93):26846-26924 (May 14, 1998)



Similarly, little additional refinement was done in the iteration of Step 2, again resulting in missed opportunities to determine NE or NLAA outcomes that would simplify the overall process and conserve precious resources, particularly in the Services. Another difficulty with Step 2 is improper application of the weight of evidence process, so that any one of the lines of evidence can result in a LAA determination regardless of the weights of the other lines. Such improper application results in species not being screened out that are not at risk and that should be removed from further evaluation.

Using the concept of fitness of an individual in risk hypotheses unfortunately makes the hypotheses less explicit and therefore somewhat ambiguous, or, at the least, difficult to follow throughout the analysis. Reference to the fundamental attribute changes of mortality, reduced growth, and reduced reproduction would be more explicit and more directly related to lines of evidence for both direct and indirect effects. Furthermore, they are consistent with ESA's basic definitions for protection of listed species. Any additional non-fundamental attribute change such as altered behavior should be considered only when it can be linked to the three fundamental ones with exposure-response evidence that would reasonably be expected to occur and measurable in the environment. EPA has relied on the fundamental endpoints for decades when conducting appropriate risk assessments and this follows the recommendations in the 1998 Guidelines. The BE documents do not explain why the change to the concept of fitness is necessary to improve the process, and there is no underlying policy or guidance to support such a drastic and unprecedented change. Under the OPP process that is codified in Agency-wide risk assessment policy, EPA is capable to determine the likelihood of its action to affect a listed species not only under FIFRA but under many other statutes. It is inefficient, and contradictory to the Agency's approach to risk assessment, to make unneeded and untested changes to such policy at an Office level.

5.2 Decision Criteria (i.e. NLAA, EECs, RQs)

No effect ("NE"): Absence of co-occurrence alone is not the sole criterion for a no effect determination. Co-occurrence with the action area as the sole criterion for a no-effect determination requires consideration of what is meant by both the "action area" and "range/habitat." Action area attributes include site of application and off-site drift, runoff, etc. Range/habitat attributes include amount that is in protected areas (e.g., parks, refuges, preserves), buffers between fields and range/habitat, etc. A definition of "overlap" also is required at this Step, and not delayed until Step 2. As it currently is applied, the criterion is inefficient and unnecessarily sends more work to later in the process where resources are limited.

Not Likely to Adversely Affect ("NLAA"): The rationale for selecting the threshold range overlap of less than 1% is not given. Because 100% of species and Critical Habitat exceed the threshold it does not appear to be a good discriminator, probably due to the overstated spatial extent of potential use. EPA can make this decision using the guidance of the Counterpart Regulations and their supporting policy and review documents.



Estimated Environmental Concentrations (“EECs”): The Step 1 framework used in the draft BEs is unnecessarily protective in that actual exposure potential is ignored. Instead, the spatial extent of use sites for both crop and non-crop uses is greatly overstated so that the overlap analysis for a potential MA determination is applied in a precautionary, not predictive, manner. EECs are estimated by predictive models and applied to the overstated spatial extent. The predictive models have both extreme conservativeness in model and scenario assumptions and large uncertainties in model accuracy and scenario relevance for listed species and their habitat requirements. Finally, none of this construct has undergone peer review or a validation process to insure the predicted exposures are fit for purpose, as described in the 2009 Guidance on the Development, Evaluation, and Application of Environmental Models, EPA/100/K-09/003 [2009 Guidance]]. In particular, there are serious acknowledged errors with the approaches used to model aquatic and terrestrial exposure. Given the recommendations in the 2009 Guidance and acknowledgment of the errors, the models applied to the BEs are not valid. Rather than use grossly overstated predicted exposure values from faulty models as a line of evidence in the risk characterization, this portion of the analysis should instead rely on environmental monitoring data, when available, for these chemicals that have been in use for many decades. Bringing in model predictions as a line of evidence should be deferred to the future when adequate models are available.

5.3 Deficiencies in the “Step 2” Approach

In Step 2, many of the deficiencies listed for Step 1 are carried over. Of particular importance are the overstated spatial use footprints and unvalidated conservative exposure predictions. Many refinements could be brought into both Steps 1 and 2 to improve exposure predictions and effects profiles and apply general aspects of biological attributes in the assessment to eliminate species from further consideration by assigning a NE determination based on low probability of occurrence of adverse effects at the appropriate spatial and temporal scales. Registrants are demonstrating this in a case study that uses an efficient, logical, reproducible, and scientifically valid processes to address a stepwise process of sorting and documenting NE/MA and LAA/NLAA findings.

EPA’s assessments, on the other hand, use generalized opinions of “possible” (but unlikely) exposures exceeding “possible” (but not probable) effects thresholds expressed using terms such as “could,” “might,” or “is possibly.” These loose definitions all then combine with consistent EPA conclusions: that any level of exposure exceeding a threshold results in ecological adversity, risk is claimed even when a particular assessment has great acknowledged uncertainty. Ranges of values are used instead of distributions and associated probabilities. The conclusions are not predictive, and are only precautionary, with no probability framework, as recommended in the 1998 Guidelines and the 2013 NAS report. The most useful risk hypotheses are those expressed in the form: “an exposure (x) to stressor (y) could, with probability (p), cause an adverse response (z) of magnitude (m) in receptor (q)”.⁴¹

⁴¹ Hope BK, Clarkson JR. 2014. A strategy for using weight-of-evidence methods in ecological risk assessments. HERA 20:290-315



Unfortunately, the BEs do not contain such useful risk hypotheses using quantitative risk characterization described in this manner.

Why is the absence of useful risk hypotheses important? The risk characterization for Step 2 depends upon a series of highly conservative assumptions to test the simplistic risk hypotheses. Because probabilities are multiplicative, not additive, the final result is an extremely unlikely event.⁴² It may not be necessary or prudent to be so conservative throughout an assessment that even a highly unlikely event will result in a LAA determination. For example, use of a 1-in-15 years 4-day average for an estimated water concentration as an exposure value will result in much more LAA determinations than use of a 21-d, 60-d, or 90-d average (for 1-in-15 years). Because the estimation of these average concentrations includes a concatenation of multiple conservative input values, the probability that a listed species will experience such a concentration is very small. When the impact represents a sublethal or indirect effect, the probability that such an effect would occur after only 4 days of exposure is even less, and this needs to be taken into consideration with respect to determining if the effect is likely to impact survival, growth or reproduction. And finally, the probability that a sufficient number of individuals experience such an exposure by being present and exposed to the 1-in-15 year, 4-day occurrence in such a way that actually results in decreased fitness, to a level that could possibly affect the population growth, is vanishingly small.

For Step 2 (NLAA/LAA), it is very difficult to discern what effects values are being used. The species-specific summary tables in Appendix 4-3 state the types of endpoints that are available, but are not very clear about which one was used (e.g., NOAEC or LC20?) nor do they present the actual numeric value. The effects endpoint used in Step 2 should no longer be the ultra-conservative 1-in-a-million threshold, so a policy decision needs to be in place to define what level of effect will be acceptable. Policy guidance at OPP, and agreed to by the Services, is in place for this determination and should be honored conceptually (HED ERA SEP) or modified formally through FIFRA procedures for substantive policy changes. Additionally, a spreadsheet showing the effects values used for each species in Step 1 and Step 2 would be very helpful, as would including the numeric values into each of the species-specific summary files in Appendix 4-3.

The current step 2 aquatic methodology does not utilize the detailed locations and use site information in a substantive way, and thus ignores best available data. Instead, aquatic exposure was estimated using assigned bins (based loosely on assumed habitats generically) and the HUC2 location, with no further refinement to account for known habitat, known location and actual waterbodies identified as supportive of specific species.

Best available data need to be incorporated into the assessment. The NAS Panel identified what several of these (mostly government) databases are and also recognized that data developed by registrants was a potentially useful source of knowledge. Further refinement and filtering should be employed based on

⁴² Hope BK. 2012. Exposure gone "wild": A call for rational exposure scenarios. HERA 18:485-487



such data, including (1) designated critical habitat, (2) land cover representative of known habitat; (3) known waterbodies in the vicinity of a given species where an aquatic input is important; (4) actual potential use sites adjacent to or up gradient from these locations; and (5) customized exposure assessment to reflect the actual conditions relevant to the species (soil type, weather, crop types, wind direction, elevation, etc.). While the use of multiple data sets may seem daunting, their aggregation and assimilation into the may affect determination can, and in several cases has, been accomplished.⁴³

What are called “step 2 exposure estimates” in these assessments are not improvements on the basic crop-species action area determination that is done in Step 1. They actually backtrack from Step 1 because EPA has applied unjustified caveats, illogical areas of concern, and unproven taxonomic relationships when used as “surrogate data” in assumptions that attempt to overcome the deficiencies in the presented methodologies. The exercise is less meaningful as a screen than are the current methods in use, some of which are described in the Overview Document, and it cannot be carried forward to step 3 as if it met species specific endpoint needs. Unfortunately the conservative assumptions, noted numerical issues, assumed drainage area to receiving water ratios, and limitations of using a fixed receiving water body (regardless of actual habitat or hydrology in location of species), do not constitute a “refined analysis.”

5.4 Modeling

5.4.1 Aquatic

The aquatic modeling approach followed in Step 2 remains an ultra-conservative, screening level assessment. If this approach had been followed with corrections leading to sensible inputs and models appropriate for the types of water bodies being simulated (e.g., a flowing water model for Bin 2, Bin 3, and Bin 4), then this analysis would have been reasonable for Step 1 of the assessment. The approach followed does not make use of the species critical habitat, known location, or known waterbody information appropriate for a Step 2 analysis for making LAA/NLAA determinations. Nearly all of the comments concerning the BE aquatic modeling throughout this document have been focused on the Step 2 methods, as screening level aquatic exposure modeling was omitted from Step 1. As has been discussed throughout this document, the most significant flaws in the BEs aquatic modeling approach followed in Step 2 include:

- Overly generalized scenarios assigned to represent vast HUC2 hydrologic regions with significant variability in agronomic and soil conditions.
- Minimal accounting for climate variability in regions where significant differences are found over very short distances.

⁴³ For example, through FESTF’s Information Management System.



- Use of a single, highly conservative PRZM scenario to represent all locations of a given crop group within a HUC2, regardless of the proximity or relevance of this scenario to the range of an individual species.
- Highly conservative, screening level assumption for critical model inputs such as Percent Cropped Area (PCA) and Percent Treated Area (PTA) that are implausible for many of the aquatic habitats modeled and not relevant to the species being assessed.
- Representation of watersheds larger than a field size while maintaining a homogenous areas from both the landscape characteristics standpoint and an agronomic standpoint.
- The extrapolation of a single field and static receiving water model to scales both larger (Bin 3, Bin 4) and smaller (Bin 5) than these models were intended to represents.
- Incorporating assumptions concerning watershed drainage area and resulting drainage area to normal capacity ratios (DA/NC) that violate the conceptual model of a static water body with no overflow, and completely break physical laws by assuming massive instantaneous inputs of pesticide runoff load into a receiving water with no change in water volume.
- The inappropriate use of VVWM, a model designed to represent a static water body with overflow, to simulate open channel flowing systems.
- Uniform application timing across large watersheds oversimplifies the realism occurring at the watershed scale and species-specific level of assessment, overstating acute exposure substantially.
- The lack of a probabilistic element to the exposure modeling that accounts for variability and uncertainty in model inputs and assumption as recommended by the NAS report.
- Not utilizing actual habitat data provided in the BE to prove appropriateness of tools employed for the species being evaluated or modify assumptions to develop species relevant scenarios.

Overall, a considerably more refined, species-specific aquatic exposure modeling approach is necessary at Step 2 in order to have sufficient data required to make LAA/NLAA decisions for each species.

5.4.2 Terrestrial

For assessments of endangered birds and mammals in its biological evaluations of malathion, chlorpyrifos, and diazinon, EFED introduced the TED model. According to information provided by EFED, the TED model is a umbrella-type model that compiles several of EFED's more standard, well-known terrestrial organism exposure and risk models, such as T-REX, T-HERPS, SIP, and STIR into a single model. However, information provided in the BEs suggests that EFED has added components to the TED model that are not included in the underlying models that TED is based on, such as additional feed item categories, and a food chain biomagnification component. Unfortunately, the TED model is not readily available for download and trial from EPA's pesticide website, or its pesticide models website.



Therefore, a direct evaluation of the TED model is only if the model can be verified as the intended on for use.⁴⁴

Based on EFED's description of the TED model, this model is expected to have similar underlying assumptions and limitations as those found in the T-REX, T-HERPS, SIP, and STIR models. These assumptions include scaling factors that may inaccurately modify toxicity endpoints (LD50 and NOAEL values) of endangered species, and inaccurate assumptions concerning the daily feed ingestion rates of different sized (weight) animals. Based on EFED's description of the TED model, we presume that these same inaccuracies are included in the TED model. Key assumptions of the T-REX and T-HERPS models, which we also presume are included in TED, also include assumptions that birds and mammals ingest only a single category of feed item daily (e.g., eat 100% short grass), and that 100% of an animal's diet is consists of pesticide-treated feed. Although the T-REX and T-HERPS models include modules for evaluating both maximum estimated residues on feed items and median or typical estimated residues on feed items, in recent registration reviews EFED has based its assessments only on maximum estimated residues; therefore we presume EFED made the same assumption about endangered animals only ingesting feed items with maximum estimated residues for its TED analyses.

Another significant limitation of the T-REX and T-HERPS models, which we presume is also included in TED since TED is apparently based on these models, is that the T-REX and T-HERPS models compare both acute and chronic toxicity endpoints to a single-day, acute EEC. For chronic effects, this represents a mismatch of toxicity and exposure, particularly since the vast majority of chronic toxicity effects do not result from a single-day exposure, but rather the chronic toxicity endpoints used by EFED are based on continuous exposures of birds and mammals to a pesticide over a period of months, primarily in reproduction studies. Detailed examples are given in the terrestrial modeling sections that follow.

Also, the T-REX and T-HERPS models assume that residues from multiple applications, which EFED modeled for its malathion, diazinon, and chlorpyrifos endangered species assessments, are linearly additive on feed items. Again, based on EFED's description of TED, we presume this assumption also applies to EFED's TED model. This assumption of residues from multiple applications being linearly additive may greatly simplify the modeling exercise, but based on actual residue trials, this assumption is greatly overestimates residues on feed items and is unrepresentative of potential exposure under actual use conditions.

EFED conducted higher tier probabilistic risk assessments for 13 listed bird species using TIM to estimate acute risk, but these results are more or less presented as an "example" of what could be done - - but has not been done. Such refinement should be performed for the listed terrestrial species and, with aquatic modeling, for aquatic species, that failed in their Tier I analyses - which, as it now stands in the draft BEs, represents the vast majority of listed terrestrial species. However, prior to applying TIM to a refined assessment, the model itself requires correction of the inherent flaws that are currently in it.

⁴⁴ The TED model is available through Internet sources, but the location is obscure and the status of the model is not known.



The BE develops thresholds around the most sensitive sublethal endpoints irrespective of the demonstrated relevance of the endpoints for risk characterization (e.g. AChE inhibition in *H. azteca*). Sublethal thresholds should be based on growth, survival, reproduction and growth endpoints that can be used in Steps 2 and 3.

5.5 Monitoring Data

EPA acknowledged that monitoring data is available and provides information that could be included, qualitatively, in the weight-of-evidence approach to support potential exposure pathways and exposure concentrations. However, monitoring data were not included as a line of evidence in the assessments. Monitoring data should be used more explicitly in the analysis. For example, a systematic analysis of the proximity of detections to species locations should be conducted and included as a line of evidence. The highest monitoring data concentrations reported were 1 to 3 orders of magnitude lower than the modeled aquatic EECs. However, this was not accounted for to assess the realism of the new modeling scenarios and approaches. The significant discrepancy should have been reason for assigning high uncertainty to the modeled EECs and led to the development of alternative modeling approaches.

Other models also need further evaluation, validation and vetting. EPA indicates that preliminary investigation of bias factors and SEAWAVEQ is underway. These tools would enhance the ability to use general monitoring data quantitatively. Mosquin's work, published in the Journal of Environmental Quality in 2011 should be considered in developing robust bias factor approaches.⁴⁵ Additionally, the WARP model, a conservative exposure screening tool, was used to estimate concentrations of chlorpyrifos and diazinon. WARP-predicted concentrations provide valuable information about potential locations and upper bounds on the magnitude of pesticide exposure that should be considered in a weight of evidence approach. Although EPA conducted an analysis with WARP, it did not factor that analysis into LAA/NLAA determinations. WARP concentrations were orders of magnitude lower than concentrations predicted with PRZM/VVWM models. Discrepancies of this magnitude require explanation and further investigation to determine which, if either, of the models may be appropriate for exposure modeling.

⁴⁵ Estimation of Upper Centile Concentrations Using Historical Atrazine Monitoring Data from Community Water Systems. Paul Mosquin, Roy W. Whitmore and Wenlin Chen. Journal of Environmental Quality 41 (3): 834-844.



6.0 CONSIDERATION OF FURTHER REFINEMENT

6.1 Aquatic

6.1.1 Aquatic Modeling Parameterization and Species Relevance

The revised approach to modeling aquatic exposure is conceptually not different from screening level models reviewed in the Overview Document. As a screening level assessment, the use of standard scenarios is in line with typical FIFRA risk assessment methods, and standard scenarios are appropriate to use as an early screen of high-end exposure values representing a “worst case” situation, but they do not support a final determination of exposure to endangered species. The source scenarios that have been in use in screening level assessments were developed to represent major production areas, using soil with a moderate to high runoff potential and incorporating agronomics of the specific high production crop area. For the endangered species assessments, EPA made only one modification to these scenarios, and that modification is of questionable scientific relevance, and definitely not a modification that brings any specificity to endangered species assessment. Changing the standard crop scenarios to HUC2 scenarios was accomplished only by a modification of the weather station source, and it is not correct to assume the source is representative of the HUC2 area as a whole; nor does this change do anything to make the model more “specific” to endangered species. In fact the opposite is true. In many cases, only one or two scenarios are used to represent the entire country (Table A3-1.4, Attachment 3-1). It is critical to note the deficiency of this approach because the use of scenarios in this manner does not lead to a representative exposure value for soil, weather, or cropping practices in agricultural areas that may be in proximity to listed species. A focused assessment on species of concern must rely on the best available data for the location of the species of concern. While the biological evaluations espouse an extreme volume of exposure estimates, it is impossible to identify if any of these values are relevant to the unique and localized conditions where a given endangered species is found. The use of HUC2 as an organizational unit for scenarios, weather stations, and crop parameterization is not appropriate for species with known locations. For an exposure relevant to the species perspective, the core data must be better tailored to the species location.

Aquatic exposure modeling at Step 2 should move beyond simple screening level approaches that use a single conservative PRZM simulation to predict EECs in flowing water bodies draining from heterogeneous watersheds. Step 2 aquatic exposure modeling approaches should include the following:

- Representation of the heterogeneous landscape through explicit simulation of the land uses and soils that comprise a given watershed.
- Spatially explicit predictions of EECs that can be associated with species habitat locations.
- An accounting for variability in pesticide application timing that occurs at the watersheds scale. Over larger areas, it is likely that applications timing will vary over a wider window than only the



application interval, and would have significant effect on peak EECs. Accounting for application date variability should be considered in Step 2 aquatic exposure modeling.

- Both PCA and PTA should be incorporated for all 3 flowing bins, as well as the medium and large static bins. These factors are used by HED in their risk assessments for human exposure. PCAs can be calculated based on aggregate use site footprints from multiple crop groups, and PTA, while based on state-level data, can be meaningfully included in the determination of EECs at Step 2. PCAs should be determined independently for each habitat bin and be based on appropriately sized watersheds (e.g., NHDPlus catchments for Bin 2, HUC10 to HUC12 for Bin 3). The use of a state-level PTA should transfer directly to the larger Bin 4 watersheds, which are at the same scale. State-level PTA can be used in several ways for Bin 2 and Bin 3, either as multiplicative factor to PCA, or, by splitting the EECs into percentages assuming worst case 100% PTA and the remainder assuming 0% PTA (no exposure) to then determine the likelihood of exposure at the level associated with use on all labeled crops. Because PCA and PTA are so important in determining exposure concentrations, these factors should be considered rigorously at Step 2, rather than waiting until Step 3.
- Padilla et al.⁴⁶ provides a proposed methodology for using the PRZM and SWAT in refined site-specific modeling of flowing water habitat.

6.1.2 Other Observations

Data and parameters that are representative of the species' local habitat and chemical use footprint should be used. Specifically:

- Rainfall patterns across a HUC-2 significantly vary. Assigning a few SAMSON weather stations for each HUC-2 is not representative of the vast diversity it presents. For an expanded, but simple view, there are about 242 SAMSON weather stations across the United States⁴⁷ that represent data from 1961-1990. Gridded rainfall data from NOAA (NCEP/NCAR)⁴⁸—available at a higher resolution of 2.5 x 2.5 degrees and covering a more recent time period (1961 – till date) and data derived from these rainfall grids closest to the use footprint area and species locations are more representative of the local area, and incorporates the most current trends in rainfall patterns in that area.

⁴⁶ Padilla, L., M. Winchell, N. Poletika, and P. Havens. 2016. A watershed-scale refined flowing water exposure modeling approach for endangered species assessments. USEPA Environmental Exposure Public Meeting, Arlington, VA. May 9, 2016

⁴⁷ https://www.epa.gov/sites/production/files/documents/CITY_MAP.PDF

⁴⁸ <http://www.esrl.noaa.gov/psd/data/reanalysis/reanalysis.shtml>



- Soils data and classification provided by SSURGO (with the most recently released gSSURGO) offers the details needed for a more robust and local scale assessment.
- Cropping patterns based on planting and harvest dates are key to identifying the chemical use footprint area. USDA's Plant Hardiness Zone (PHZ) map⁴⁹ provides sub-state level information, based on temperature, to identify areas that are most likely suited for growing crops. USDA's Field Crops Usual Planting and Harvest Dates⁵⁰ reported at the state level indicate the dates/periods in which the crops are planted and harvested in most years based on 20 years of historical crop progress estimates, as well as the knowledge of industry specialists. The above two datasets in conjunction provide a more detail assessment and local understanding of the crops to determine chemical use footprint and time of chemical application (which can be particularly important if species presence or increased vulnerability has a temporal aspect)

6.2 Endpoints and Thresholds

The results of reliable and relevant studies⁵¹ on organism responses to target chemical exposure need to be used to set the thresholds that are then inputs for modeling. Therefore, ensuring that those threshold values are set appropriately is critical to the outcome of the risk assessment exercise. The draft BEs do not use best available data, relevant studies or valid estimates to set these thresholds, as is described in detail below.

Toxicity to taxon from exposure to other chemical stressors of concern and non-chemical stressors (*e.g.*, temperature) are discussed. However, stressors often most critical to the health and recovery of the species are not considered. Either the specific action should be considered (individual registration) or the stressors that are most likely to put species at risk should be mentioned and added into the effects determination (*e.g.* habitat loss, urbanization, weather events, invasive species etc.). This is critical to understanding if the action has an impact on the baseline for the species of concern. For example, if a Registration Review action is to specify the number of treatments that can be made per season, and an added buffer to prevent soil erosion, when before that action, use was unlimited and there was no buffer, then the Registration action improves the baseline rather than further degrades it.

It is stated that the EPA does not typically request toxicity studies for amphibians from pesticide registrants, but rather uses data on freshwater fish to represent potential effects to amphibians in the

⁴⁹ <http://planthardiness.ars.usda.gov/PHZMWeb/>

⁵⁰ <http://usda.mannlib.cornell.edu/usda/current/planting/planting-10-29-2010.pdf>

⁵¹ The validity or adequacy of a study is often thought of as having 2 components: reliability and relevance. Reliability is "the inherent quality of a test relating to test methodology and the way that the performance and the results of the test are described." Relevance is "the extent to which a test is appropriate for a particular hazard or risk assessment." (Technical Guidance Document on Risk Assessment, European Commission, 2003; Klimisch et al., 1997)



aquatic phase. If amphibian data is available and sufficiently relevant and reliable it should be used as surrogate information for listed amphibians.

Questions or uncertainties can occur related to using surrogates to represent potential effects to individual listed species. However, a penalty should not be imposed for having effects data of increased relevance. If data is available for a species more taxonomically similar than other laboratory species tested then this data should be used if of sufficient quality.

7.0 COMMENTS ON THE USE/USAGE OF CHLORPYRIFOS, DIAZINON, OR MALATHION

EPA has adopted a “*weight-of-evidence approach to risk assessment when making determinations (for listed species and designated critical habitats) for the three pilot ESA pesticides (chlorpyrifos, diazinon, malathion).*” In implementing this approach, EPA assumes that a given compound is used on every acre of crop, every year, for the maximum number of applications permitted on the label at the maximum rate. Yet, compelling evidence, readily available and previously used by EPA, disproves these assumptions.

As shown in Figure 8, the U.S. Census of Agriculture shows that a total of 100.7 million acres of agricultural land were treated by **any** means to control **all** insects in 2012, which is 25.8% of cropland and 11.0% of the land in farms. The same Census information is available for each of the 386 6-digit HUC watershed basins.

USDA has extensive pesticide use survey data, including acres treated, application rates, and number of applications per treated acre. USDA data are available for individual pesticides for crop and state combinations where pesticides are typically used. Private survey data are also available, such as that collected annually by GfK.⁵²

Major field crops account for much of the total usage of the three pilot pesticides, chlorpyrifos, diazinon, malathion (CDM), but the percentage of acreage of major field crops treated is small. Survey statistics reveal that CDM was used on 2%, 14%, 0.9%, and 0.3% of the U.S. acreage of wheat, hay, cotton and rice, respectively. Acreage treated is higher for specialty crops such as oranges, strawberries, cherries and apples, averaging 92%, 94%, 27% and 32%. Percentage of acreage treated with CDM is lower, in some instances much lower, than the above numbers, to the extent that individual pesticides are used on the same land and to the extent that multiple treatments are made.

⁵² The only major data gap is the extent of usage of individual pesticides on the same acre, but this gap could be closed in future USDA/EPA/NASS surveys.



Regional percentages vary considerably by crop. For example, in Michigan in 2011, chlorpyrifos was used on 67% of the apple bearing acreage, but only 9% of cherry acreage.

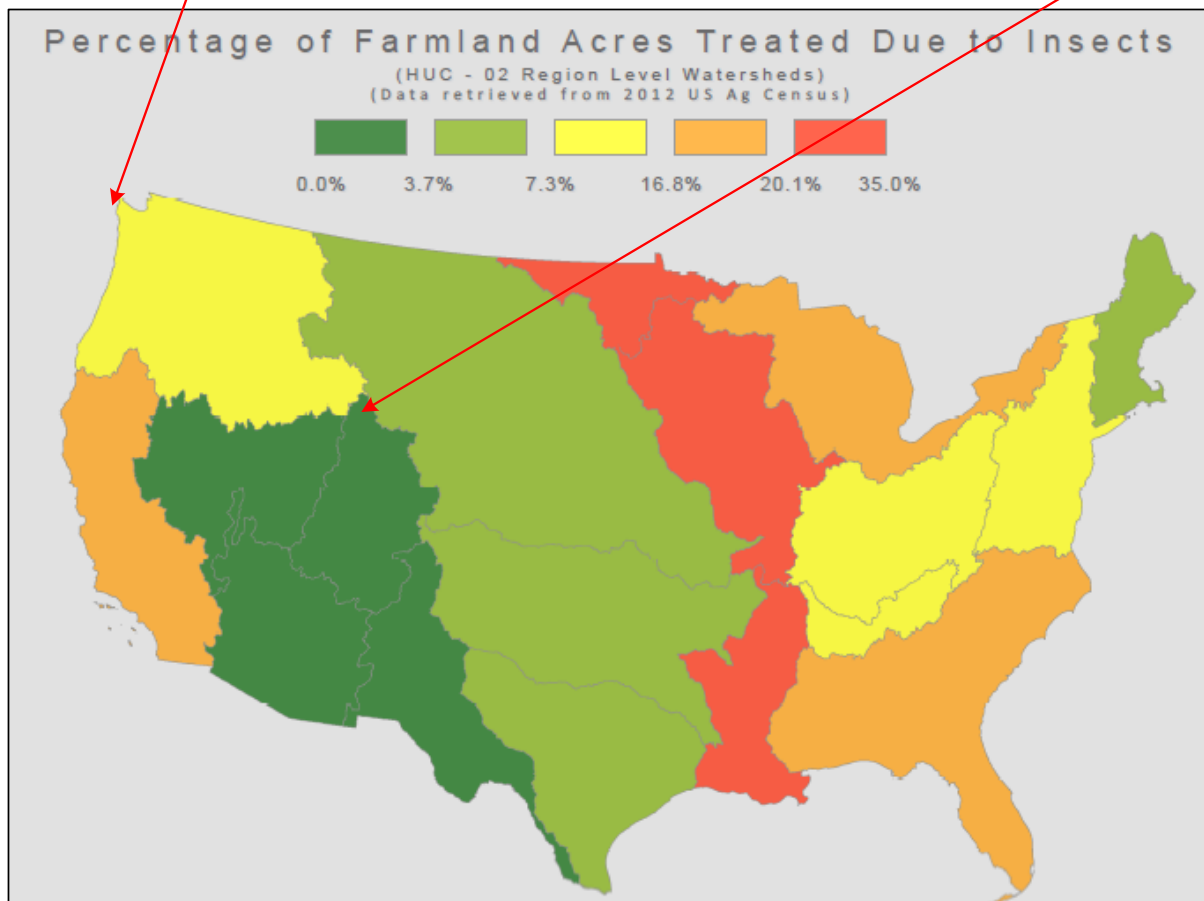
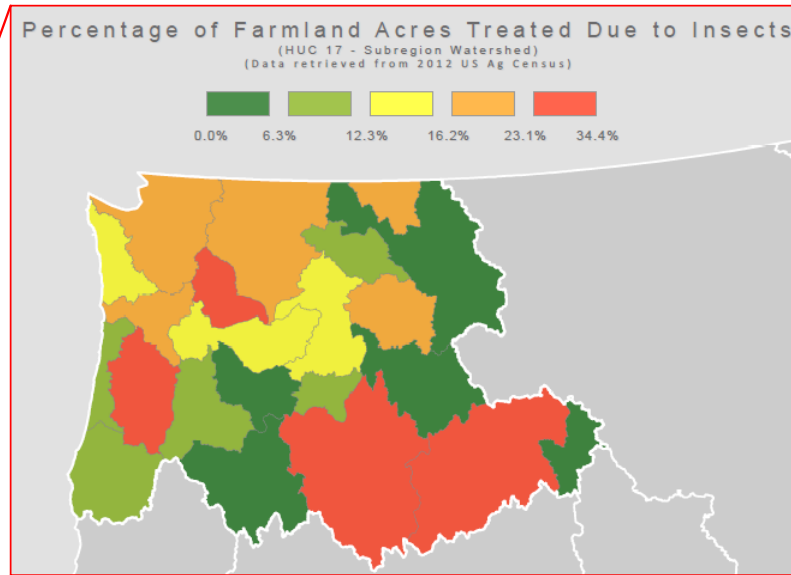
USDA survey data reveal that the average number of applications of pesticides is generally below the permitted maximum. For example, in recent years the average number of applications of Malathion on winter wheat was 1.1 with 2 permitted, on alfalfa 1.4 actual and 6 permitted, on strawberries 2.5 actual and 4 permitted, and on oranges 1.3 actual and 3 permitted.

Applying EPA's usage assumptions to all permitted crops at maximum rates at maximum number of applications gives total A.I. numbers that are orders of magnitude higher than actual use. For example, Malathion usage under the EPA assumptions would total 1,042 times actual agricultural use of it in recent years.

EPA recently used publicly available data to profile usage of individual pesticides, including chlorpyrifos, diazinon, malathion. Yet, they have not incorporated this data into ESA risk evaluation. By assuming that a given compound is used on every acre of crop, every year, for the maximum number of applications permitted on the label they have placed an elephant on the "weight of evidence" scales.



Figure 8. Percentage of Farmland Acres Treated Due to Insects According to 2012 Census of Agriculture



Representing the Crop Protection Industry



8.0 NON-TECHNICAL COMMENTS OF A BROADER REGULATORY NATURE CONCERNING THE REGISTRATION REVIEW PROGRAM OR THE ENDANGERED SPECIES PROGRAM

In its Instruction letter, EPA notes, “The interim scientific methods used in these draft BEs were developed collaboratively with the U.S. Fish and Wildlife Service (USFWS) and the National Marine Fisheries Service (NMFS), hereafter referred to as the Services, based on recommendations from the April 2013 National Academy of Sciences.” EPA goes on to state, “As part of this effort, the U.S. Department of Agriculture has provided information on crop production and pesticide uses.” The Panel released its report to the Agencies in April 2013 and to the public in May 2013. In August of 2012, OPP called for public comment on Registration Review and the ESA consultation process,⁵³ an exercise that resulted in a final guidance document issued on March 19, 2013⁵⁴, just a little over a month before the NAS Panel report was issued. In light of the developments since the publication of the Panel report, commitments made in this document are very important to revisit when addressing what is needed in policy development. A few points that EPA made that have either not been followed up on, or that are yet to be addressed are:

- A greater role for USDA
- Earlier involvement of stakeholders in the review process, to include
 - Use patterns defined by the registrant in advance of risk assessment by the Agencies
 - Clarifying product use and use practices through additional meetings with states, USDA and commodity groups
 - Consideration of early risk reduction, including key efficacious rates, existing conservation practices
- Consideration of pesticide use and usage data, solicited from USDA and grower groups
- Increased use of the informal consultation process

OPP notes that these changes and the involvement of growers and state lead Agencies “*will also ensure that the protection measures are economically and technologically feasible.*”⁵⁵ In discussing consultation, OPP notes, “*in developing draft RPAs and RPMs, the Services should include only those risk reduction measures that EPA has the authority to impose and should then defer to EPA to implement these measures using its existing statutory authorities. For example, the Services might identify a level of exposure below which jeopardy would not occur. In response, EPA will implement changes to the pesticide registration that will reasonably ensure that the specified level of exposure is not exceeded.*”⁵⁶

⁵³ FR 77: 49792 (August 17, 2012)

⁵⁴ Enhancing Stakeholder Input in the Pesticide Registration Review and ESA Consultation Processes and Development of Economically and Technologically Feasible Reasonable and Prudent Alternatives. EPA (March 19, 2013)

⁵⁵ *Ibid*, page 8.

⁵⁶ *Ibid*, page 11.



For efficiency and regulatory compatibility, an agreement on negligible risk for pesticide assessment could define both “acceptable risk” under FIFRA and “no jeopardy” under ESA. Thence came the “NAS Panel report,” Assessing Risks to Endangered and Threatened Species from Pesticides.⁵⁷ During the interim since the issuance of this report, CLA has been dealing mainly with the science of ecological risk assessment in response to the findings of the Panel and the developments at the Agencies in light of it. That is a correct response, because the Panel was instructed, and repeatedly reminded report readers, that their mission was “science only” and not to address policy. However, in absence of addressing policy needs in response to the Panel report, by default the report becomes a policy guidance document – a role for which it was not intended nor for which it is fit. That has become quite clear in the 3 years of operations under OPP’s “Interim Approach”⁵⁸ to Registration Review.

9.0 DOCUMENT ERRORS AND TECHNICAL CORRECTIONS

It appears that EPA and the Services worked together to attempt to improve risk assessment efficiencies through automation and aggregation of resources. However, this relegated the biological evaluation to a database exercise without the usual thoughtful and meaningful application of scientific process and a subsequent evaluation in a tiered framework. The biological evaluation included an impressive amount of effects data and species data. However, this detailed level of information and species scale resolution is lost when endpoints across species, within a taxonomic group, assigned as mean and worst case values regardless of species attributes or sensitivity. The attempted efficiencies instead cause the biological evaluations to suffer scientifically by merging species endpoints, focusing only on worst case exposure values across use patterns, not sufficiently conducting spatial analysis, and only focusing on an individual scale views. To use the data and resources effectively, species must be well represented and specifically addressed so that there is not a subsequent high “failure rate” that defeats the intent of the NAS Panel’s tiered risk assessment process.

The biological evaluations need to improve in their establishment of a relationship between effects data and relevance to individuals and species at representative environmental exposure concentrations. It is not established in the biological evaluations how thresholds used in the assessment translate into a likely or not likely effects determination (i.e. ecological consequence is not established). Units are often missing in tables and graphs in the effects section. Additionally, the biological evaluations do not appreciate the wealth of data on many species that is available for the OPs. Rather than rely on effects data for the best taxonomic surrogate for the listed species, the BEs rely on the lowest effect value for a taxonomic group. This increases the uncertainty and reduces the value of the assessment.

⁵⁷ NRC NAS Committee on Ecological Risk Assessment under FIFRA and ESA, “Assessing Risks to Endangered and Threatened Species from Pesticides

⁵⁸ Interim approaches for National-Level Pesticide Endangered Species Act Assessments Based on the Recommendations of the National Academy of Sciences April 2013 Report. EPA, November 15, 2013.



CLA's chapter-by-chapter comments in this section have iterated needed biological evaluation improvement recommendations, such as:

- Use taxonomic association between the listed species and laboratory species for appropriate effects thresholds, organize listed species by habitat type and apply appropriate exposure analysis based on these habitats, apply probabilities to risk analysis to get a better view of risk, consider effects endpoints against the frequency, magnitude and duration of exposure potential. This range of information provides a good relative range of risk. Otherwise all combinations are lumped together and a poor risk analysis is the result.
- The biological evaluation should exclude inflammatory and irrelevant discussions regarding illegal (or unknown) product applications relative to incident reports. Further, only the association between current use patterns and confirmed incident reports should be considered.
- The Interim Approach document and CLA concur on sublethal effects: they are relevant only when associated with growth, survival and reproduction. The BE spends a considerable amount of time collecting such information, arraying the range of information, and then assuming that the lowest effects values for these measurements translates into whole organism and species effects, with little explanation or justification of how this is related to growth and survival of the species.

Overall, the biological evaluations lack consideration for causal linkage between effects, exposures and risk to individuals or species as a whole. This lack of analysis has resulted in a process that is highly ineffectual in reaching what could be interim no effect/NLAA decisions. This is especially concerning as the analysis now, presumably, moves to the Services with little progress or certainty on how the remaining analysis can be conducted effectively. The conclusion that a pesticide is likely to impact all taxonomic groups and nearly all species within those groups defeats the value of the tiered approach that the NAS Panel recommended, and ultimately provides a disservice by putting undue burden on the Services and communicating to stakeholders an unrealistic and incomplete view of risk. Chapter by chapter examples of specific observations follow. These are complete because time and the nature of the documents did not allow a thorough review and findings reported by EPA could not be tracked back to specific data and references in many cases. We provide these comments in addition to those that the individual registrants may observe and report.

9.1 Chapter 1 and its Attachments and Appendices

9.1.1 Chapter 1 Body

- Section 1.2.1 and Fig. 1-2: AOP shows potential pathway from AChE inhibition to individual mortality, but is entirely qualitative. Has there been any quantitative analysis to relate AChE inhibition to assessment endpoints?



- Section 1.2.2, p. 23, paragraph on bioconcentration: based on log Kow, possible bioconcentration; but rapidly metabolized, so bioconcentration limited; how will this inconsistency be accounted for in the weight of evidence?
- Fig 1-3, p. 25: “Dotted lines indicate exposure pathways that have a low likelihood of contributing to ecological risk.” Exposure by ingestion of contaminated food is one such exposure pathway. But effects on aquatic plants leading to “reduction in algae and aquatic plants” and “Reduction in primary productivity” would also seem to have a low likelihood of contributing to risk, and should be shown with dotted lines. (Same comment applies to terrestrial conceptual model in Fig 1-4, p. 26.
- Risk hypotheses “link directly to the protection goals of Step 2” (p. 27). These are phrased qualitatively: “reduces the fitness of an individual” based on direct or indirect effects, or “adversely impacts PCEs” for designed critical habitat. Quantitative criteria are needed: reduced fitness by how much? With what likelihood? How frequently? Where?
- Fig 1-5, p. 28: The 3-step process is a logical decision tree but the figure doesn’t indicate the criteria for “yes” and “no” findings at each step. P. 29: Step 1 is the process leading to a No Effect/May Affect decision. There are several pathways to a No Effect finding: e.g. no county-level co-occurrence, no action area co-occurrence, not toxic on or beyond treated field, habitat preferences preclude exposure, species is already subject to a protection program. It is unfortunate that the BE takes only one of these pathways – action area co-occurrence – as the sole indicator of No Effect.
- The BE says (p. 29) “Step 1 does not identify the degree of effect that would be anticipated (e.g., an insignificant effect), or whether the risk of exposure of [sic] adverse effects is unlikely to occur (e.g., a discountable effect), where the action area overlaps with listed resources.” But this is not in fact the case. The degree of effect is indeed identified: a 1-in-a-million likelihood of individual mortality, or no observable chronic effect. And the risk of exposure is presumed, based on EFED modeling practices: 1 year in 10 (or 15) in a worst-1-in-10 watershed under worst case assumptions (e.g. percent treated area), though the assumptions are mostly implicit and are rarely mentioned. Without such comparisons of assumed effect and exposure, how can the off-site extent of the action area be determined? And if the assumed effect and exposure metrics are available, why not use them directly as part of the Step 1 process (e.g., as described in the Overview Document)?
- What are the criteria for effect magnitude and exposure probability that would result in a finding of no co-occurrence with the action area? And for species that move to Step 2, what are the criteria for NLAA vs LAA?
- Section 1.4.1.1: “Information on effects thresholds and results from offsite transport models will be used to establish the spatial extent of the action area.” So effects thresholds and exposure modeling are involved in Step 1.
- Section 1.4.1.1.a, p. 31: Using 5-year aggregation from CDL, will move to 10-year as data become available. Are 5- to 10-year-old data applicable to current crop distributions?



- Section 1.4.1.1.a, p. 31: If NASS acreage exceeds CDL, expand CDL until it equals NASS. “This reduces land cover mapping errors by adjusting the extent of each category to the best available census values.” What is the evidence that NASS is more accurate than CDL? How much does this add to the action area, fractionally?
- Boundary of action area is based on lowest exposure concentration found to be toxic. What is the allowable exceedance frequency of this exposure concentration? What is the implicit exceedance frequency, given the model assumptions?
- Where is the Downstream Dilution Tool presented?
- “The 1-in-15-year EEC is the maximum annual, daily aquatic concentration, developed from the 30-year distribution of maximum annual daily environmental concentrations, which, on average, will be exceeded only once every 15 years. It is important to note that for any single 15-year period, the 1-in-15 year value may be exceeded more than once, or not at all.” It is also important to note that the EECs are based on conservative, multiplicative assumptions and are much higher than nearly all measured concentrations in the environment.
- The most sensitive taxon for spray drift, terrestrial invertebrates, was the basis for the spray drift action area for all taxa (p. 35). The spray drift action area therefore overestimated the area where other, less sensitive taxa would be at risk.
- p. 36, last paragraph: “This process is iterative [i.e., if additional information becomes available during the course of conducting Step 2 that indicates that the action area is not adequate for the action(s) being assessed, the action area can be revisited].” If such additional information indicated that the action area had been overestimated, it is possible that a species finding would need to be changed to No Effect (no co-occurrence with the action area) in Step 2.
- Section 1.4.2, p. 37: “Exposure values are based primarily on fate and transport model results that assess the range of labeled uses of the pesticide (rates, methods).” EECs from EFED are taken without a discussion of what they represent – i.e., high-end estimates under worst-case assumptions.
- Section 1.4.2.1, Table 1-7, Lines of evidence for risk hypotheses: The phrases in the column headed “Line of Evidence” are actually assessment endpoints, not lines of evidence, and are more than a matter of terminology. “Lines of evidence” and assessment endpoints are two related but very different things. There is no indication in this table about criteria for weighing lines of evidence, or criteria for making decisions based on overall weight of evidence. (Repeated on p. 45-46)
- Section 1.4.2.2, WoE: Approach begins with risk hypotheses “that link directly to the protection goals of Step 2.” Nothing about the protection goals of Step 2 appears to be in this chapter, nor in Attachment 1-9. (See separate comments on Attachment 1-9.) The elements of the WoE are plausible risk hypotheses (A/B/C as presented on p. 45-46) and the criteria for expressing “confidence” as “relevance” and “robustness” are reasonable in principle. The key to the success of the WoE will be setting appropriate criteria for relevance and robustness such that they discriminate between species at greatest risk (LAA) and species at less risk (NLAA) or no risk (No Effect). If the criteria in Table 1-8 are set in such a way that the great majority of species fall



into the highest risk category, the criteria are in need of adjustment because they fail to adequately discriminate the degree of risk to different species. What is needed here are meaningful criteria and a clearly defined way to weight them.

- Section 1.4.2.2.a.1, p. 49: What is the scientific basis for the 1-in-15 year EECs? What is the scientific basis for this exposure criterion? How many receiving water sites are presumed to receive this amount of exposure with this frequency, and how many are acceptable?
- Section 1.4.2.2.a.1, p. 50: “Since general monitoring studies are not coordinated with specific applications of pesticides, it is unlikely that they capture the maximum concentrations that exist in the environment.” There is no logical connection between coordinating monitoring with specific applications, and capturing the maximum concentrations. Given sufficient numbers of samples, there is a finite probability of including some that are near the upper end of the distribution, and this probability is increased by monitoring programs that target locations and times where pesticide concentrations are likely to be greatest.
- Section 1.4.2.2.a.1, p. 51: If monitoring data exceed effects thresholds, confidence is increased that the chemical poses a risk. But the opposite should also be the case: if a substantial amount of monitoring data, representing a wide range of potential exposure conditions including high-exposure situations, indicate that concentrations are far below the model predictions (EECs), confidence should be increased that the chemical poses no risk (or confidence in the exposure estimate should be decreased). It is true, as stated at the end of this section, that “risk cannot necessarily be precluded,” but that is not the appropriate criterion in this case – not that risk can ever be precluded, but that it can be kept within acceptable levels.
- Section 1.4.2.2.b.1, p. 53: Use of open literature data is appropriate if properly screened for scientific method and completeness of documentation. Studies that fail objective evaluation criteria should be given little or no weight in the WoE process. Since many open literature studies use non-standard test designs, relevance is also an issue for consideration before incorporating data into the risk assessment. Therefore greater transparency is needed with regard to data reliability and relevance of the data used in the assessment (see section 10.8).
- Section 1.4.2.1.b.1.2, p. 53: The threshold for indirect effects is 10% effect (EC10 or LC10) for the 5th percentile species (or the most sensitive species). This is an extremely conservative threshold except for listed species with obligate dependencies on other species. Most predators are unlikely to be affected by a 10% reduction in 5% of their prey species, and habitat is unlikely to be affected by a 10% reduction in 5% of the plant species.
- Section 1.4.2.1.b.1.2, p. 54: The threshold for sublethal effects is the lowest available NOAEC that can be “quantitatively or qualitatively linked” to survival and reproduction. A plausible AOP can be constructed for AChE inhibition, or many other sublethal effects, but these do not generally account for the degree of actual linkage to the assessment endpoints of growth, survival and reproduction. Thus this threshold essentially becomes the lowest concentration causing any measurable effect on any parameter in any study. As such, this line of evidence should be rated low for relevance.



Comments on Attachment 1-9, Weight of Evidence Approach

p. 2: The “lines of evidence” (A1 through A5, etc.) are actually risk hypotheses or assessment endpoints, not lines of evidence. The individual lines of evidence are more or less the rows in Table 1-7 of Chapter 1. Each line of evidence can be evaluated for relevance/robustness of EECs and effect thresholds.

p. 3: Under exposure relevance, a bullet should be added to address confidence in model accuracy and realism, and the need to quantify the inherent bias resulting from conservative assumptions. Based on the two bullets now under exposure relevance, essentially any SWC output is considered High Relevance by default, regardless of the implicit and explicit assumptions. Likewise, “robustness” would be high for any compound with a complete set of efate data; very few compounds would have field-scale monitoring data so this robustness criterion is unlikely to come into play.

p. 4: AOP and other approaches to linking measured effects to assessment endpoints are more relevant when quantitative linkages are known; this implies that such approaches are less relevant when quantitative linkages not known, and should not be rated high for relevance in that case.

9.1.2 Chapter 1 Attachments

Attachment 1-12:

Attachment 12 references ‘Supplemental Information 3’ as an excel file “Compiled Aquatic Invertebrates.xlsx” which appears to have the information that led to the assignment of bins. This file does not seem to be available with the other documentation. An attachment 1-12 Supplement “Supplemental Information 1: Federally Listed Aquatic Invertebrate Database (XLSX) is provided which we assume to be the same document and the reference needs to be corrected.

The background information included batch run control files, for example “1. Chlorpyrifos Aquatic Modeling Batch Input File (ES)111615.csv” and they were sufficient to replicate USEPA output. The xlsx version of the inputs were not included in documentation. Based on the example provided on the provisional models web page, the xlsx version contains additional functionality for creating run rows and it would be desirable to have the XLSX version as well as the CSV:

- Aquatic bin information (field size, waterbody area, depths, etc.) are auto-populated with a lookup function from a tables found in a hidden tab.
- Guidance for choosing inputs is provided in row 1
- A read me tab is provided with directions for creating input sheets and running batch files in PWC.



Rice scenarios are not publically available (although available by request). Cranberry scenarios are not available publically and should be released thru typical EFED channels.

9.2 Chapter 2 and its Attachments and Appendices

2.1 Introduction to Fish and Aquatic-Phase Amphibian Toxicity:

It is stated that the EPA does not typically request toxicity studies for amphibians from pesticide registrants, but rather uses data on freshwater fish to represent potential effects to amphibians in the aquatic phase. If amphibian data is available and sufficiently relevant and reliable it should be used as surrogate information for listed amphibians.

Issues or uncertainties exist related to using surrogates to represent potential effects to individual listed species. However, a penalty should not be imposed for having effects data of increased relevance. If data is available for a species more taxonomic similar than other laboratory species tested then this data should be used if of sufficient quality.

2.2 Threshold Values for Fish and Aquatic-Phase Amphibians

This section discusses using separate thresholds from studies conducted with the technical grade and formulated product when available. Toxicity values are used from the lowest toxicity threshold if they are from studies with the formulated product. What is missing is a discussion of the relevance of the formulated products when environmental fate and dissipation are considered.

Regarding indirect effects thresholds for listed species that consume fish and aquatic-phase amphibians, there is no practical relationship stated or justified between the thresholds used and their actual relationship on indirect effects. Indirect effects thresholds should be stated.

Summary arrays of effects data are grouped by the type of effect (*e.g.*, behavior, reproduction, mortality), and present the range of values for each effect type. The square symbols represent mean endpoint values and the bars represent the data range. The decision to combine data in this way results in losing the detail and resolution needed to advance the risk assessment in a pragmatic way. By combining data (all fish for example) into such arrays the ability to associate the data between appropriate surrogate species (lab to listed species) is lost. The mean values from all studies are assumed to be more relevant otherwise. Further, the relevance of each endpoint group is lost when all fish species are combined together in this way.

2.3 Summary Data Arrays for Fish and Aquatic-Phase Amphibians

Inter-laboratory variability in effects data can result in a range of responses for the same species tested. These difference can be accounted for by using a geometric mean of the values from the distribution.



The toxicity data for each taxon are generally presented as summary data arrays developed using the Data Array Builder v.1.0. This array does little more than illustrate the range of effects responses. It would be appropriate to use this range in sensitivities to characterize the range of risk across listed species of interest and not assume worst-case values are representative (e.g. class, order, family) associations.

Data in these arrays are grouped by the type of effect (e.g., behavior, reproduction, mortality), and present the range of LOAECs and NOAECs (NOAECs must have a corresponding LOAEC to be represented in array) for each effect type. The relevance of behavior responses are not related to whole organism responses and thus species and population level responses. Further the distribution of effects is necessary to later understand the impacts at the population level. It is not clear how this detail will be appropriately translated for use in STEP 3.

There is much discussion regarding impacts of effects such as AChE inhibition, effects to sensory systems, and impacts on behavior. This include references to BCM=Biochemical, CEL=Cellular, PHY=Physiological, BEH=Behavioral and POP=Population responses. The linkage of this to survival and reproduction is not established and it is unclear how population responses could be determined based on this information or how these observations can be related to species populations. This appears to be mostly a database exercise without a critical technical review of the effects characterization.

2.3, paragraph 1: Every data point from every study is presented in the data array summaries. How is this information useful? How does inclusion of all this data alter the conclusion of the mean toxicity value for each effect?

2.4 Lines of Evidence for Fish and Aquatic-Phase Amphibians

Acute mortality (LC₅₀) data for chlorpyrifos are available for 40 fish species and 8 amphibian species based on studies submitted by the registrant or identified in the ECOTOX database. However, the value of this data is lost when these studies are lumped together because the understanding of sensitivity ranges within a taxonomic group is not possible. Therefore, the compiled data do not allow for appropriate association between relevant mortality data and a given listed species. The ultimate loss is the level of resolution needed for risk characterization and, if necessary, risk mitigations that may be needed for a species of concern.

Species effects data and endpoints from studies ranging in exposure durations were considered (e.g. 96 hours up to 10 days) and were included but it is not clear how the differing exposure durations may be translated into comparisons to exposure modeling and risk analysis.

The mortality thresholds for Freshwater and Estuarine/Marine Fish (E/M) fish are based on the 1-in-a million of the HC₀₅ based on the Species Sensitivity Distribution (SSD). Again, compiling data into an SSD and calculated HC₀₅ causes the resolution of the species to species level comparison to be lost. This is



especially problematic when there is a wide range of sensitivities across taxonomic groups. In effect, what the SSD would allow you to do is assign species to those data on the distribution and using exposure it would be a way to group species for further analysis but not to stop fully.

Using the lowest value from the distribution for indirect effects builds unrealistic conservatism into the assessment. The assumption here is the amphibian prey base is uniformly sensitive at lowest concentrations resulting in a much more significant impact to prey base. This conservative view should be corrected for (or captured) in the risk characterization.

Data arrays in some cases focus on the sensitive end of the toxicity spectrum. Some of these data may be relevant to some listed species but the selection for the most sensitive end of the distribution for all listed species does not make scientific sense.

Section 2.4.1, Amphibians: How are these studies weighted, if at all? It's difficult to place much confidence in 2 LC50 values where the variance is larger than the LC50 itself (LC50 = 121.87 µg/L ± 346.68; LC50 = 205.24 µg/L ± 543.75). This is likely due to the large spacing in the dose concentrations (50, 500, 5000 µg/L), so it's not clear why these studies are considered the most reliable data.

2.4.2.2 Effects on Reproduction of Fish and Aquatic-Phase Amphibians

Similar to other sections, a data array is presented with endpoints and ranges. Species sensitivities are illustrated but not effectively used in the risk analysis. This makes sense from a "may effect" analysis stand point at the taxonomic level but not in a "likely to adversely affect" sub taxonomic scale analysis.

2.4.2.3 Effects on Behavior of Fish and Aquatic-Phase Amphibians

Lowest toxicity values for behavior impacts are presented with no reference to how these translate to actual whole organisms (individual effects). Even if these were translated into whole organism (then species to populations) they would not be expected to be representative to all species. A relationship between behavioral effects across a range of exposure values varied with time is also not established. The Interim Approach states that relevance of sublethal effects (e.g. behavior) is limited to whether or not they can specifically be linked to impacts on individual survival and reproduction. Not making an association between sublethal effects and impacts on growth, survival and reproduction clearly does allow for the assessment of these sublethal effects later to the population level.

Section 2.4.2.3: A behavioral LOEC of 625 µg/L is given for 'the number of movements in the Zebrafish.' What does this mean, and how is this interpreted as a behavioral effect? And how is this correlated to a whole organism effect?

Section 2.4, page 2-18



The chlorpyrifos BE bases hazard values and thresholds for estuarine/marine fish on values collected for freshwater fish, even when marine fish toxicity values are available.

Section 2, page 2-27 and elsewhere

As acknowledged in the chlorpyrifos BE (page 2.51), aquatic phase amphibians may be less sensitive than fish. Despite this, fish hazard values are used when more sensitive, even when amphibian values are available. This is inappropriate for risk assessment and we recommend that the BE be modified to use relevant hazard data for amphibians.

Section 2, page 2-35

The behavioral endpoints identified, *e.g.* spontaneous swimming and feeding behaviors, are of questionable value for risk assessment. The selection of these inappropriate values is propagated to the marine environment as there are no similar studies with E/M species. We recommend that ESA assessments be based on conventional endpoints at screening-level Steps 1 and 2. Step 3 should incorporate higher-tier methods including probabilistic estimates and population models.

Section 2, page 2-38

Cholinesterase inhibition is a highly sensitive biomarker of exposure and is the mode of action of the OPs. However, ChE inhibition should not be used as an effect measure in the place of conventional endpoints, particularly at Steps 1 and 2.

Section 2.5, page 2-46

The BE effects table eliminates some studies that are not reported in units other than $\mu\text{g/L}$, mg/L , *etc.* It appears that some of these could be converted to $\mu\text{g/L}$ in order to expand the database for use in risk assessment.

2.6 Concentrations Where No Effects Were Observed in Fish and Aquatic-Phase Amphibian Studies

This section illustrates a wide range of values across species where effects were not noted in studies with constant exposures (unlike field conditions). However, the conservative nature of these effects data are not adequately used and expressed in the risk assessment towards discussion of risk probability.

2.7 Incident Reports for Fish and Aquatic-Phase Amphibians

This section as a whole is highly speculative, misleading, inflammatory and without adequate context. The section mentions that some incidents were attributed to compound uses but highlights clear cases



of illegal or undetermined uses. This is inappropriate in that illegal uses are not supported by the label and therefore not appropriate for the risk characterization. The section does not provide context as to the relevance of the uses to current labeled uses.

Section 2.7, page 2-49

In reviewing the incident database, the BE concludes that “exposure pathways for chlorpyrifos are complete and that exposure levels are sufficient to result in field-observable effects.” In fact, the reporting of 110 incidents over a 35-year period (1974 – 2009) represents a very small proportion of the applications of chlorpyrifos that have been made. All of the specific use examples described in the BE were of uncertain legality and/or could not be conclusively attributed to chlorpyrifos. The low number of incidents, especially in recent years, brings into question how approved uses of chlorpyrifos are related to the dramatic environmental effects portrayed by the BEs.

Tables 2-1, 2-2 and throughout

The data selected for use, often including those identified as the most sensitive species, may be of questionable value and acceptability. In addition, as many of the endpoints are not from standardized studies, the hazard values are not consistent and are generally not correlated to duration of environmental exposures. We recommend that the EPA and Services use standardized GLP studies in the identification of most sensitive species. For other consolidation purposes (*e.g.* SSDs) in higher tiers of risk assessment, additional studies of confirmed scientific rigor may be helpful.

Table 2-3: How are these studies weighted? Are guideline studies given more weight than ECOTOX studies?

3.0 Effects Characterization for Aquatic Invertebrates

Section 3, page 2-52 and throughout - The BE develops thresholds around the most sensitive sublethal endpoints for invertebrates irrespective of the demonstrated relevance of the endpoints for risk assessment (*e.g.* AChE inhibition in *H. azteca*). Sublethal thresholds should be based on growth, survival and reproduction endpoints that can be used in Steps 2 and 3.

3.1 Introduction to Aquatic Invertebrate Toxicity

Much like other effects sections it is stated that the data is used for weight-of-evidence approach. However, this approach is no more than completing data arrays which illustrate the range of sensitivities but then only calculates or pulls from the lowest end of the spectrum. It appears that the weight-of-evidence approach is only showing the range of potential effects from various endpoints.

3.2 Threshold Values for Aquatic Invertebrates



Lethal thresholds are derived from SSDs of survival from aquatic invertebrate acute toxicity studies, while sublethal thresholds are based on the most sensitive sublethal effects identified among registrant-submitted studies and open literature in the ECOTOX database. This section states that the lowest effects values were used across studies to calculate thresholds and construct SSDs, without respect to the relevance and reliability of the endpoint, the underlying quality of the study, or the relevance of the study to the species of concern. A weight of evidence approach and adequate effects characterization of species sensitivity is lost with this approach.

3.3 Summary Data Arrays for Aquatic Invertebrates

The presented data arrays contain a wide range of endpoint and ECx values. The section does not include a discussion of statistical quality of these extrapolated values (especially those at the low end or beyond the data). This evaluation and associated use of these data are paramount regarding the relevance and reliability of the data and thus the assessment as a whole.

The data arrays do not provide an association between the actual data use and what is plotted. Units are often missing from figures and tables.

3.4.1 Effects on Mortality of Aquatic Invertebrates

The mortality thresholds for freshwater and E/M aquatic invertebrates are based on the 1-in a million of the HC05 based on the SSD. This approach is highly conservative and inappropriate as it does not effectively characterize the sensitivity of many of the species for which data is provided. A threshold calculation for sensitive species should only represent a threshold for species more closely associated to them taxonomically at the screening scale.

Similar to fish, there is an extensive range in species sensitivity across invertebrates. This range is illustrated but not effectively used in the risk characterization.

Section 3.4.1, Mollusks: Shell deposition studies are designed to monitor for shell growth. EC50 values derived from these studies are effects thresholds, and are not intended to be used as a proxy for mortality. Also, it was mentioned that this study did not meet validity criteria for control shell growth.

Section 3.4.2 and throughout

The chlorpyrifos BE incorporates isolated results from qualitative studies into the analysis and weight-of-evidence assessments, without consideration of their relevance or reliability. The conclusions of the BE should be based on high-quality, quantitative studies that are reliable and relevant to the assessment and species of interest.



Section 3.4.2.1: It's interesting that data based on mosquito testing is included, given that chlorpyrifos is used as a mosquito adulticide. Does inclusion of this data skew the overall species sensitivity analysis?

Section 3.4.2.2: A *Daphnia* reproduction LOEC based on reduced offspring in the 1st brood doesn't mean much. Neonate production in the 1st brood in cladocerans is variable, therefore a longer-term study is recommended to allow for 3-4 total broods. Based on what is summarized in this paragraph, there were no differences in reproduction in subsequent broods. This is a good example of why all data requires a critical technical review and understanding of life history patterns before it can be used quantitatively in the BE.

Section 3.4.2.2: The third paragraph in the chlorpyrifos document discusses a marine 35d early life stage toxicity study, but never mentions the species.

3.8 Summary of Effects to Aquatic Invertebrates

It is stated that chlorpyrifos is highly toxic to aquatic invertebrates with less sensitivity exhibited in the mollusk group. Again, there is a range of sensitivity and this statement and the overall analysis loses this resolution.

Figure 3-1: What does NOC stand for?

Table 3 -5: This table is very difficult to interpret and should be better portrayed.

Section 3, Page 4-3

EPA discusses species range geospatial files that were provided by USFWS and NMFS. Could these files please be made publicly available for comment by stakeholders, and so that entities outside of EPA and the Services can conduct the same type of assessment?

Section 2, Page 4-2 and throughout

EPA makes numerous highly conservative assumptions and decisions in arriving at step three as illustrated by the fact that 97% of species for chlorpyrifos are judged "Likely to Adversely Affect" even when the assessment is refined in step 2. The combinations of conservative assumptions and estimates used by EPA regarding potential effects and modeled exposures will result in theoretical scenarios that are far in excess of reasonable environmental exposure and risk. We recommend that EPA and the Services reconsider their ESA methodologies to incorporate additional tools to be able to arrive and environmental realistic and scientifically defensible conclusions around ESA.

Section 4, Page 4-5



The BEs state that an NLAA determination is made for those species and/or designated critical habitats for which the use site (including off-site transport) and range overlap at 1% or more after rounding for significant digits. This approach is very simple and expeditious and relies on little species information. However, (1) the scientific basis for the 1% overlap is unclear, appears to be arbitrary, and needs to be improved, and (2) exceeds the accuracy of the underlying data. This value should be adjusted by species, by the size of the action area, and the location, characteristics and area of the critical habitat. An approach involving such species-specific information would introduce a level of realism, and result in more scientifically-based decision at the Step 1/Step 2 stage. We request that such methods be incorporated into the ESA process at Step 1 and Step 2.

Section 4.4: In other sections of these comments, we have noted how no qualifying evaluation of literature was made. This specific example demonstrates that:

Threshold for aquatic vascular plants, chlorpyrifos: this was based on an open literature study “Removal of chlorpyrifos by water lettuce (*Pistia stratiotes* L.) and duckweed (*Lemna minor* L.)” published in the International Journal of Phytoremediation (study ID 155150, Prasertsup and Ariyakanon, 2011). Based on the information available in ECOTOX, and the information provided in Chapter 2 (page 96-97), this study does not meet many of the factors that “should be considered as important in determining the utility of an open literature study” (page 23 of the Open Literature Guidance), as follows:

Open Literature Guidance	Study ID 155150
Nature of the test substance – must indicate exact nature	Not clear if TGAI or formulated product, % a.i. not reported.
Species age, sex, size, life stage, and source reported	Chapter 2, Table 4-1 says the species are from Thailand. No other information given.
Number of organisms tested per concentration and number of concentrations tested	No information
Exposure method, route, frequency and duration	7 day duration, no other information. Analytical verification of exposure apparently not done
Suitable number of controls	A control is reported to be included, but no other information
Performance of test species was typical	No information
Observations (nature, time, of occurrence, severity, etc.)	Relative growth rate was based on wet weight and dry weight; frequency of measurement not provided.
Husbandry conditions	No information. Test apparently conducted in water, not growth medium, so plants may have been under sub-optimal conditions with respect to nutrients
Statistical methods	NOEC for population growth rate reported at



	level of significance <0.05. No further information.
Complete and accurate description of procedures and evaluation of results	ECOTOX record provides limited information. Some additional information in Chapter 2.
Important information missing	Yes
Other useful information: test chemical properties, water quality, negative and solvent control performance, endpoint selection	Temperature, pH, conductivity and dissolved oxygen are reported (single values only). Dissolved oxygen is not a relevant water quality parameter for aquatic plant toxicity studies.

It is noted that this was the only study providing quantitative endpoints for vascular aquatic plants and chlorpyrifos. Nevertheless, due to the weaknesses in this study, it should not be used to make conclusions of “high risk with medium confidence” to most of the listed aquatic plants.

Example 2: Threshold value for terrestrial plants, chlorpyrifos. The selected threshold for Step 2 for post-emergent exposure to monocots was based on an open literature study which reported effects on reduced photosynthetic rate and reduced stomatal conductance in corn at 0.999 lb a.i./A (Study ID E064451, Godfrey and Holtzer, 1992). It is not clear how these endpoints are translated into the assessment endpoint of growth and reproduction. Also, based on the information available in ECOTOX, and the information provided in Chapter 2 (pages 224-225), this study does not meet many of the factors that “should be considered as important in determining the utility of an open literature study” (page 23 of the Open Literature Guidance), as follows:

Open Literature Guidance	Study ID E064451
Nature of the test substance – must indicate exact nature	Lorsban 4E, % a.i. not reported.
Species age, sex, size, life stage, and source reported	<i>Zea mays</i> , Pioneer 3377, 58 days post planting
Number of organisms tested per concentration and number of concentrations tested	No information on replication. Only one treatment rate tested.
Exposure method, route, frequency and duration	Dosed by ground spray 2 times during a 14-day period. Study done in 2 different years. No analytical verification of exposure.
Suitable number of controls	No information
Performance of test species was typical	No information
Observations (nature, time, of occurrence, severity, etc.)	Physiologic parameters measured with a portable photosynthesis system at 2 or 3 time points after treatment.
Husbandry conditions	No information.



Open Literature Guidance	Study ID E064451
Nature of the test substance – must indicate exact nature	Lorsban 4E, % a.i. not reported.
Species age, sex, size, life stage, and source reported	<i>Zea mays</i> , Pioneer 3377, 58 days post planting
Number of organisms tested per concentration and number of concentrations tested	No information on replication. Only one treatment rate tested.
Statistical methods	LOAEC reported at level of significance <0.05, ANOVA (general linear models). No further information.
Complete and accurate description of procedures and evaluation of results	ECOTOX record provides limited information. Some additional information in Chapter 2.
Important information missing	Yes
Other useful information: test chemical properties, water quality, negative and solvent control performance, endpoint selection	“Silty clay loam soil” but no details; pH, organic matter, moisture, etc. not reported.

Given that this study is deficient in a number of the criteria, that it reports physiologic measurements not correlated to the standard measurements of growth (dry weight and shoot height), and that post-emergence data for monocots are available from standard guideline studies (MRID 49307201), this study is inappropriate for selection as a threshold. The guideline studies show that chlorpyrifos does not affect monocot vegetative vigor with IC25 values >5.7 lb. a.i./acre. Note that the full MRID citations are stated to be given in Appendix X, but Appendix X was not provided.

Other examples of studies that were used by EFED without providing detail on the robustness of the study:

- Malathion (aquatic plants)– Chapter 2, Page 109 – direct effect evaluating dissolved oxygen production and cell density in green algae; EFED indicated this study would be used to represent all aquatic plants. This effect is not related to effects related to growth, reproduction or survival.
- Chlorpyrifos (aquatic plant)–Chapter 2, page 89 – the most sensitive threshold value for “all aquatic plants,” for direct effects, is based on decreased photosynthesis in a nonstandard species of marine diatom. This effect is not related to effects related to growth, reproduction or survival.

9.3 Chapter 3 and its Appendices

Chapter 3, Section 3.4.2.2: The third paragraph discusses a marine 35d early life stage toxicity study, but never mentions the species.



Chapter 3, Figure 3-1: What does NOC stand for?

Chapter 3, Table 3 -5: This table is difficult to interpret and needs to be better organized and portrayed.

Chapter 3, Section 2.8: Aquatic Modeling Results, CPY

- Issue: EECs regularly exceed the solubility limits
 - BE Statement: “While EECs would not normally be expected to exceed the water solubility limit, variations in waterbody conditions (e.g., pH, temperature, turbidity, hardness) could be different than those used to determine the solubility, such that EECs could be above the water solubility limit; however, concentrations are not expected to be orders of magnitude higher than the solubility reported for laboratory studies.”
CLA Comment: This is overall an inappropriate justification for EECs being higher than laboratory solubility. Note that in case of CPY, the solubility values used were already on the very high end of laboratory results (see Giesy & Solomon for additional solubility values⁵⁹).
 - BE Statement: “EECs derived using the PRZM5/VVWM for Bin 3 (moderate flow aquatic bin) and Bin 4 (high flow aquatic bin) exceeded the solubility of chlorpyrifos by several orders of magnitude for some modeled use scenarios and Bin 4 EECs are greater than Bin 3 EECs which are greater than Bin 2 EECs.”
CLA Comment: This indicates towards a flaw in the overall flowing water modeling approach. A given screening level flowing modeling approach should be appropriate for a range of watershed and water body sizes. The fact that Bin 3 and Bin 4 EECs were determined to be unreasonable point towards the broader limitations and inappropriateness of the flowing water modeling approach that was followed.
 - BE Statement: “Moreover, the concentrations for these flowing bins are several orders of magnitude higher than that derived for the statics bins which have no outlet for the release of pesticide.”
CLA Comment: This should be true for Bin 2 as well. EECs in the small static habitat (Bin5) should be higher than Bin 2. This is not the case in the BEs, which is an indication of the unreliability of the flowing water EECs, including Bin 2.
 - BE Statement: “Limited data are available for edge of field concentrations and should not be used to characterize an upper bound exposure estimate.”
CLA Comment: While limited monitoring data is available for edge of field concentrations, modeled edge of field concentrations are available for every PRZM5 simulation that was run. These values provide an extremely conservative upper limit for every flowing water simulation that was run. These modeled edge of field dissolved concentrations were often 2 orders of magnitude lower than the Bin 2 peak EECs that ranged from 17 ppb to 69,900 ppb (for chlorpyrifos).

⁵⁹ Giesy J, Solomon K. 2014. Ecological Risk Assessment of Chlorpyrifos in Terrestrial and Aquatic Systems in the United States. New York, NY: Springer International Publishing.



- BE Statement: "Use of solubility is more appropriate for defining an upper bound exposure estimate."
CLA Response: The need to use solubility as an upper bound exposure estimate demonstrates the flaws in the modeling approach. Modeled edge of field concentrations (from a conservatively parameterized model) serve as better upper bound for a low flow water body.
- BE Statement: "Existing atrazine monitoring data sets are used to evaluate the relative magnitude in EECs as a chemical moves successively from Bin 2 to Bin 4".
CLA Comment: There is no data presented to support that the relationship between atrazine from small headwater streams to larger rivers would apply directly to the 3 OPs. Atrazine's environmental fate properties differ from the 3 OP, including solubility, adsorption to sediment, and degradation rates, all of which impact the fate of the pesticide over time in a flowing water system. For the monitoring datasets assessed (AEMP and Heidelberg University), atrazine is used on crops with very high PCA (corn/sorghum) and is very widely used on those crops (a high PTA). These differences in use pattern between atrazine in the watersheds evaluated and the 3 OPs make the transfer of conclusions on atrazine to the 3 OPs unsubstantiated.
- BE Statement: "As such, a qualitative approach is being considered where Bin 2 EECs are generated using the PRZM5/VVWM".
CLA Comment: An appropriate quantitative approach, that accounts for pesticide environmental fate properties, should be applied to arrive at EECs more relevant to specific pesticide chemistry. Furthermore, the use of erroneously high EECs for Bin 2 to derive Bin 3 and Bin 4 EECs further propagates this error to additional species that are found in these habitats.
- BE Statement: "Bin3 EECs are characterized as being conservatively 5 and 10 times lower than the Bin 2 EECs, and the Bin 4 EECs are characterized as being conservatively 5 and 10 times lower than the Bin 3 EECs."
CLA Comment: This statement is unclear as written. Based on review of the detailed model results in Appendix 3-4f, it appears that EECs for bin 3 are 5 times lower than Bin 2 and that EECs for Bin 4 are 10 times lower than Bin 2. This needs to be clarified.

Chapter 3, Section 2.9: Aquatic Modeling Sensitivity Analysis

- The aquatic exposure sensitivity analysis was only conducted for environmental fate parameters and application dates only. Given that the flowing water scenarios and modeling approaches are brand new, a sensitivity analysis that included the following parameters would have been more useful:
 - Water body dimensions
 - Water body flow rates within the range of the Bin
 - Watershed area
 - Flow-through options



Chapter 3, Section 2.13.1: Uncertainties in Aquatic Modeling and Monitoring Estimates, Surface Water Aquatic Modeling

- One source of uncertainty described for static water modeling was the effects of increase or decreases in pond levels. Significant increases in levels occurred for many of the static bin scenarios as a result of the very large watersheds. This source of uncertainty should be quantified.
- For flowing water bodies, changes in volume were only associated with surface runoff, and did not account for shallow subsurface flow (interflow) or groundwater contributions (baseflow). These components of streamflow are accounts for more than 50% of total stream flow in many regions of the country and should be accounted for.
- Multiple conservative assumptions for spray drift modeling were described, including the constant 10 mph wind speed, always blowing from the treated field to the receiving water body, no interception of drift on vegetation or other barriers, or BMPs followed by applicators. While these conservative assumptions are appropriate for Step 1 modeling, Step 2 modeling should incorporate more realistic assumptions through a probabilistic analysis of the likelihood of a range of spray drift conditions.
- EECs were shown to be very sensitive to application dates. The uncertainty in application dates should be accounted for in Step 2 modeling in order to achieve a more comprehensive probability distribution of annual maximum EECs to compare against species end points.

Chapter 3, Section 2.6.1: Urban Exposure Model (Chlorpyrifos)

- The neighborhood conceptual model:
 - A large trash storage area (a use site for CPY) equal to the size of a driveway is part of the conceptual model. Are there examples available of such large trash storage areas on these sized lots? The size seems like an unreasonably large area to devote to trash storage.
 - The fence in this scenario leads to a very large application area. It is unclear what the 6 ft represents, and it is unrealistic to have these fences surrounding every lot in a neighborhood. The applications to the wood fence are treated as impervious surface applications. There is no evidence that PRZM simulations using an impervious curve number value produce predictions of pesticide transport from wooden surface with an acceptable degree of accuracy. The application to the wooden fence is at a rate equivalent to 72.5 lb/acre, which seems extreme.
 - A utility easement is present on a portion of every house lot, which does not seem possible for an entire neighborhood of ¼ acre lots.
 - All the uses of CPY modeled result in effective treated area assumption of 100% of every house lot. There is a 0% likelihood of this assumption happening in reality. Additional data concerning urban outdoor pesticide use should be considered in order to estimate an upper bound on the percent treated area for the urban scenarios.

Chapter 3, Section 2.6.1: Urban Exposure Model (Malathion)



- The neighborhood conceptual model:
 - Patios and other treated areas are not to scale on map. 1,000 ft² would be the same size as the house footprint and nearly as large as the garden area. Given that this is a significant area, it should be better justified and represented to actual scale on the diagram to provide a better sense of the proportion of the lot these use sites occupy.
 - Dimensions of important features would be helpful
 - Total percent treated area with malathion is equivalent to 38.8% of the neighborhood. What is the likelihood that a residential watershed (of any size) would be 38.8% treated with malathion? Based on our understanding of malathion sales and residential use, this likelihood would be close to 0%. Is this an appropriate screening level scenario assumption?

Attachment 3-1, Background Document: Aquatic Exposure Estimation for Endangered Species, Section:
3.1.3.3.3 Estuarine and Marine

- EPA approximated pesticide exposure in offshore (bin 9) and tide pool (bin 8) marine environments using pesticide concentrations calculated for large stream/riverine-type habitats (bin 4, 7) and smaller stream/pond-type habitats (bin 2, 3, 5), respectively. Neither tidal pools nor offshore marine environments can be modeled appropriately with existing habitat scenarios which represent homogenous watershed areas draining into pond-like receiving waters that are either static or allow for steady inflow/overflow. In addition, the EPA standard scenarios defining weather, crop, and soil characteristics for each HUC2 are based on inland environmental conditions and therefore are not representative of coastal environments.
- Tidal pools by definition are filled with seawater by tidal flushing and wave activity during high wind/storm conditions. The ratio between tidal pool volume and watershed area is not determined by rainfall/runoff. Therefore the watershed sizes from the surrogate habitats, determined through regression of flow rate with drainage area (flowing bins) or precipitation/evaporation/runoff balance to maintain a specific normal capacity (static bins), are not relevant for modeling pesticide exposure in tidal pools and are not related to tidal pool capacity. Tide pools, which vary in volume and may be completely submerged during high tide, do not experience runoff loadings from a fixed-size watershed into a fixed-volume receiving body in the way that classical ponds and streams do. In addition to having inappropriate watershed areas for tide pools, the habitat surrogates have inappropriate flow regimes for tidal pools. The rate, magnitude, and frequency of tidal flushing cannot be approximated by the steady stream inflow/outflow rates used in the surrogates. Tides will have a leading order impact on dilution of any pesticide residues present in the pool via partial or total exchange of the tide pool water with seawater every 12 hours. This is fundamentally different and requires a different modeling approach from the continual, constant baseflow assumed for flowing habitat bins. The watershed areas and flow rates are critical parameters influencing pesticide loadings and dilution in receiving waters. The values of these parameters determined for the static and flowing habitat bins that were used as surrogates for tidal pool habitats are not representative



of tidal pools and require a different modeling approach. Both watershed area and diluting flow/exchange rates for tidal pools are not determined by the same factors used to define these values for the surrogates. In addition, for tidal pools these values are highly variable over the course of one day and these variations are not captured with the surrogates.

- Like tidal pools, the dimensions of offshore marine environments and their drainage areas are not determined by regional variations in precipitation/evaporation/runoff balances or typical stream baseflow rates. Offshore marine habitats are tidally influenced and therefore vary in volume due to twice daily mixing and flushing of the tides. Their volumes and the related dilution of any pesticide residues are determined more by their connection to bays, seas, and oceans (large reservoirs of residue-free seawater) than runoff from upland drainage areas. For offshore marine environments, the drainage areas contributing pesticide loadings will vary for each receiving water body based on local geography and elevation. There is not a generalized relationship between the volume of the marine environment and the contributing coastal watershed area. Specific offshore habitats of interest need to be modeled with spatially explicit characteristics by water models capable of simulating tidal effects.

9.4 Chapter 4 and its Appendices

This chapter of the BE is titled “Effects Determination,” although it is, in reality, the Risk Characterization section. It is understandable that such a title was selected if one is to assume the objective of the work is to determine if there is No Effect/May Effect (Step 1) or NLAA/LAA (Step 2); however, it creates confusion with Section 2 (Effects Characterization). Therefore, it is recommended that the section be renamed “Risk Characterization” to be in keeping with the Agency’s *Guidelines for Ecological Risk Assessments*. Note that section 8 of this chapter is already titled “Risk Analysis” so it makes sense to realign the name of the Section as well.

The Risk Characterization for Step 2 depends upon a series of highly conservative assumptions. Because probabilities are multiplicative, not additive, the final result is an extremely unlikely event.⁶⁰ It may not be necessary or prudent to be so conservative throughout as a highly unlikely event will result in a LAA determination. For example, use of a 1-in-15 years 4-day average for an estimated water concentration as an exposure value will result in much more LAA determinations than use of a 21-d, 60-d, or 90-d average (for 1-in-15 years). Because the estimation of these average concentrations includes a concatenation of multiple conservative input values, the probability that a listed species will experience such a concentration is very small. When this represents a sublethal effect, the probability that such an effect would occur after only 4 days of exposure is even less. And finally, the probability that a sufficient number of individuals experience such an exposure that actually

⁶⁰ Hope BK. 2012. Exposure gone “wild”: A call for rational exposure scenarios. HERA 18:485-487



results in decreased fitness to a level that could possibly affect the population growth is vanishingly small.

Step 1 No Effect/May Effect determinations are based on the overlap between the crop “action area” and the species range/habitat. While considerable information is provided about how the crop “action area” is determined, less information is available about species range/habitat and what is meant by “overlap.”

The term “habitat” is not described. Perhaps the ESA definitions of “habitat” and “critical habitat” can be incorporated into this document.

A definition of “overlap” is needed for Step 1 – Does the “action area” have to overlap all of the species range/habitat or just a portion of it? If a portion, how much? What if they are contiguous (i.e., share a boundary but there is no overlap)? Note that this is defined later in Step 2 as a minimum of 1% overlap, but there is no such definition provided for the initial, Step 1 analysis.

It is surprising that the degree of overlap between the “action area” and the range/habitat is not considered until Step 2 (Section 4 of Chapter 4). This should be part of Step 1.

There is no rationale presented for why 1% overlap was chosen as the minimum required overlap for inclusion as “May Effect”. The 1% value seems to be highly conservative, as most of the species do not utilize 100% of their range/habitat 100% of the time. An argument could be made for considering 5% as the *de minimus* amount as well. Please provide a rationale for why such a conservative value was chosen, and include this in the Step 1 analysis.

Attachment 1-6 provides the information about “Action Area” and “Overlap.” In the malathion and diazinon documents, this is an EXCEL File (same file in both documents); it has not been incorporated into the chlorpyrifos document although all the information is in the file.

The tab “RangeOverlap” is exactly the same in the malathion and diazinon files. Is this correct? If not, which chemical are these numbers referencing?

Why does the “Species percent overlap with each use site” on the “RangeOverlap” or CHOverlap” tabs not include offsite transport? Therefore, how is the “site” defined? Is the “action area” (= site + off-site transport) defined somewhere else (it does not appear to be available in an EXCEL format).

Section 4 of Chapter 4 says “see **ATTACHMENT 1-6** for results of the Step 1 and Step 2 analysis.” This is not an accurate statement as the overlap does not include the entire “action area.”



Since **ATTACHMENT 1-6** incorporates the RESULTS of the analysis (i.e., overlap of species range/habitat with chemical use areas), it should be in Section 4 not in Section 1 (Problem Formulation). These data provide the basis for the No Effect/May Effect determination (assuming that “site” and “action area” are the same).

While the **ATTACHMENT 1-6** EXCEL file is helpful in displaying the percent overlap between use area and range/habitat (although it should be “action area” not “use area” and incorporate the calculated off-site movement), presentation of the data via maps would be very helpful, particularly if landcover were included as a filter (e.g., one could then see how agricultural fields might be overlapping with salt marshes).

Chlorpyrifos is assumed to be used in 100% of the US and all its territories for mosquito control and “wide area use.” This seems like an oversimplification, particularly since “wide area use” is defined as “e.g., for ants and other misc. pests.” It calls into question whether one should really assume that 100% of the range/habitat of species such as the polar bear and the grizzly bear should be considered at risk, as substantial portions of their habitats are not associated with areas of intensive human use. Similarly, the short-tailed albatross only occasionally nests on Midway Island in areas protected for bird nesting, where mosquito control and “wide area use” are unlikely (major nesting colonies are outside the U.S., in Japan). Another example is the band-rumped storm petrel that is a pelagic forager and nests on steep cliffs on volcanic islands; these will not be subject to spraying for control of mosquitoes, ants, or other misc. pests. A Step 1 No Effect/May Effect analysis should encompass more in-depth thinking about the actual locations/ecology of each of the species.

A list of the species for which a “not toxic” conclusion can be made should be included. These species will obviously not need to be carried through the risk assessment so a “NLAA” determination can be made early in Step 2. This should be its own Section in Chapter 4. If the chemical is toxic to all species, then the Section will simply state that fact and move on. “Not toxic” is as defined by EPA (i.e., no effects at the upper testing limit).

For Step 2 (NLAA/LAA), it is very difficult to discern what effects values are being used. The species-specific summary tables in Appendix 4-3 state the types of endpoints that are available, but are not very clear about which one was used (e.g., NOAEC or LC20?) nor do they present the actual numeric value. The effects endpoint used in Step 2 should no longer be the ultra-conservative 1-in-a-million threshold, so an argument needs to be presented as to which level of effect will be acceptable. A spreadsheet showing the effects values used for each species in Step 1 and Step 2 would be very helpful, as would including the numeric values into each of the species-specific summary files in Appendix 4-3.



Similarly, there is no indication of what risk quotient is considered “high,” “medium,” or “low” so it is impossible to determine if there is consistency in assignment of these values across all lines of evidence for all species, routes of exposure, etc. How large of an effect is needed and how much of an exceedance of the EEC is required to place the risk in each of these risk categories?

Section 4 – this sentence is confusing: *“An effects determination of “Not Likely to Adversely Affect” is reached for 14 species that were “presumed extinct” based on information gathered in review of the 5 year status review but did not meet the additional criteria to receive a no effect determination.”* What are the “additional criteria” that need to be met to receive a “no effect” determination?

Section 4 – it is not clear why exposure is not discountable for Midway Islands, the pacific island coral atolls, and Mona Island of Puerto Rico, all of which have less than 150 people living on them. Midway Islands, in particular, are primarily protected wildlife refuges so assumption of exposure to any of these pesticides is not warranted, or the separate ESA analysis that is conducted through wildlife refuge management procedures should be considered.

The granular exposure analysis for Step 2 (**Appendix 4-6**) for diazinon uses a toxicity threshold for birds set at 1-in-a-million. This is highly conservative and should only be applied in Step 1, not here in Step 2.

Sublethal effects such as behavioral changes, cholinesterase depression, etc. are deemed “relevant to species fitness.” However, the degree of relevance can vary as there are thresholds of change below which the organism can adapt. These thresholds have not been identified in Section 2; rather, the value selected appears to be the lowest value where there was a significant difference from control values. This is a large uncertainty for most of the sublethal endpoints (although there are some studies for cholinesterase effects in fish that could be referenced to determine the appropriate behavior-associated threshold, for example) and should be included where determining the weight to put on the “relevancy” criterion in the WoE approach.

We are unable to find the location where the EEC is compared to the toxicity threshold value for each of the endpoints being considered (mortality, cholinesterase, behavior, etc.), so it is not clear how a “high,” “medium,” or “low” risk categorization is determined. It would be very helpful to have these values and the associated risk ratio presented in the species-specific tables in **Appendix 4-3**. A large amount of information is presented in Sections 2 and 3 from which these values are derived, but it is nearly impossible to correctly attribute the numbers from those tables to the rows in the species-specific tables in Appendix 4-3.

Example 1: Diazinon Appendix 4-3q. Plant: *Asplenium diellaciniatum* (10586), has a “high risk” determination. There is no way to understand why – perhaps because the LOAEC is exceeded for post-emergent or for “alternate rate” applications? But by how much is it exceeded – 2X, 10X,



100X? And how much of a growth reduction does this represent (10%, 50%)? What are the NOAEC and LOAEC values that were used?

Example 2: Chlorpyrifos Appendix 4-3aa) - "*Pritchardia remota*", a growth direct effect was identified as a "high risk", with "low confidence". While it is not clear in the biological evaluation chapters how a "low" confidence rating impacts the weight of a risk conclusion, the "high risk" conclusions are overly conservative based on multiple lines of evidence that should have been included as well as made transparent in the risk determination. This overly conservative approach was a contributing factor that resulted in nearly all listed species resulting in a "likely to adversely affect" call in each of these insecticides' assessments.

To refine this analysis, the levels of conservatism in each of these WoE documents should be defined. Using the *Pritchardia remota* example, the "high risk" direct effect to growth is determined based on a threshold value NOEC being exceeded for the minimum and maximum application rates at the site of a pre-emergent application, however, for post-emergence application, the LOEC value is not exceeded for the upper bound rate (4 lb a.i./A). Based on *Pritchardia remota* habitat being located in scattered, small groves in two valleys with elevations of 50 to 800 feet (information from WoE Appendix 4-3aa), actual in field or edge of field exposure would be rare at most. It is not evident that this WoE document included overlap in critical habitat from the CDLs, or the TED tool analysis for distance that risk extends from edge of field.

*For pre-emergent exposure at the site of application, the NOAEC is **not** exceeded for the minimum application rate of 0.5 lb a.i./A and the NOEC is **not** exceeded for the upper bound rate of 3 lb a.i./A. The LOAEC is not exceeded for the minimum application rate and the LOAEC is not exceeded for the upper bound rate....For post-emergent exposure, the NOAEC **is** exceeded for the minimum application rate of 0.5 lb a.i./A and the NOAEC **is** exceeded for the upper bound rate of 3 lb a.i./A. The LOAEC **is** exceeded for the minimum application rate and the LOAEC **is** exceeded for the upper bound rate....Considering alternate rate: etc with various scenarios either exceeding or not exceeding NOAEC and LOAEC values.*

The **sea turtle** analysis (Section 7.1) has multiple assumptions that result in an unrealistic conclusion of risk for these species and misses some key issues to consider when assessing exposure potential:

Exposure to bird eggs is used as a surrogate for exposure to turtle eggs, although there is no mention of the fact that birds have hard, calcified shells while turtle eggs are covered in a leathery membrane with no outer shell. Although it is known that bird eggshells are permeable, the degree of permeability likely differs significantly between bird and turtle eggs. This area of uncertainty should be acknowledged.

Extrapolation of dermal exposure estimates between birds and turtles should not be done. Bird skin is very thin without much fat in the dermis whereas turtle skin is more leathery and has greater



amounts of subcutaneous fat. The permeability of the skin likely differs significantly (although turtle skin may be similar to the skin on bird legs/feet). Additionally, the percent of the turtle's body where chemical may come in contact with skin is significantly different than that of birds.

Estimation of uptake of chemical into the near-shore diet significantly overestimates turtle exposures. There is a large uncertainty in the KABAM estimate of the BCF for uptake into plants and the result is a highly conservative (i.e., high) value.

The estimate of the amount of chemical likely to be in near-shore waters is also highly uncertain. This is acknowledged in the following paragraph (emphasis added):

It should be noted that these estimates of exposure are highly conservative, since they represent the highest 4-day and 21-day average concentrations estimated in a 15 year period. They do not necessarily represent concentrations that would be considered likely in estuary and near shore areas. In addition, there is a high degree of uncertainty related to the representativeness of bins 2, 3 and 5 for estuaries. It is questionable whether bin 2, which is a small water body has sufficient volume to hold an adult sea turtle.

Furthermore, as stated in the diazinon Chapter 4 (emphasis added):

The EECs for these bins reflect contributions from both runoff and spray drift from treated areas adjacent to the waterbodies, which may not typically occur for intertidal and subtidal nearshore waterbodies *which have beaches between the treated area and the leading edge of the water....*[and for bins 2 and 3] there is uncertainty in how well these bins reflect the *turbulent, mixing nature of the 12-hour tidal cycle* and the EECs that may be present. ... it is uncertain if these surrogate bins could hold sufficient volume to contain adult sea turtles."

It is important to note that juvenile **turtles** are not reared in near-shore tidal pools as stated in the document. The hatchlings move into the open ocean waters and juveniles float with the currents for 1+ years, returning to near-shore habitats as subadults.

In **Section 7**, the malathion and diazinon documents say to See Table 2-6.1 of Chapter 2 for the avian dietary thresholds for mortality and sublethal effects; the chlorpyrifos document says to see Table 2-5.2 in Chapter 2 for these values. These tables do not appear in Chapter 2.

In **the sea turtle** analysis, EPA dismisses the monitoring data that show low concentrations of chemical in estuaries. However, these comprise a valid line of evidence and should be considered as part of the estimation of risk to chronic exposure (peak exposures may not have been captured). The measured values are significantly lower than the modeled values.

In the malathion document, the only exposure route where toxicity thresholds are exceeded for **sea turtles** is for uptake into fish in areas where the Rio Grande empties into the Gulf of Mexico (because of



the assumption of high use on cotton crops). Only the loggerhead and olive ridley turtles are known to consume fish, so an assumption of risk is made because the EECs will exceed a sublethal threshold toxicity value. However, there are two reasons to discount this finding:

- i. Fish is a minor dietary component of the diet of both these species who prefer to forage on shrimp, mollusks, crabs, and jellyfish. Therefore, it should not be assumed that fish comprise 100% of the diet so the total exposure will be significantly less [It is worth noting that the Stellar sea lion risk analysis in Section 7.3 did take into account the fact that fish are only one of a variety of organisms that comprise their diet, and assumed 100% fish consumption is inappropriate].
- ii. Secondly, the loggerhead and olive ridley turtles do not forage in the area where the Rio Grande discharges into the Gulf of Mexico. In fact, the olive ridley is not found in the Gulf of Mexico. U.S. foraging areas are along the California coast, where cotton cultivation does not occur. The loggerhead turtle forages along the U.S. Gulf Coast, but is found primarily *east* of the Mississippi delta.
- iii. Furthermore, the analysis shows that any direct effect via spray drift is discountable more than approximately 30 m from the application site, and most coastlines do not have agriculture that close to the shore (see above comment).

Green **sea turtles** are found only in southern Florida, so direct effects via uptake to plants should be represented only by HUC 3 – the highest calculated value for HUC 3 is less than that of HUC 13 which provided the high value used in the analysis. Furthermore, the 21-d, 60-d, and 90-d average peak values for 1-in-15 years are much, much lower than the 4-d average peak value. Therefore, the conclusion for direct effects for all species for all three chemicals should be “low risk” and NLAA.

A similar argument can be made for the **manatee**, which is found only in HUC 3. Potential for freshwater exposure should be based on the values from this HUC alone. Again, the 4-d average value for 1-in-15 years may be overly conservative for this analysis and a more realistic value of 21-d, 60-d, or 90-d average should be chosen.

Dermal exposure in pinnipeds (**Section 7.3**) should not be comparable to terrestrial species due to the subcutaneous blubber layer. Similarly, polar bears and otters have different fur properties to reduce water penetration and therefore reduce the amount of contact of the chemical with the skin. This uncertainty, combined with the uncertainty of exposure potential, is sufficiently great that the dermal exposure route should not contribute to an LAA conclusion for marine mammals.

Inhalation exposure for marine mammals (**Section 7.3**) has an additional level of uncertainty than what was mentioned in the text. Because most marine mammals make protracted dives underwater, they have mechanisms for reducing blood flow to extremities and differing breathing patterns when surfacing. This likely will change the metabolism and distribution of chemical through the



body, making extrapolations of toxic effects highly uncertain. The inhalation exposure route should not contribute to an LAA conclusion for marine mammals.

The analyses for the cave-dwelling terrestrial invertebrates (**Section 7.4**) are identical for all three chemicals. They are all based on assumptions of exposure of dietary items occurring outside the caves and then being washed into the cave or moving into the cave after exposure and prior to death. There is no quantification of this probability, and the statement is made that “there is a potential for effects if exposure occurs, but it’s not clear what effects would occur at the exposure concentrations.” The probability that is not addressed is the frequency with which prey items might be exposed to lethal concentrations of the pesticides sufficiently close to the cave mouth that they would be transported into the cave, with the exception of documenting that leaf litter could have lethal concentrations for several days. Mapping the overlap of the “action area” with the species habitat should be based on the location of cave mouths, not on the general area within which caves occur (e.g., karst formations of west Texas). This will narrow the area of overlap and help identify the type of agricultural activity that is most likely to result in pesticide use near/around cave mouths. The distance from an agricultural field to a cave mouth should be included in the analysis. Given all the assumptions and qualitative nature of the assessment, the Confidence for direct effects should be “low,” not “medium” as shown in **Table 4-21**.

Section 8 of Chapter 4 (Refined Risk Analysis for Bird Species) is an excellent example of how the Step 2 analysis should be approached for all species, data and resources permitting. It provides a probabilistic analysis that is much more informative than the simple deterministic risk ratio and follows the recommendation of the NAS report. Unfortunately, there are deficiencies in model implementation and parameterization.

The Refined Risk Analysis in **Section 8** of Chapter 4 highlights the importance of the selection of the level of effect, e.g., HC05 versus HC50. More information needs to be provided in the prior sections about why particular thresholds are selected.

The Refined Risk Analysis in **Section 8** also provides information about risk of mortality versus risk of reproductive effects, and how these relate to timing of the pesticide applications. These are extremely useful pieces of information as they can be brought to bear on the LAA/NLAA determination for each typical enduse product (which may have different application times) and for each species life history (i.e., some species are more affected by increased mortality and others by reduced reproduction).



For example, the red-cockaded woodpecker is not particularly vulnerable to mortality of individuals, but reproductive effects on the small breeding populations could be significant.⁶¹ So a determination of a low risk of mortality may result in a NLAA while a similar level of risk of reproductive effects might constitute a LAA determination.

Chapter 4 Page-Specific Comments:

Section 3, Page 4-3

EPA discusses species range geospatial files that were provided by USFWS and NMFS. Could these files please be made publically available for comment by stakeholders, and so that entities outside of EPA and the Services can conduct the same type of assessment?

Section 2, Page 4-2 and throughout

EPA makes numerous highly conservative assumptions and decision in arriving at step three as illustrated by the fact that 97% of species for chlorpyrifos are judged "Likely to Adversely Affect" even when the assessment is refined in step 2. The combinations of conservative assumptions and estimates used by EPA regarding potential effects and modeled exposures that will result in theoretical scenarios that are far in excess of reasonable environmental exposure and risk, even if some of the individual estimates alone are scientifically defensible. We recommend that EPA and the Services reconsider their ESA methodologies to incorporate additional tools to be able to arrive and environmental realistic and scientifically defensible conclusions around ESA.

Section 4, Page 4-5

A "Not Likely to Adversely Affect" determination is made for those species and/or designated critical habitats for which the use site (including off-site transport) and range overlap is less than 1% after rounding for significant digits. This approach is very simple and expeditious and relies on little species information. However, the scientific basis for the 1% overlap is unclear, appears to be arbitrary, and could be improved. This value should be adjusted by species, by the size of the action area, and the location, characteristics and area of the critical habitat. An approach involving such species-specific information would introduce a level of realism, and result in more scientifically-based decision at the Step 1/Step 2 stage. We request that such methods be incorporated into the ESA process at Step 1 and Step 2.

⁶¹ See: US Fish and Wildlife Service. Red-cockaded woodpecker recovery. Updated November 17, 2015. http://www.fws.gov/rcwrecovery/files/RecoveryPlan/executive_summary.pdf.



10.0 TOPICS FOR CONSIDERATION AND COMMENT

10.1 Identification of "Best Available" Spatial Data to Represent Potential Pesticide Use Sites and Species Locations (Attachments 1-2 and 1-3)

Comment 1: Species location data

Section 1.4.1.2 states that species ranges used in the co-occurrence analysis were provided to EPA in the form of Geographic Information System (GIS) spatial files by the FWS and NMFS. However, the spatial data files used to assign "May Affect" determinations were not made available in the documents posted by EPA. As such, a thorough analysis of these files as "best available" is not possible. EPA has released neither a subset of their data nor a summary file indicating species presence at even the county level for all species. Attachment 1-6 contains a "SpeciesRegions_DraftBE" worksheet but very broad region-presence information is provided. State and county presence is provided for some, but not all of the species addressed in this assessment, in the biological information attachments (Attachments 1-11 through Attachment 1-21), but it is not clear how this geographic range information relates to the species location files received from the Services and used in the overlap analysis in Section 1.4.1.3. Additionally, the co-occurrence analysis should consider temporal aspects as migratory species are only present in a given area for a limited duration and thus may not overlap with the action area when a pesticide is applied.

Species presence by Hydrologic Unit Code region (HUC 2) is provided in Attachment 1-10 but the rationale behind using this very large unit and the methodology supporting these pairings is not provided. Presumably a county or sub-county GIS overlay of HUC regional boundaries against species range was performed, but this is not stated. Section 1.4.1.2 in Chapter 1 states that "FWS requested from the species experts in their Regional and Field Offices the most refined range data (*e.g., sub-county level where possible*) for all listed species under their jurisdiction." Based on the information provided in Attachment 1-10, less than 10% of the species range files used to assign HUCs were refined beyond county. Using county-level range data for over 90% of the species analyzed dramatically over-estimates extent and results in inaccurate and unrealistic HUC assignments (*e.g., Sei Whale in HUC 11*). The highest resolution data available should be used to represent species range; county-level range data is not considered the "best available". As requested by EPA and USFWS for this pilot, the FIFRA Endangered Species Task Force (FESTF) generated and provided species range maps for each listed species⁶². These range maps included location data refined to the sub-county level obtained from NatureServe and other sources.

Comment 2: Potential pesticide use sites - Agricultural

⁶² See MRID 49515201. FESTF Species Maps – Phase 1; MRID 49643401. FESTF Species Maps – Phase 2; MRID 49880801. FESTF Species Maps – Phase 3.



The use of the Cropland Data Layer (CDL) to represent potential pesticide use in agriculture is generally considered the best available at the national level. However, there are other datasets, from the Washington State Department of Agriculture and the California Farmland Mapping & Monitoring Program, for example, which would provide more accurate data at the state, regional, or local level.

Also, it is commonly known that each CDL dataset contains “data errors” and varying degrees of data accuracy⁶³. Of note are “spurious pixels,” or small areas that are misclassified as agricultural land. These are individual or a few pixels assigned a land use that is inconsistent with their location (such as one pixel on the shore of a pond that is classified as a particular crop). When multiple CDL datasets are aggregated, the spurious pixels (which are typically in different places for each dataset) can artificially expand the use site footprint and result in use site areas outside the true agricultural land footprint. Attachment 1-3 notes that “several methods have been employed to minimize data errors within the CDL” but according to the process definition, the CDL 2014 Cultivated Layer contains all the misclassified pixels (spurious pixels) for the 2014 data year, without any refinement or validation for actual true crop use areas. Therefore, using the CDL 2014 Cultivated Layer as a cultivated land mask to adjust for land cover mapping errors clearly includes known errors in the resulting adjusted land use areas. When “refining” a dataset, the same dataset containing known errors should not be used as mask; this process inherently propagates known data errors. A more suitable dataset to use for the cultivated mask would be the high-quality National Land Cover Dataset (NLCD) 2011, in which land use data have been validated, and spurious pixels removed from areas that cannot support agriculture. The NLCD 2011 has been relied upon for non-agricultural land use types in these evaluations, so the use of this dataset should be acceptable as an alternative data source. Further, if the NLCD 2011 Cultivated Crops (Class 82) and Pasture/Hay (Class 81) could be used as a general refinement for the aggregated CDL for crops and hay/forage production, then the spurious pixel problems within the CDL datasets may be greatly reduced, resulting in a use site footprint that is more consistent with actual land use.

Comment 3: Potential pesticide use sites - Agricultural

Appendix 1-6 describes that label limitations and geographic use restrictions (e.g., many diazinon uses are only allowed in Texas) were taken under consideration when modeling HUC 2 regions but there is no indication that these limitations and restrictions were considered when developing crop footprints. To incorporate the “best available” data, crop footprints should be limited to registered uses and incorporate geographic use restrictions to accurately reflect the extent of potential pesticide use.

Comment 4: Potential pesticide use sites – Non-agricultural

- The potential use site footprint for mosquito control was not spatially defined and assumed to be in every HUC and every watershed. However, data available from the American Mosquito Control Association, registrants, and mosquito control districts indicate at the county and sub-

⁶³ https://www.nass.usda.gov/Research_and_Science/Cropland/metadata/meta.php



county level where applications for mosquito control take place. This would be a much more realistic spatial footprint representing mosquito control use.

- Appendix 1-6 notes via table footnote that the land cover class “other trees” includes only Christmas trees (with possible misclassification). It is not clear if the land cover class “other trees” was actually utilized to represent the extent of Christmas trees and how the “possible misclassification” was resolved.
- Golf course is listed as a use in the chlorpyrifos master use table, but the dataset representing the spatial extent of this use is not included in Attachment 1-3 with the other non-agricultural uses. Therefore, it is not possible to comment on this dataset as “best available” to represent this use.
- Attachment 1-3 notes that NLCD Developed or Open Space Developed land use categories were used to spatially represent certain non-agricultural label uses, but it is not clear which non-agricultural label uses are represented by which land use category. Details should be provided for these non-agricultural uses.
- The use site footprints for nursery uses were derived using a proprietary business database, Dun and Bradstreet. It is difficult to evaluate its use and conclusions drawn from the data without the knowledge of what the data looks like or the metadata associated with the dataset. Efforts should be made to make this data publicly available. Additionally, the description for this footprint notes that SIC codes “018” and “526” from Dun and Bradstreet were buffered by their facility size. Despite efforts to map production facilities only, this method overstates potential use in nurseries because it includes businesses with no nursery facilities, such as those that are strictly lawn supplies stores, and nursery facilities that do not use pesticides at all.
- Spatially mapped rangeland is the best characterization of the Cattle Ear Tag use. However, tags are used only when pest pressures are high and very minimal off-target exposure is expected. It is suggested the spatial extent of this use be refined with cattle density information from USDA or the Census of Agriculture Farm Survey to determine the presence of cattle and potential exposure.

10.2 Methods Used to Identify Potential Overlaps (and Extent) of Species Locations and Potential Use Sites and their Applications in Effects Determinations Made in Steps 1 and 2 (Attachment 1-6)

This section reviews the details of the use of spatial datasets, methods, and tools to determine the co-occurrence of the Action Area and species locations. The evaluation focuses on the general approach and is generic to all three OPs. However, because the total use site footprints for malathion and chlorpyrifos were considered to be the entire US, the methods presented for the more limited use site footprint of diazinon served as a better example for review. The primary document considered in this evaluation included:



- 1.) Chapter 1, Section 1.4.1.1: Action Area
- 2.) Attachment 1-4: Process for Determining Effects Thresholds
- 3.) Attachment 1-6: Co-Occurrence Analysis
- 4.) Appendix 3-5: Downstream Dilution Tool Results

Comment 1

Document 1: Chapter 1, Problem Formulation, Section 1.4.1.1c: Action Area, Off-site transport area;
Document 2: Attachment 1-4: Process for Determining Effects Thresholds, Section 1: Effects Thresholds for the Action Area (Step 1)

The determination of the spray drift off-site transport distance component of the Action Area was based on spray drift modeling results and effects end points for the most sensitive aquatic or terrestrial species. For the aquatic species, the Expected Environmental Concentrations (EECs), due to spray drift, for the most sensitive aquatic habitat bin (Bin 5) was compared against the end point for the most sensitive taxa (aquatic invertebrates) to determine the spatial extent of the off-site transport distance. The use of both the most sensitive aquatic habitat, which many aquatic species do not occupy as habitat, and the effect endpoint that is several orders of magnitude lower than is appropriate for many taxa, results in an Action Area that is scientifically inaccurate for many taxa and species. The extent of the spray drift based off-site transport zone and associated Action Area needs to be specific to the aquatic habitat characteristics a species occupies and based on relevant conservative effects end points.

Comment 2

Document: Appendix 3-5: Downstream Dilution Tool Results, Section: Use of Downstream Dilution Tool in Step 1

The downstream dilution approach is a simplified method to estimate pesticide concentrations in flowing water bodies downstream of potential use sites. The method assumes that conservative estimates of pesticide concentrations in small high vulnerability water bodies occur on all potential use site areas across the landscape, and that all non-potential use site areas of the landscape would generate pesticide concentration of 0. From this simple perspective, the "Percent Treated Area" (PTA) upstream of any point along a stream network can be used to estimate the pesticide concentration by diluting the estimated concentration from the use site areas by the contributions from the areas with no pesticide use. Implicit in this approach is the notion that all land uses, soils, and slopes within a watershed generate runoff and baseflow contributions to streamflow at the same rate. The approach also does not account for any environmental fate processes of the pesticide during its travel downstream. The EPA recognizes these simplifying assumptions in the BE documents, however it should be emphasized that the downstream dilution methodology is not a substitute for a processes-based watershed hydrologic model which is necessary to achieve more accurate estimates of pesticide concentrations in flowing water bodies.



Comment 3

Document: Appendix 3-5: Downstream Dilution Tool Results, Section: Use of Downstream Dilution Tool in Step 1

In determining the downstream dilution component of the Action Area, EPA assumed “the rights-of-way scenario (a protective scenario utilizing a high curve number)” to generate an EEC that was assumed to reflect the entire use site footprint of the active ingredient (AI). A “right-of-way” use is an off-label use pattern for the AI for which this scenario was applied, and as such is not relevant to the labeled use patterns for the AI (limited to orchards/vineyards, vegetables and ground fruit, nurseries, and cattle ear tags). The determination of EECs for use in the Step 1 downstream dilution analysis should be based upon EECs associated with scenarios for labeled uses of the AI being assessed.

Comment 4

Document: Appendix 3-5: Downstream Dilution Tool Results (Diazinon), Section: Use of Downstream Dilution Tool in Step 1

In determining the downstream dilution component of the Action Area, EPA assumed the single highest EEC simulated for Bin 2 from all HUC2 watersheds and the worst case use pattern (239,000 µg/L) represented potential EECs from the entire use site footprint. Had this EEC been generated from a specific PRZM scenario associated with one of the diazinon labeled uses (see comment 1 above), it should not have been extrapolated to crop group footprints and HUC2 for which it is not relevant. At the Step 1 downstream dilution analysis, EECs relevant to specific crop groups and HUC2 watersheds should be used in the analysis. The screening level aquatic exposure modeling for individual crop groups and HUC2s has been designed to be extremely conservative, so to select a single EEC that is the highest EEC generated nationally (from an irrelevant scenario) and apply it universally across the entire use site footprint is not scientifically valid. Conservative screening level EECs that are crop group and HUC2 specific should be used in defining the downstream dilution component of the Action Area.

Comment 5

Document: Appendix 3-5: Downstream Dilution Tool Results (Diazinon), Section: Use of Downstream Dilution Tool in Step 1, Use of Downstream Dilution Tool in Step 2

The use site footprint used in the Step 1 downstream dilution analysis included uses that do not pose a runoff exposure potential issue. In the diazinon BE, a use site footprint for Cattle Ear Tags was included for downstream dilution concentration calculations, yet the Cattle Ear Tag use was later acknowledged not to pose a runoff risk (Section “Use of Downstream Dilution Tool in Step 2”). Downstream dilution concentrations can only be based upon uses that have the potential to contribute to flowing water exposure resulting from runoff. Uses that do not contribute to runoff should be excluded from the site footprint use in downstream dilution.



Comment 6

Document: Appendix 3-5: Downstream Dilution Tool Results (Diazinon); Section: Use of Downstream Dilution Tool in Step 1

The Step 1 downstream dilution analysis used the “lowest aquatic endpoint” to calculate the dilution factor (DF) in determining the Action Area for all species. This lowest aquatic endpoint based on the most sensitive species taxa can be significantly different from end points for the remaining taxa and species. The use of an end point for an aquatic invertebrate (the most sensitive taxa for the OP pesticides) to determine a downstream dilution factor for aquatic plants (with endpoints several orders of magnitude higher than aquatic invertebrates) is not scientifically justifiable. The determination of downstream dilution Action Area extent should be relevant to the species being assessed, and should be minimally based on effects end points related to the taxa to which the species belongs. Not accounting for the wide range in species sensitivities to exposure results in inappropriate “May Effect” determinations for species that are insensitive to the pesticide exposure.

Comment 7

Document: Appendix 3-5: Downstream Dilution Tool Results (Diazinon), Section: Use of Downstream Dilution Tool in Step 2

The Step 2 downstream dilution approach followed in the EPA BE is a refinement of the Step 1 downstream dilution, and is similar to the recommended process for Step 1. It is unclear from the BE description of how EECs used in the downstream dilution analysis are determined for areas where multiple crop group footprints overlap. A conservative approach would be to take the EEC from the crop group with the highest EEC. A more realistic approach would be to take the average of the EECs for the crop groups that overlap at each pixel, accounting for the variability and uncertainty of the crop group footprints.

Comment 8

Document: Appendix 3-5: Downstream Dilution Tool Results (Diazinon); Section: Use of Downstream Dilution Tool in Step 2

The Step 2 (and Step 1) downstream dilution approach followed in the EPA BEs used a worst case 1-in-15-year annual maximum instantaneous peak concentrations from Bin 2 simulations with PWC as the basis for calculating downstream EEC dilution factors. The instantaneous peak concentrations from the PWC model can be several orders of magnitude higher than the daily average concentrations and are based on a physically impossible conceptual model of exposure in flowing water bodies (see comments on aquatic exposure modeling). The use of instantaneous peak EECs in the downstream dilution analysis is conceptually incorrect, because a nominal amount of travel time is required for landscape-generated pesticide loads to reach downstream locations. In addition, given that the landscape model (PRZM) used



to generate the pesticide and runoff loads operates on a daily time step, daily average pesticide concentrations should be used as the basis for downstream dilution calculations.

Comment 9

Document: Appendix 3-5: Downstream Dilution Tool Results (Chlorpyrifos), Section: Use of Downstream Dilution Tool in Step 2

Downstream dilution was not conducted for chlorpyrifos "because of the widespread use of chlorpyrifos and the uncertainty with where the adultimate, wide area, and non-agricultural uses could occur, the entire United States is considered the action area for chlorpyrifos for Step 1." (Appendix 3-5) The same rationale was applied for Step 2. However, the tables in Appendix 1.6 and 1.7 indicate that developed uses (which include mosquito control) are separate from agricultural uses and there are certain agricultural use sites that are not applicable across all HUCs (e.g., rice and grassland). If applications are excluded for rice and grassland in some HUCs then it would not be valid to assume use on these sites for any purposes, including mosquito control. Some agricultural crop groups, where use is restricted, should not be assumed to allow a non-Ag, wide area use in Action Area delineation and co-occurrence analysis.

Comment 10

Document: Chapter 4 Effects Determinations, Section 4: Effects Determinations of NLAA - No overlap

A NLAA is determined in Step 2 for all species and/or designated critical habitats with less than 1% co-occurrence with use sites (including off-site transport). There is no information provided for the basis of the 1% cut-off. It is unclear whether the 1% co-occurrence represents a level of co-occurrence determined to be ecologically significant, or if there is other justification in the selection of 1% as opposed to a different nominal threshold (e.g., 0.5%, 2%, 5%). In addition, this implementation of a co-occurrence threshold does not account for uncertainty in the CDL pixel classifications. Recent approaches for developing probabilistic crop footprints using the CDL would help quantify the likelihood of co-occurrence while accounting for dataset uncertainty (see Budreski et al., 2015).⁶⁴

⁶⁴ Budreski, K., M. Winchell, L. Padilla, J. Bang, and R. A. Brain. 2015. A Probabilistic Approach for Estimating the Spatial Extent of Pesticide Agricultural Use Sites and Potential Co-Occurrence with Listed Species for Use in Ecological Risk Assessments. *Integr Environ Assess Manag*. doi: 10.1002/ieam.1677.



10.3 Estimation of Exposure in Various Aquatic Environments (Bins) that have been Regionally Delineated and the Parameterization of Bins and their Relevance across the Landscape

10.3.1 Static Waterbody Parameterization and Methodologies

The reliance on a fixed volume waterbody led the Agency to make other adjustments (DA/NC/watershed size) that introduced extreme concentration predictions. For static bins specifically, the approach of calculating a large watershed to support a static environment introduces errors and led to clearly higher predicted concentrations in dry climates (Table 3-7 chlorpyrifos). The assumption that a large land area could drain to a single puddle, pool, or pond is not documented and their existence supporting endangered species is not documented or justified. One approach would be to start the assessment where the species exist and guide the tools and assessment to look up gradient from there, rather than adjusting watersheds to support an environment that may not exist. More specifically:

- The documentation notes that watershed (and therefore) DA/NC are set to achieve a DA/NC around $15 \text{ m}^2/\text{m}^3$ for Bin 5 and a D/ANC “around” 12 and $5 \text{ m}^2/\text{m}^3$ for Bin 6 and 7, respectively. There is limited discussion of the source or meaning of these values and no discussion as to how an endangered species habitat is represented by the waterbody or these back calculated “watersheds” to support the static waterbody.
- Evaluating the supplemental information reveals a large range of predicted field sizes/watershed derived from the 242 meteorological station simulations (using the single high runoff soil scenarios, MSCorn). The data was organized by HUC2 and a selection of the 10-25% result for Bin 5 or 75% result for Bin 6 and 7 within each HUC was made, to achieve the desired outcome (it appears achieving the desired final median value was the goal, how these different percentiles were selected is unclear). The data reveals a large range of calculated DA/NC within and across HUC2 areas and highlights that as an organizational unit, the HUC2 is not correct when the goal is predicted concentrations in a localized hydrologic feature.
- Table A 3-1.8 summarizes the selection of Watershed/DA/NC for each HUC2 which is the percentile result within the HUC2 region. Furthermore, describing the results based on the median value simplifies the range and actual impact of the values in the assessment. A more accurate description of the DA/NC values being used in the assessment can be found in the table below, and it is important to note that the HUC2 areas with the larger defined watersheds (drier climates) also resulted in higher predicted concentrations.



DANC (m ² /m ³)	Bin 5	Bin 6	Bin 7
Median	14.9	11.0	5.7
Average	154.7	44.2	18.3
Max	1255.0	204.0	78.8
Min	0.8	0.6	0.3
StdDev	311.4	66.9	25.7

10.3.2 Regional Aquatic Modeling Parameterization and Species Relevance

HUC-2 is too large an area and any single representation derived from it unrealistic. A HUC-2 represents a drainage basin and has no commonality with rainfall, soils, or crop patterns, it simply identifies which major river or coastal water the land area eventually drains to.

10.4 Evaluations of Exposures in Flowing Water Bodies and in Non-Freshwater Habitats

10.4.1 Flowing Water Systems

The Bin 2 flowing water Peak EECs are impossibly high (and admittedly so per EPA public statements) due to flaws in the modeling design and assumptions. This is readily demonstrated by comparing the model-simulated 1-in-15-year edge-of-field concentrations to the receiving water concentrations generated by flawed flowing water modeling. The edge-of-field concentrations produced by model simulations that have been described in the Overview Document and vetted through the Science Advisory Panel are several orders of magnitude lower than the resulting peak receiving water concentrations for all scenarios. This is an impossible situation: that the dilution process produces receiving water levels of pesticide residues several orders of magnitude higher than edge of field levels. Dilution physically cannot result in concentration and it is inappropriate to put forth values generated by these untested models as meaningful.

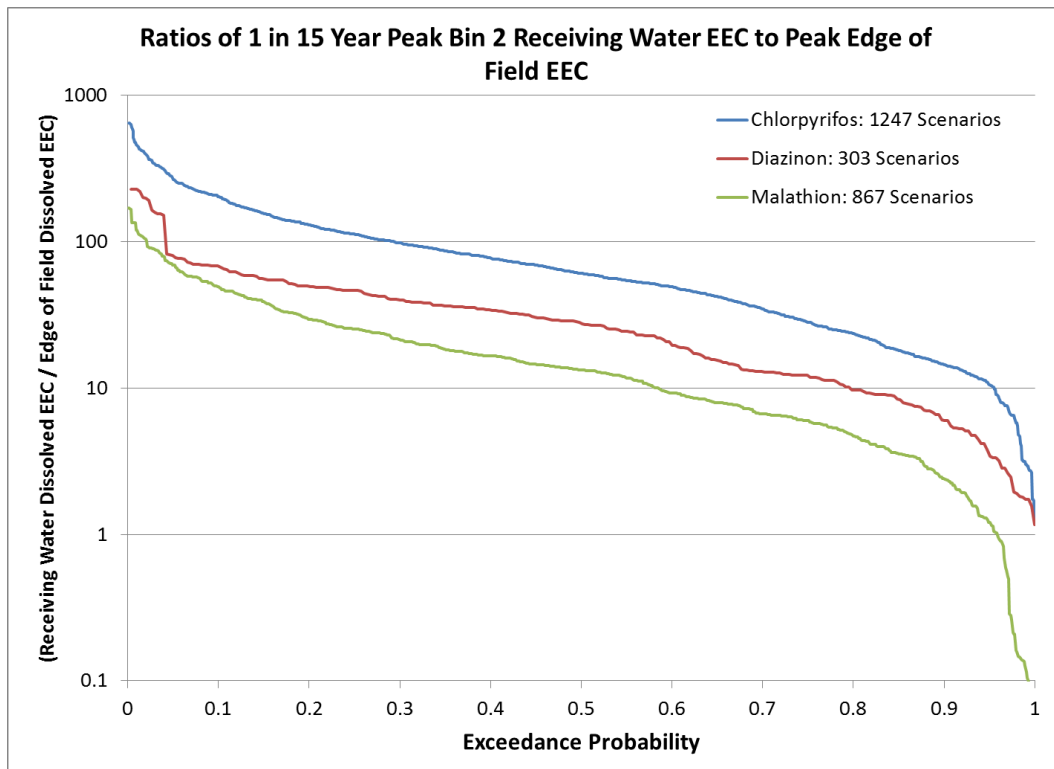
As a way of evaluating model output using results from the chlorpyrifos assessment, the distribution of EECs generated are shown for a population of 1,056 chlorpyrifos scenarios (excluding seeds and urban uses) in Figure 9 below. These results are compared in the figure to both dissolved and total edge of field concentrations.

As a way of quantifying the degree of the over-prediction of Bin 2 EECs, the ESA PRZM/VVWM scenarios were re-run for each of the 3 OPs with drift set to a negligible amount (drift fraction of 0.001), thus ensuring that 1 in 15 year annual maximum peaks would be runoff driven rather than drift driven. These Bin 2 receiving water EECs were then compared against the matching 1 in 15 year edge of field concentrations. For a pair of EECs associated with a single ESA scenario, the ratio of the receiving water 1 in 15 year peak EEC to the edge of field 1 in 15 year EEC was calculated. Any ratio with a value of



greater than 1 indicated a receiving water EEC higher than the edge of field EEC. The results of this analysis are summarized in the figure below for all chlorpyrifos, diazinon, and malathion. Receiving water concentrations are 10x or more higher than edge of field concentrations for 95% of chlorpyrifos scenarios, 80% of diazinon scenarios, and 58% of malathion scenarios. To have receiving water concentrations of a diluent several orders of magnitude higher than the contributing source of the diluent is physically impossible and does not comply with logic or the conventional understanding of the dilution process. A more credible expectation is for dissolved receiving water EECs to be lower than edge of field dissolved EECs. Deriving EECs for larger flowing water bodies (Bin 3 and Bin 4) from these erroneous Bin 2 EECs propagates this error to affect a greater number of species than need be included in the assessments. The flaws in this modeling must be corrected.

Figure 9. Ratios of 1 in 15 Year Peak Bin 2 Receiving Water EEC to Peak Edge of Field EEC



The errors in the screening level flowing water EECs can be addressed by taking the following steps:

- Replacing VVWM with a receiving water model designed to simulate pesticide fate and transport in a flowing channel. The Soil and Water Assessment Tool (SWAT) has this capability and has been shown to produce realistic peak exposure values for small, medium, and large flowing water bodies.



- Incorporating a baseflow rate equal to the minimum of the flow range associated with each habitat bin. Bin 2 baseflow = 0.001 (m³/s); Bin 3 baseflow = 1 (m³/s); Bin 4 baseflow = 100 (m³/s)
- Constraining the watershed areas to that which can drain into a main channel within 1 day.
- Applying Percent Cropped Area (PCA) adjustments at a minimum to Bin 3 and Bin 4
- Winchell et al.⁶⁵ provides a proposed methodology for using the SWAT in screening level modeling of generic flowing water bins.

Aquatic exposure modeling at Step 2 should move beyond simple screening level approaches that use a single conservative PRZM simulation to predict EECs in flowing water bodies draining heterogeneous watersheds. Step 2 aquatic exposure modeling approaches should include the following:

- Representation of the heterogeneous landscape through explicit simulation of the land uses and soils that comprise a given watershed.
- Spatially explicit predictions of EECs that can be associated with species habitat locations.
- An accounting for variability in pesticide application timing that occurs at the watershed scale.
- Incorporation of Percent Treated Area (PTA) that acknowledges that 100% of potential use sites do not get treated with a given pesticide.

Additional comments on the specific aspects of the flowing water conceptual model and parameterization are provided below.

1. Attachment 3-1: Background Document: Aquatic Exposure Estimation for Endangered Species, Section 3.1.2: Revised Conceptual Model and Approach for ESA

- Figure A 3-1.1: The conceptual model for flowing water habitat exposure assumes that the treated field (a square the size of the entire watershed) is adjacent to the water body across the watershed's entire length. As watersheds become large (Bin 3 and Bin 4), this conceptual model of a watershed deviates more significantly from reality. This is in part what led to increasingly higher and more extreme EECs for the modeled Bin 3 and Bin 4. The watershed conceptual model should be reconsidered to better reflect actual watershed characteristics at Step 2 modeling.

2. Attachment 3-1: Background Document: Aquatic Exposure Estimation for Endangered Species, Section 3.1.3.2: Aquatic Habitat Bins

⁶⁵ Winchell, M., L. Padilla, N. Pai, N. Poletika, P. Whatling and P. Havens. 2016. Development and evaluation of a screening level flowing water exposure modeling approach for endangered species assessments. USEPA Environmental Exposure Public Meeting, Arlington, VA. May 9, 2016.



- The Bin 2 parameters do not represent the vast majority of low flow habitats. The Bin 2 description is as follows, “Some examples of low flow habitats include springs, seeps, brooks, small streams, floodplain habitats (oxbows, side channels, alcoves, etc.), dendritic channels that occur within exposed intertidal areas, and distributary channels in estuaries on the incoming and slack tides.” The assumed velocity, of 1 foot per minute, is indicative of an extremely slow moving, essentially stagnant, waterbody, and does not reflect the vast majority of types of habitat listed that Bin 2 represents. Hydraulic calculations using Manning’s equations and a very high roughness value of 0.125 result in a channel slope of 0.00001, which is not typical of headwater streams across the US.
- Flowing Bin Flow Rates. The flow rate characteristics of each Bin were given a range (e.g., Bin 2 = 0.001 – 1 m³/s). In each of the 3 flowing bins, the lowest flow rate was assumed when parameterizing the models. While this is conservative, given that the flow rates vary over 3 orders of magnitude within a single Bin, it would be good to evaluate the effects of higher flow rate assumptions when modeling in Step 2.

3. Attachment 3-1: Background Document: Aquatic Exposure Estimation for Endangered Species, Section 3.1.3.3: Watershed Sizes

- Watershed areas for the flowing bins were estimated using a regression of area versus average annual flow from the USGS NHDPlus dataset. The watershed sizes estimated, while accurately representative of the targeted flow rates of the three bins (0.001, 1, and 100 m³/s) are not appropriate sized drainage areas for use with the VVWM model.
 - VVWM received runoff and pesticide loads from PRZM on a daily time step. The watershed areas derived based on the NHDPlus regression equations result in watershed areas that exceed a 1-day unit hydrograph base time for Bin 4 scenarios. Watershed areas need to be constrained to areas that can fully drain into a main channel within 1 day.
 - The watershed areas for the flowing bins resulted in drainage area to normal capacity ratios (DA/NC) of 1.) Bin 2: 334 – 7725, 2.) Bin 3: 604 – 11218, 3.) Bin 4: 739 – 35614. The EXAMS/VVWM model has long been used to simulate receiving waterbody pesticide concentrations from much smaller watersheds, with DA/NC of 5 (EPA farm pond). The use of VVWM for scenarios with such extremely different ratios of drainage area to water body size needs to be validated to ensure that results are appropriate for such a different hydrologic scenario.

4. Attachment 3-1: Background Document: Aquatic Exposure Estimation for Endangered Species, Section 3.1.4.2.1: Issues Modeling Medium- and High-Flowing Waterbodies, Modifications Considered But Not Incorporated

The following improvements for Bin 3 and Bin 4 modeling were explored but not considered:



- **Incorporation of Base Flow:** This section discussed approaches to incorporating baseflow into the model parameterization for Bin 3 and Bin 4. Several methods for estimating baseflow rates were presented, however no conclusions were presented as to a preferred approach to estimate baseflow. Estimation of baseflow for Bin 2 was not mentioned in this discussion. Baseflow occurs for all sized streams across much of central, eastern, and more humid parts of the western US. If there were no baseflow, we would not have a flowing aquatic habitat, with the exception of storm events. For screening level flowing water modeling, a nominal baseflow rate should be set for each of the 3 flowing habitat bins. Given that the flow rates that define each bin span a range (e.g., 1 to 100 m³/s for Bin 3), the lowest flow of the range should be assumed to represent the baseflow condition. Flows below this amount do not define the bin of the species that inhabit that class of waterbody. During storm events, the flow rates will increase to levels above the minimum defined for the bin.
- **Percent Use Area and Percent Use Treatment Adjustment Factors:** The concept of using Percent Crop Areas (PCA), Percent Use Areas (PUA) or Percent Treated Areas (PTA) was proposed. It was noted that these factors are used by HED in their risk assessments for human exposure. Both PCA and PTA should be incorporated for all 3 flowing bins. PCAs can be calculated based on aggregate use site footprints from multiple crop groups, and PTA, while based on state-level data, can be meaningfully included in the determination of EECs at Step 2. PCAs should be determined independently for each habitat bin and be based on appropriately sized watersheds (e.g., NHDPlus catchments for Bin 2, HUC10 to HUC12 for Bin 3). The use of a state-level PTA should transfer directly to the larger Bin 4 watersheds, which are at the same scale. State-level PTA can be used in several ways for Bin 2 and Bin 3, either as multiplicative factor to PCA, or, by splitting the EECs into percentages assuming worst case 100% PTA and the remainder assuming 0% PTA (no exposure) to then determine the likelihood of exposure at the level associated with use on all labeled crops. Because PCA and PTA are very important in determining exposure concentrations, these factors should be considered rigorously at Step 2, rather than waiting until Step 3.
- **Adjustment of Water Body Length:** It was suggested that increasing the water body length, accounting for a meandering stream, would increase the water body volume and thereby reduce the DA/NC ratios lowering the EECs to more realistic levels. Given the current use of VVWM, this effect would be true, but it is the underlying use of the VVWM to model a flowing channel that leads to the problems of unrealistically high EECs with large DA/NC ratios. The extreme EECs with higher DA/NC ratios do not occur with models designed to simulate channel flow, such as SWAT (see Winchell et al., 2016⁶⁶). We suggest that water body length is not the main issue to address for the flowing scenarios.
- **Spreading Out Applications:** Based on testing with diazinon, it was determined that spreading out applications across a window equivalent to a retreatment interval, did not have a significant

⁶⁶ Winchell, M., L. Padilla, N. Pai, N. Poletika, P. Whatling and P. Havens. 2016. Development and evaluation of a screening level flowing water exposure modeling approach for endangered species assessments. USEPA Environmental Exposure Public Meeting, Arlington, VA. May 9, 2016.



effects on peak EECs. Over larger areas, it is likely that applications timing will vary over a wider window than only the application interval, and would have significant effect on peak EECs. Accounting for application date variability should be considered in Step 2 aquatic exposure modeling.

5. Attachment 3-1: Background Document: Aquatic Exposure Estimation for Endangered Species, Section 3.1.4.2.2: Issues Modeling Medium- and High-Flowing Waterbodies, Modifications Explored and Incorporated into Modeling

The following improvements for Bin 3 and Bin 4 modeling were explored and incorporated into a case study:

- **Curve Number Adjustment:** An area weighted CN value at the HUC12 level was calculated and used to select a 90th percentile CN value for each HUC2 and used as the representative CN value for Bin 3 and Bin 4. This is a very sound approach to obtain appropriately conservative CN values. A suggestion would be that for Bin 4, a larger watershed (say HUC8 or HUC6) 90th percentile CN value be used to more appropriately represent that Bin. In addition, this approach should be applied for Bin 2 as well, using NHDPlus catchments as the watershed units from which to select a 90th percentile CN value for each HUC2.
- **Daily Flow Averaging:** The option of using daily flow averaging (instead of 30-year flow averaging) in VVWM was evaluated. It was mentioned that the SAM model was used for daily flows, but it is very unclear how this was done. It is assumed that simply the VVWM flow averaging period was simply changed from 30 years to 1 day to implement this option, and that SAM was not required. Nevertheless, the daily flow averaging option should be a requirement for modeling of all flowing bins (2, 3, and 4).
- **Adjustment of Water Body Dimensions:** This option used a “representative length” of 40 m to represent that water body length (as opposed to the full watershed length) and adjusted the volume accordingly. This “representative length” has been proposed for use in SAM. While this option may be effective at lowering EECs closer to realistic levels, it is conceptually inaccurate, and would not be necessary if an appropriate flowing water receiving model were used to route the flows (such as SWAT).
- **Use of Daily Average EECs:** The use of daily average EECs instead of peaks was proposed. One argument was that typical toxicological studies have a 48-hour duration. For this reason, daily average or 48-hour average EECs should be used instead of peak EECs for all aquatic habitat bins. There are no toxicological tests that measure effects of instantaneous exposure.

6. Attachment 3-1: Background Document: Aquatic Exposure Estimation for Endangered Species, Section 3.1.4.4: Modifications Evaluation, Pilot Chemicals, Final Approach



- The alternative modeling approaches for Bin 3 and Bin 4 modeling were evaluated for atrazine and compared against monitoring data. The proposed approaches resulted in the most refined EECs comparing relatively favorably with atrazine monitoring data, so the refinement approaches were tested for the 3 OPs.
- Based on testing with the 3 OPs, the alternative modeling approaches produced similar results for some cases. Values higher than solubility were still noted and cases of Bin 4 and Bin 3 EECs higher than Bin 2 were also noted. The lack or improvement in EECs is being driven by the limitations in receiving water model (VWWM). Several of the alternative modeling approaches considered (CN, daily average EECs) are good ideas for refinement and should be pursued in conjunction with a more appropriate receiving water model.
- Final Approach to Calculating Bin 3 and Bin 4 EECs: Based on review of atrazine monitoring data, scaling factors were derived to estimate Bin 3 and Bin 4 EECs from the model-simulated Bin 2 EECs. While these values are more appropriate than the modeled values that had been calculated, there are limitations to this approach:
 - There is no data presented to support that the relationship between atrazine from small headwater streams to larger rivers would apply directly to the 3 OPs. Atrazine's environmental fate properties differ from the 3 OP, including solubility, adsorption to sediment, and degradation rates, all of which impact the fate of the pesticide over time in a flowing water system.
 - For the monitoring datasets assessed (AEMP and Heidelberg University), atrazine is used on crops with very high PCA (corn/sorghum) and is very widely used on those crops (a high PTA). These differences in use pattern between atrazine in the watersheds evaluated and the 3 OPs make the transfer of conclusions on atrazine to the 3 OPs inappropriate.

10.4.2 Appropriateness of Marine Bin Surrogates

Neither tidal pools nor offshore marine environments can be modeled appropriately with existing habitat scenarios which represent homogenous watershed areas draining into pond-like receiving waters that are either static or allow for steady inflow/overflow. In addition, the EPA standard scenarios defining weather, crop, and soil characteristics for each HUC2 are based on inland environmental conditions and therefore are not representative of coastal environments.

Watershed areas assumed for the marine scenarios (based on their surrogate freshwater scenarios) are not relevant to these habitats. The watershed sizes from the surrogate habitats, determined through regression of flow rate with drainage area (flowing bins) or precipitation/evaporation/runoff balance to maintain a specific normal capacity (static bins), are not relevant for modeling pesticide exposure in marine habitats and are not related to their capacity.



The habitat surrogates have inappropriate flow regimes for the marine bins. The rate, magnitude, and frequency of tidal flushing and currents cannot be approximated by the steady stream inflow/outflow rates used in the surrogates freshwater bins. For tidal pools, tidal action will have a major impact on dilution of any pesticide residues present in the pool via partial or total exchange of the tide pool water with seawater every 12 hours. This is fundamentally different and requires a different modeling approach from the freshwater bins.

10.5 Evaluation of Exposure to Terrestrial Organisms, including Dietary and Non-Dietary Routes of Exposure

It is unclear what ecotoxicological endpoints were used in the risk assessments for terrestrial invertebrates (i.e., what specific toxicity test results were used to assess the risk for each listed species). There appears to be no foundation for the conclusions reached. Given these examples of uncertainties, it is surprising that the Agency concluded that there is a high degree of certainty in their risk predictions.

The statement is made "*There are currently no toxicity data available for terrestrial plants and chlorpyrifos that have mortality as an endpoint.*" This is a false statement, there is data from the pre-emergent and post-emergent studies. If mortality occurred it would be noted. A more accurate statement would be that no mortality occurred in these studies at the rates tested.

The information in the weight of evidence tables for the different species indicates that there is information available for a given number of species, but does not state what the endpoints were or how they were used. This makes it impossible to validate the application of data for terrestrial invertebrates. Therefore, a thorough scientific analysis that ties assigned values to species is required. Evaluation of results based on currently available information is otherwise impossible. It is also unclear what toxicity information was used to predict the sensitivity of snails.

10.6 Use of Species Sensitivity Distributions to Evaluate Effects (Attachment 1-5)

When developing the SSD for avian acute oral toxicity, page 2 of Attachment 1-5 describes that acute oral LD50 endpoints are normalized to represent birds with a body weight of 100g prior to curve fitting. This normalizing step is not necessary since each endpoint is reported in dimensions (mg/kg bw) that are normalized for body weight. The question of relevant exposures (perhaps conservative to start) and impacts to a subset of the prey based (ecological meaningful) have to be addressed.



EFSA,⁶⁷ the AF is reduced from 100 (for acute) and 10 (for chronic) to 3-6 (the selection of the AF depends on the data set size and the distribution of the data along the SSD curve). Since it has been shown (and would be expected) that listed species are not inherently more sensitive to chemicals than non-listed species, application of the SSD approach in the listed non-target organisms would be expected to have a similar reduction of the AF from one-in-a-million. How much it needs to be reduced to will require further understanding of how much the uncertainty plays in the data extrapolation.

SSDs have been used in chemical risk assessment for at least 20 years as an improvement over basing risk characterization on the most sensitive species. Theoretically, an SSD depicts the fraction of species in a community that would be affected (by a specified amount, e.g. 50% mortality) at a given exposure concentration, and the HCx represents the concentration causing a certain magnitude of community effect. EFED uses SSDs in this way for assessment of indirect effects (e.g. indirect effect on fish diet due to direct effect on invertebrate community). Ecologists have supported the use of SSDs for this purpose, with certain qualifications. The big question, of course, is relating overall community impact to the fraction of species affected, something that has not been done in the Agencies' evaluation of indirect effects.

In the BEs, the SSD is also being used for a different purpose, namely, to make an assumption that an untested listed species is equal in sensitivity to the xth percentile of tested species. This seems sensible, in principle, and when enough data are included in the SSD it avoids the assumption that the listed species is as sensitive as the most sensitive tested species (which may be an outlier). One can argue about what percentile is appropriate to assume – 5th percentile (HC5) seems highly conservative, since it has been shown (and would be expected) that listed species are not inherently more sensitive to chemicals than non-listed species. Theoretically, the SSD describes an underlying distribution of species sensitivities, and only makes sense if species are compared using similar endpoints. There is no point comparing an NOEC for lethargy for Species A with an LC50 for Species B – the comparison doesn't yield anything about which species is more sensitive. Lethal and sublethal endpoints cannot be mixed and endpoints measured at different life stages (e.g., time-wise) should not be mixed as well. It is clear that only the same set of the endpoints from different species can be used to generate a valid SSD. The protection level on which the assessment relies determines what endpoints need to be focused on for the SSD to reduce the uncertainty in data extrapolation. The same group of endpoints (survival; growth; reproduction; off-spring survival, etc.) that are used for the SSD will be helpful in determining a reliable HCx that is relative to the protection level.

The data set that can be used for SSD is in general from the standard studies with a range of species that are tested. There are two factors that should be considered for the data quality in regard to the usage in SSD.

⁶⁷ EFSA Journal 2013; 11(7):3290



1) Whether the group of the non-target organisms tested are sensitive to the product (depends on the mode of action).

For these three biological evaluations, the most useful/sensitive groups of non-target organism are recognized as arthropods (by literature). If the SSD approach is applied to non-target plants, its use should be evaluated with respect to bringing more certainty to the risk assessment.

2) The quality of the data set that is used in the SSD also depends on the selection of the data, and the process of the data within the same species.

Following the Agency test guideline that is designed for generating pesticide endpoints, and, where used, having assurance that published data are reliable and relevant, are crucial steps to ensuring whether the data generated are reliable or reproducible. A consistent set of toxicological endpoints needs to be used in a given set of SSDs in order to reduce the uncertainty. If there are several qualified studies available for the same species, a geometric mean approach should be used to reduce the inter-lab and inter-study variation of the endpoints, which ultimately helps reduce uncertainty.

3) The number of data points that are used for SSD are very crucial to the meaning of an SSD as well. At least 5 data points are needed for vertebrates, and 8 data points are needed for invertebrates according to the practical guidance published by EFSA.⁶⁸

Specifically:

- Conceptually, it is more relevant to use the SSD in a higher tier assessment and not as the source for the one-in-a-million value. The one-in-a-million value would be more appropriate as a tier 1 assessment using the lowest endpoint, if relying on one data point only for a specific taxa (fish, aquatic invertebrates, etc.). In this scenario, the uncertainty would be higher than using data from multiple sources. With the SSD, the uncertainty level is significantly lower because many species are used to generate the regression curve to calculate HCx values. Since the uncertainty is much lower when using an SSD, it is proposed that the level of uncertainty (i.e., one-in-a-million) should be adjusted to reflect this decrease in uncertainty. The uncertainty level should be based on the life history patterns of each listed species; a cookie cutter approach would not be applicable.
- The quality of the SSD is only as good as the quality of the data used to construct the regression line. Data quality has been extensively discussed in previous sections. This includes the selection of test organisms tested once by one author (e.g., snake-head catfish, rainbowfish, etc.) and/or sublethal effects that cannot be correlated to whole organism response (survival, growth, reproduction).

⁶⁸ Ibid



- The authors are trying to do too much with the SSDs. That is, the significance of a specific point of interest is diluted or lost because the authors include all data in the SSDs (i.e., horizontal lines) that are of questionable quality at best. It would be more reasonable to apply a set of criteria to all data for a single species/AI (things like flow-through vs static, measured vs nominal, etc.) and identify the most reliable endpoint, and if there are multiple equally reliable endpoints for a particular species/AI, take the geometric mean to use for that species in the SSD. It is also possible to generate SSDs using Oracle Crystal Ball software to select an endpoint for each species from the range of available endpoints, or from the confidence interval around an individual value, as one way to incorporate the uncertainty around individual data points into the SSD.
- Each SSD should be constructed with similar endpoints. That is, sublethal endpoints should not be mixed with survival (or immobility) endpoints; it is also not advisable to combine survival endpoints from different study times (i.e., 24 hr, 48 hrs, etc.). Likewise, it is not advisable to construct SSDs consisting of different sublethal effects. There is obvious biasing (either direction) when obvious sublethal effects (i.e., lethargy) are mixed with less obvious and less proved/replicated/relevant sublethal effects (PAM fluorescence for aquatic plants).
- SSD summary statistics do not provide units.

10.7 Characterization of Toxicity Data from Registrant Submitted Toxicity Data and Literature, and Utility of Sublethal Effects Data

Crucial to any risk assessment are the underlying data for the effects and exposure characterization, which are then used in the risk characterization and weight of evidence approaches. The data used must be robust in both its reliability and relevance to the risk assessment paradigm and its associated assessment endpoints, attributes and risk hypotheses. The Numerous papers have provided examples of such an evaluation process (Klimisch et al. 1997,⁶⁹ Durda and Preziosi 2000,⁷⁰ Hobbs et al. 2005,⁷¹ Schneider et al. 2009,⁷² Breton et al. 2009,⁷³ Hall et al., 2012,⁷⁴ Van Der Kraak et al. 2014,⁷⁵ Beasley et al.

⁶⁹ Klimisch H-J, Andreae M, Tillmann U. 1997. A systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data. *Regulatory Toxicology and Pharmacology* 25:1-5.

⁷⁰ Durda JL, Preziosi DV. 2000. Data quality evaluation of toxicological studies used to derive ecotoxicological benchmarks. *Human and Ecological Risk Assessment* 6:747-765.

⁷¹ Hobbs DA, Warne MSJ, Markich SJ. 2005. Evaluation of criteria used to assess the quality of aquatic toxicity data. *Integrated Environmental Assessment and Management* 1:174-180.

⁷² Schneider K, Schwarz M, Burkholder I, Kopp-Schneider A, Edler L, Kinsner-Ovaskainen A, Hartung T, Hoffmann S. 2009. ToxRTool, a new tool to assess the reliability of toxicological data. *Toxicology Letters (Amsterdam)* 189:138-144.

⁷³ Breton RL, Gilron G, Thompson R, Rodney S, Teed S. 2009. *A new quality assurance system for the evaluation of ecotoxicity studies submitted under the new substances notification regulations in Canada*. *Integrated Environmental Assessment and Management* 5:127-137.



2015,⁷⁶ Moermond et al. 2015⁷⁷). What is common in nearly all of these schemes is a thorough evaluation of study with a clear documentation on the evaluation of each study.

In general, there is concern for the lack of completeness and transparency in the evaluation of the data use in the risk assessment. Comments on the reliability and relevance of the underlying data in the assessment and how it can be used are found throughout this document. Some examples of sections where more information can be found are:

- Section 9: Documentation errors and technical correction – provide examples of where a more thorough review of the data is needed and where the relevance of the data needs to be better established or explained.
- Section 10.7: Use of SSD
- The following sections to 10.8 reviewing published literature, incident data, and sublethal effects.
- Section 10.9: Use of mortality effects thresholds based on chance effects – discussed the transparency needed in grouping of the data and establishing thresholds
- Section 10.11: Weight of evidence – discusses the importance of an *a priori* approach to the risk assessment paradigm and the weighing of the data.

10.7.1 Submitted and Published Literature Data

The results of data characterization as described in BE Appendix 1-4 and 2-3 cannot be evaluated. The utility of the data array builder is questionable because the open literature review summaries (Appendix 2-3) are incomplete and cannot be linked to data in the “array.” The data array builder is questionable because many studies have either not been reviewed or have not been included in the review summary and the reason for doing such is not explained. For studies included in the review and used quantitatively in the BEs, Appendix 2-3 contains the following warning (in capital letters), “some of the ‘quantitative’ studies still need to go through secondary review.” It is very concerning that endpoints used in the BEs have been selected from studies that have not undergone thorough review. Data from these studies should not be relied upon until the study results are evaluated.

⁷⁴ Hall, A.T., B.D. McGaughey and J.A. Gagne. 2012. *Data Quality, Reliability, and Relevance Standards for Ecological Risk Assessment: Recommendations for improvements to pesticide regulation in compliance with the Endangered Species Act*. In K.D. Racke, B.D. McGaughey, J.L. Cowles, A.T. Hall, S.H. Jacksons, J.J. Jenkins and J.J. Johnson (Eds), *Pesticide Regulation and the Endangered Species Act*. American Chemical Society, Washington DC.

⁷⁵ Van Der Kraak GJ, Hosmer AJ, Hanson ML, Kloas W, Solomon KR. 2014. Effects of atrazine in fish, amphibians, and reptiles: An analysis based on quantitative weight of evidence. *Critical Reviews in Toxicology* 44(S5):1-66.

⁷⁶ Beasley A, Belanger SE, Otter RR. 2015. Stepwise Information-Filtering Tool (SIFT): A method for using risk assessment metadata in a nontraditional way. *Environmental Toxicology and Chemistry* 34:1436-1442.

⁷⁷ Moermond C, Kase R, Korkaric M, Agerstrand M. 2015. CRED: Criteria for reporting and evaluating ecotoxicity data. *Environmental Toxicology and Chemistry* DOI:10.1002/etc.3259.



EPA states that it has followed its guidance on the use of open literature studies,⁷⁸ with modifications as described in Chapter 1, Attachment A-8. EPA's guidance on the web site (hereinafter referred to as Open Literature Guidance) states that a study may be considered appropriate for "quantitative" use if it meets following:

- a) is reported in, or can be converted to, environmentally-relevant exposure units;
- b) the endpoint is consistent with endpoints currently used to derive risk quotients (e.g. LC50 for acute exposure in fish);
- c) "sufficient information *must* (emphasis added) be provided in the open literature to substantiate and/or independently evaluate whether the study conclusions (i.e. dose-response) and endpoints are accurate", and
- d) the endpoint from the open literature study must be lower (i.e., more sensitive) than the lowest endpoint from a comparable registrant-submitted study for all taxon with the exception of aquatic-phase and terrestrial-phase amphibians, freshwater mollusks, reptiles, and for derivations of SSDs for other groups.

Chapter 1, Attachment A8 states that EPA used ECOTOX as a first screen. Given that the criteria for inclusion in ECOTOX are very minimal, this is appropriate for casting a wide net and as an initial screen.

The second level of screening should have been OPP's evaluation of the appropriateness of a study's use in the assessment, as described above. However, the only criterion that seemed to be retained was item (a), regarding the units. Data with environmentally-relevant exposure units were used to build the arrays; the other criteria were not applied. This means that studies with very low reliability and relevance can become part of the arrays, essentially rendering the arrays useless. Uncertainty is supposed to be included in the effects arrays (Chapter 1, page-51 in the chlorpyrifos BE) – but how uncertainty is used is not discussed. While the concept of arrays, to allow presentation of a lot of data visually, is appealing, by including essentially all open literature without critical evaluation, EPA has presented a meaningless compilation of data.

Returning to discussion of the open literature studies that EPA used for development of thresholds, it is stated that these were "intensively reviewed" and only the study providing the lowest effect information was reviewed in detail. It is further stated in Chapter 1, Attachment A8 that the studies used for thresholds had to meet the definition of "quantitative" as described in the Open Literature Guidance. This means they would have had to meet the criteria given above in "b", "c" and "d." For plants, if data were insufficient to construct an SSD (as was the case for the OPs, and as mentioned

⁷⁸http://www.epa.gov/pesticides/science/efed/policy_guidance/team_authors/endangered_species_reregistration_workgroup/esa_evaluation_open_literature.htm



previously SSDs were not even considered for plants), only studies providing lower effect levels than registrant-submitted guideline studies were even considered (criterion “d”). This ignores potentially high quality plant effects data that could constructively inform the assessment, because at best, they would only be shown in the arrays (along with data of questionable quality). Moreover, EPA does not seem to have followed criteria “b” and “c” as illustrated by the examples provided in Section 9.0 (Document Errors and Technical Corrections, review of BE section 4.4).

The presented data arrays contain a wide range of endpoint and ECx values, in absence of a discussion of statistical quality of these extrapolated values (especially those at the low end or beyond the power of the data). This type of evaluation is paramount to subsequent use of these data, regarding the relevance and reliability of the data and thus the assessment as a whole. In fact, the data arrays do not provide an association between the actual data used for assessment and what is plotted in the array. Units, source and other critical items of data integrity are often missing from figures and tables.

Effects data considered in the BEs, or any risk assessment, should be relevant to the assessed organism(s) and robust (reliable). Therefore, effects data should only be considered if: (1) studies are on test organisms that are relevant to the organism(s) being assessed; (2) it is representative of effects from environmentally relevant exposure routes and durations; (3) data are produced from a suitable test system with appropriate test conditions; (4) results are reproducible; and (5) the study produces endpoints that can be quantitatively or qualitatively linked to survival, growth, or reproduction. Not all effects data considered in the draft BEs meet these criteria. As stated on pp. 2 of Attachment 1-4, *“Endpoints generally are: a) from in vivo studies that are conducted with whole organisms; b) representative of environmentally relevant exposures routes, and c) able to be quantitatively or qualitatively linked to effects on survival, growth or reproduction.”* However, some endpoints (e.g., plasma acetylcholine levels) are not linked to toxicological effects, yet are still considered as an adverse effect in the evaluations. Effects data should be re-evaluated, with additional scrutiny, for its relevance and utility in the final BEs.

Page 4 of Attachment 1-4 states, *“For deriving a threshold based on sublethal direct effects to listed animals, the lowest available NOAEC or NOAEL or other scientifically defensible effect threshold (ECx) that can be quantitatively or qualitatively linked to survival or reproduction of a listed individual is used,”* however, the draft BEs rely on sublethal endpoints such as plasma cholinesterase which have no link to growth, survival and reproduction or toxicological relevance in general.

When assessing the indirect effects to endangered birds and mammals by compound related mortality of prey items, the use of an LC50 is more appropriate than the LD50. Using the LC50 does not require the toxicity thresholds to be scaled to different size classes to accommodate a unit-to-body weight adjustment, nor is it then necessary to add the assumption that uptake and absorption kinetics of a gavage toxicity study approximate the absorption associated with uptake from a dietary matrix. Further, the assessment should more accurately portray individual species diets by assigning prey to the



appropriate size class based on the best available species information. Using only the default assumption that a 20g bird or 15g mammal feeds only on treated short grass will grossly overstate assumed exposure when compared to exposures calculated from actual dietary composition.

10.7.2 Use of Incident Data

The use and interpretation of incident data is highly speculative, misleading, inflammatory and without adequate context. It also does not follow the policy document it references.⁷⁹ Various in the three assessments, EFED mentions that some incidents, while strongly linked to compound uses, were clear cases of illegal application or undetermined sources or rates. Yet, an incident like this was used to support a risk assumption as if it were a labeled, legal use. Illegal uses are not supported by the label and therefore not appropriate for use objectively or subjectively in risk characterization. Any use of incident data needs to provide context as to the relevance of the reported event to current labeled uses and the likelihood that it was from a labeled, legal application, or if not, what kind of evidence it supplies should be described.

Foreign incident reports should not be used to supplement the available U.S. incidents. Foreign incidents may not represent U.S. agronomic practices, products, uses, or exposures given differences in environmental conditions and therefore do not benefit the characterization of potential effects to species in the U.S.

10.7.3 Sublethal Effects Data - Animals

Sublethal endpoints in the BEs should be limited to effects with a direct link to survival, growth, or reproduction and should exclude effects that are markers for exposure with no direct link to apical adverse effects (e.g., plasma cholinesterase inhibition). Endpoints should be weighted with direct measures of survival, growth, and reproduction, receiving the highest weight depending, in part, on the level of biological organization at which the endpoint was measured. For non-apical sublethal endpoints that are part of an Adverse Outcome Pathway ("AOP") (e.g., acetylcholinesterase inhibition), weighting should in part be influenced by the number of key events identified in the study.

In the BEs, available effects data for cholinesterase inhibition are considered in the mortality and behavioral lines of evidence and in some cases may be used to define the lowest effects threshold in the assessment. Chapter 2 explains the use of cholinesterase inhibition data in this way is justified based on the AChE AOP. AChE AOPs include the key event of acetylcholine accumulation at the neuromuscular junctions of skeletal muscle or synapses in cardiac tissue. Some endpoints considered in the BEs, including plasma cholinesterase inhibition, do not inform whether this key event has been triggered. Therefore, plasma cholinesterase inhibition and other values not directly related to an event should not

⁷⁹ Policy Memo from Don Brady to Managers and Staff of EFED. Guidance for Using Incident Data in Evaluating Listed and Non-listed Species under Registration Review. October 13, 2011



be used in the assessment as values that trigger an “effect.”

The uncertainty regarding the ability to accurately predict apical outcomes from biochemical or behavioral measures must be considered based on the current quantitative understanding, and the reliance of these endpoints in the BEs must be weighted based on the uncertainties associated with them.

Controlled laboratory studies and the conditions under which they are conducted influence the outcome of a study. Confinement under laboratory conditions may alter the behavior, physical condition, and even biochemical levels in some species. The potential impact of test conditions must be intensely scrutinized when evaluating studies for sublethal effects such as predatory behavior, because captivity itself can bring on stress that results in exaggerated behavior. Thus, any stress-related indications should be validated in the field or considered in light of the stress presented by the testing environment.

10.7.4 Sublethal Effects Data – Plants

Figure 1-8.1 in Chapter 1, Attachment A8 shows a flow diagram that was used on the studies appearing in ECOTOX. This figure echoes the AOP concept, which is an attractive unifying concept but has not been worked out for many chemicals. Interestingly, the AOP is known for the organophosphate pesticides, and based on the AOP, effects on plants would be eliminated from further consideration as plants do not have acetylcholine. However, plants were not eliminated. More generically, it is not clear just how effects at the biochemical, cellular, or organ levels are to be demonstrated to be relatable to survival or reproduction, and therefore allowed to be considered further for development of thresholds.

10.7.5 Interpretation of “Sublethal” in Context of the Assessment

Sublethal effects, as recommended by the NAS Panel, will most likely require the subjective interpretation of whether these types of effects are possible and detectable in the natural environment. However, EFED instead attempted to make this a quantitative evaluation regardless of the fact that no context for environmental expression of an indirect effect actually has potential to adversely impact an individual or the population of organisms. For example, EFED used the lowest toxicity values for behavioral impacts with no reference as to how these translate to actual whole organisms (individual effects). Even if these were translated into whole organisms and then then species to populations, they would not be expected to be representative of all species and all settings. A relationship between behavioral effects across a range of exposure values which vary with time is also not established. The Interim Approach states that relevance of sublethal effects (e.g. behavior) needs to relate to the impacts on individual survival and reproduction. Making an association between sublethal effects and impacts on growth, survival and reproduction, if such a link can be made, clearly does allow for the assessment



of these sublethal effects at the population level. If the link cannot clearly be made, then the mere existence of a sublethal effect is not evidence of a likely effect on species survival and reproduction.

10.8 Use of Mortality Effects Thresholds Based on Chance Effects

10.8.1 General Issues

This section discusses using separate thresholds from studies conducted with the technical grade and formulated product when available. Toxicity values are used from the lowest toxicity threshold if they are from studies with the formulated product. What is missing is a discussion of the relevance of the formulated products from an environmental fate and relevance perspective.

Regarding indirect effects thresholds for listed species that consume fish and aquatic-phase amphibians, there is no practical relationship stated or justified between the thresholds used and their actual relationship on indirect effects. Indirect effects thresholds should be stated.

There is no justification for the conservatism in the threshold for indirect effects based on mortality. The current approach sets the threshold at the concentrations/dose estimated to cause 10% effect on to the most sensitive species, or 5th percentile most sensitive depending on data availability, of the taxa that are relevant to the specific species being assessed. This value implies that the assessed species fitness depends on 90% survival of the most sensitive relevant species, which infers an obligate relationship between the assessed species and the relevant species which is not the purpose of this threshold. Further, this approach neither considers the duration of effects and recovery potential of the relevant species nor the diversity of the diet or habitat of the assessed species.

10.8.2 Summary Data Arrays for Fish and Aquatic-Phase Amphibians

Inter-laboratory variability in effects data can result in a range of responses for the same species tested. These differences can be accounted for by using a geometric mean of the values from the distribution.

The toxicity data for each taxon are generally presented as summary data arrays developed using the Data Array Builder v.1.0. This array does little more than illustrate the range of effects responses. It would be appropriate to use this range in sensitivities to characterize the range of risk across listed species of interest and not assume worst-case values are representative (e.g. class, order, family) associations.

Data in these arrays are grouped by the type of effect (e.g., behavior, reproduction, mortality), and present the range of LOAECs and NOAECs (NOAECs must have a corresponding LOAEC to be represented in array) for each effect type. The relevance of behavior responses are not related to whole organism responses and thus species and population level responses. Further, clarifying the distribution of effects



is necessary to later understanding of the impacts at the population level. It is not clear if this detail will be appropriately translated for use in STEP 3.

The mortality thresholds for freshwater and estuarine/marine fish are based on the 1-in-a million of the HC₀₅ based on the Species Sensitivity Distribution (SSD). Again, compiling data into an SSD and calculated HC₀₅ causes the resolution of the species-to-species level comparisons to be lost. The overly conservative selection of thresholds is especially problematic when there is a wide range of sensitivities across taxonomic groups. Instead, the SSD should be utilized to assign listed species to those data on the distribution that are as closely taxonomically related as possible, then threshold exposure values could provide a mechanism to group species for further analysis but not to end the assessment and “risk conclusion” at that point. For example, in fish and invertebrates there is an extensive range in species sensitivity across species in each taxonomic group. While the range is illustrated in data arrays, it is not effectively used in the risk characterization. For example, it is stated that chlorpyrifos is highly toxic to aquatic invertebrates, but that less sensitivity is exhibited in the mollusk group. While this acknowledges the fact that sensitivity ranges are different, this fact is not applied to the overall analysis of aquatic invertebrates.

Using the lowest value from the EFED compiled distribution for indirect effects builds unrealistic conservatism into the assessment. For example, this assumption for amphibian food items would mean that the amphibian prey base is interpreted as being uniformly sensitive at the lowest concentrations in the distribution curve, resulting in a much more significant impact to prey base than if only certain prey organisms are sensitive (which is normally the case). This incorrect application of data needs to be corrected in the risk characterization.

There is much discussion regarding impacts to things such as AChE inhibition, effects to sensory systems, and impacts on behavior. This include references to BCM=Biochemical, CEL=Cellular, PHY=Physiological, BEH=Behavioral and POP=Population responses. The linkage of these responses to survival and reproduction are not established, and it is unclear how population responses are actually determined or translated to species populations. This appears to be mostly a database exercise without characterization.

10.9 Weight of Evidence Approach Used, including the High, Medium, and Low Weighting Assignments to the Various Lines of Evidence to Evaluate Risk and Make Effects Determinations

Consistently throughout the assessments, it is stated that data are used for weight-of-evidence approach. However, this approach is no more than auto-generating data arrays which compile the range of sensitivities reported in available and often unqualified studies. The “weight” is limited to again automatically extracting the lowest value from the lowest end of the data array – in many cases this value either cannot be linked to its original source, or can be demonstrated to be from a source not



validated, not relevant, or not reliable. It appears that the weight-of-evidence approach merely shows the range of potential values from various endpoints, with no scientifically defensible, robust determination of what the lowest value actually means or how it was derived.

10.9.1 A Solid Approach to Considering the Weight of the Evidence Begins with a Well Designed Problem Formulation

The overall structure of the assessment, with multiple files, makes it extremely difficult to evaluate and exactly understand what data, input, and assumptions were used in the evaluations. One coherent formulating document is needed. Usage information should be given greater weight for compounds such as these, which have been on the market for so long. The usage maps presented by EPA do not appear to show wide spread uses; in fact great stretches of the county appear to have minimal use, which does not support the Agency's assertion that the whole county is the action area. While clearly available to the Agency, the best commercial data was not used. This coupled with the fact that very different uses were combined into one "use classification" further confounds the problem formulation. The lumping of all uses together seems to be counterproductive – if there is a no overlap between a specific use and a listed species, then registration of that use should be considered to have no effect on listed species.

Furthermore, there is nothing in the problem formulation that considers mitigations already in place. Before the current "Bulletins Live" process, several hundred county bulletins had been developed by EPA in a voluntary endangered species protection program, and these carried mitigations that evolved through early OPP-Services consultation as well as subsequent interactions with field biologists and local knowledgeable experts. To completely disregard these in the assessment as existing tools that can be used to mitigate exposure fails to take advantage of much work and communication that resulted in an accumulation of knowledge that is still useful.

The risk hypotheses as currently proposed (section 1.3.2) are not set up to address the key factor/question in determining the risk to listed species. As written, there is an inherent assumption that exposure will occur, when one of the first questions that should be asked is whether or not there is a likelihood of exposure. It is only after it is determined that there is reasonable likelihood of exposure that the magnitude of exposure and potential effects should be addressed. Analysis Plan – 3 step approach, figure 1-5.

It is important to realize that the 3 step approach is not an analysis plan or risk assessment scheme, it is simply a schematic of questions that must be answered under the current regulatory scheme to meet the mandates of ESA. It is appropriate in describing the regulatory process, but is not particularly helpful in terms of planning the details of the risk assessment. The schematic also implies that risk assessment is separate for each question being answered, which is inefficient and counterproductive. In fact, the figure is opposite what the process really should be. The three "steps" or regulatory questions



should feed into a single problem formulation from which an analysis plan would then be developed to address the protection goals. A tiered risk assessment process, without the artificial limits currently being imposed, could then be applied to efficiently address all assessment parameters.

In setting up the evaluation of weight of the evidence, lines of evidence should be independent, relevant to the nature of the data under evaluation and transparent. However, the lines described are artificially inflated and for most part are not independent of each other, and therefore a weighting scheme should be defined accordingly.

10.9.2 EPA's Current Method on Weight of the Evidence Ignores Scientifically Established Approaches to this Exercise

Use of weighting in incorporating lines of evidence into risk assessment is not explicitly recommended in the 1998 Guidelines, but the three categories suggested for consideration when evaluating lines of evidence can be implemented in a weighting scheme. These categories are: (1) adequacy and quality of data, (2) degree and type of uncertainty associated with the evidence, and (3) relationship of the evidence to the risk assessment questions. There is a substantial literature on the use of weight of evidence in risk assessment, yet the BE authors referenced none of these published papers, and therefore there is no means of understanding how the apparently novel approaches in the BE relate to prior work meeting the standards of scientific peer review. Concern over addressing adequacy and quality of data (1) is mentioned above. Uncertainty (2) is expressed throughout the BEs as qualitative confidence levels of low, medium, and high. It is not clear how these levels were assigned or the method used to insure consistency of application. In other sections of these comments we point out the difficulties with the confusion between toxicity thresholds, assessment endpoints, risk hypotheses, and lines of evidence. This confusion certainly complicates item (3) in the weighting scheme. The flaws in the WOE method, and solutions to them follow.

10.9.2.1 No References to the WOE Literature (Chapter 1, Attachment 1-9)

Although WOE methods figure prominently in the draft Biological Evaluation (BE), there are no references to the existing – and not insubstantial – WOE literature in either Chapter 1 or Attachment 1-9. Thus there is no way to easily determine whether the WOE method outlined in these locations builds on (or is even cognizant of) prior research or whether it is a *de novo* approach unique to this evaluation. The latter appears to be the case here and it seems ill-advised to have embarked on a significant WOE analysis for a major biological evaluation without first having consulted the literature.

10.9.2.2 Assessment Endpoints Not Clearly Defined (Chapter 1)

It is more likely that the purpose of a WOE will be achieved when it is deployed within the context of EPA's existing – and now long established – process for conducting useful ecological risk assessments (USEPA, 1998). A key first step in this process is, as stated above, Problem Formulation, and a key



component within that is the specification of assessment endpoints. An assessment endpoint is an expression of the environmental value (species, ecological resource, or habitat type) that is to be protected. Useful assessment endpoints are defined by both the valued ecological entity (or resource) and an attribute of that entity (or resource) to protect. Although Chapter 1 of the BE addresses some aspects of Problem Formulation (e.g., the conceptual models and risk hypotheses) it does not clearly define (or even much discuss) the assessment endpoints on which these models and hypotheses depend. In Section 1.3.2, for example, statements such as “...on species and assessment endpoints...” and “...for each species and designated critical habitat...” incorrectly convey the idea that species or critical habitat (i.e., entities) are somehow different than assessment endpoints. They are not and this misunderstanding could contribute to disconnects within the WOE analysis. The large number of assessment endpoints inherent in this BE have to be inferred by parsing through Figures 1-3 and 1-4, whereas they should have been explicitly stated in Chapter 1. Having this many assessment endpoints, several of which would be hard to evaluate with existing or readily obtainable information, works against the usefulness of a WOE analysis. It would have been better to use the Planning & Scoping and Problem Formulation phases to drill down to a selected and limited number of assessment endpoints whose analysis would provide decision makers with focused and actionable information.

10.9.2.3 Lack of Consistency in Terminology (Chapter 1, Attachment 1-9)

The captions for Figures 1-3 and 1-4 alter the specification of attributes shown in the conceptual models (“For ‘Attribute Change’, the terms ‘Individual Organisms’, ‘Food Chain’, and ‘Habitat Integrity’ are equivalent to ‘Individual Fitness’, ‘Indirect Effects Related to Prey’, and ‘Indirect Effects Related to Physical and Biological Features of Habitat’, respectively.”). It is these altered specifications, and not those in the figures, that appear – highly abbreviated – in the risk hypotheses in Section 1.3.2. The risk hypotheses also introduce critical habitat as an entity, direct and indirect effects, and the terms PCE and PBF. These latter two terms seem consistent with the existing definition of “measures [of ecosystem and receptor] characteristics.” This is not just quibbling over semantics. If one of the purposes of a WOE analysis is to bring clarity and consistency to an assessment, then not clearly defining terms, changing terms *ad libitum*, introducing old terms as new terms, or allowing seeming inconsistencies between terms used in the conceptual model and those used in the risk hypotheses only detracts from accomplishment of this purpose.

10.9.2.4 Confusion of Assessment Attributes with Lines of Evidence (LOEs) (Chapter 1, Attachment 1-9)

The assessment endpoints, and not the risk hypotheses, should be the first link to the protection goals (i.e., management goals) for Step 2. Then the assessment attribute part of the assessment endpoint is linked to one or more risk hypotheses and then each hypothesis is linked to at least one LOE (here one based on toxicity - exposure and effect). Here, however, assessment attributes are incorrectly identified as LOEs - which would result in weights being applied to an attribute and not to the evidence. For example (Attachment 1-9, page A9 (PF)-3, Bullets A & B), survival (the inverse of mortality) is an



attribute, not an LOE. The LOEs refer to whether that attribute will be affected and, if so, to what extent; for example:

Assessment endpoint: Assessment entity + its assessment attribute(s).

Assessment entity: An individual of Listed Species X.

Assessment attributes (for X): Survival, growth, reproduction, behavior, sensory function, prey items, pollinators/diaspore dispersal vectors, plant habitat, obligate organisms. Note that each attribute need not (and probably should not) be weighted the same relative to the other attributes.

Risk hypothesis: Exposure to Pesticide Y used according to registered labels (includes parent active ingredient, formulations, and degradates of concern) will adversely affect one or more of the attributes related to X.

Line-of-Evidence: (1) Level of exposure and (2) effect on attribute associated with that level of exposure (i.e., the exposure/response function). Evidence for characteristics might also be included here (as is alluded to by Bullet C).

A distinction has been made between the effects of direct exposures (Bullet A - the individual comes into contact with the pesticide) and those from indirect exposures (Bullet B - something the individual needs - food, prey, habitat - comes into contact with the pesticide). This is a useful distinction but the evaluation of indirect effects on populations of prey or food species or habitat organisms is more difficult and uncertain than evaluation of direct effects on individuals. In addition, loss of individuals within a population of prey or food species – unless it is great enough to extirpate local populations of those species – should not be accorded the same level of concern (i.e., weight) as the loss of an individual listed species.

Bullet C implies that there could also be evidence for a suite of characteristics (factors that may influence or confound the behavior and location of an assessment entity, the distribution of a stressor, as well as the life history characteristics of the entity or its surrogate that may affect its exposure or response to a stressor). Again, a useful thought, but one that is much harder to evaluate reliably than direct effects. Thus, the information described in bullets B and C should be subjectively evaluated, as recommended by the NAS Panel, rather than given an objective “weight” and added to an actual WOE exercise. Finally, exactly how the LOEs implied by Bullet C were implemented is not clearly discussed in the text.

10.9.2.5 Unclear Ranking of Exposure LOE Criteria

The only LOE for exposure is the estimated environmental concentration (EEC), presumably based on predictive modeling (which model or models is not stated explicitly). The text states that two criteria (“relevance” and “robustness”) were used to evaluate this LOE. Aside from the fact that the terms “relevance” and “robustness” have been applied contrary to mainline scientific approaches (see related



discussion of relevance and robustness in other portions of these comments), there are at least four or five, but not just two, criteria:

- 1) Model developed to predict exposure in relevant habitat;
- 2) Consistency with the species habitat (the text on page A (PF)-10 speaks of the conceptual model for the exposure scenario being consistent with the habitat; however, it's not clear whether this is the same criterion as (1), an additional criterion, or a typo (predictive, not conceptual));
- 3) Exposure scenarios representative of pesticide use patterns;
- 4) Extent of chemical-specific fate data (per Appendix 4-3, this criterion includes chemical specific foliar dissipation half-life, additional exposure routes, species-specific Mineau scaling factor, AgDRIFT deposition – all of which could be treated as separate criteria);
- 5) Exposure results similar across “lines of information” (e.g., field-scale monitoring data, other model predictions) - which raises the question of why these other data and predictions were not treated as criteria in their own right?

These criteria are weighted presumably with respect to a given assessment attribute for a given assessment entity – but this is not stated and may not be understood (see comments above about what constitutes an assessment endpoint). The text on page A (PF)-10 also concatenates criteria (3) and (4) into one conditionally interlocking criterion, whereas it would have been clearer to have kept them separate.

Although the text and table on page A (PF)-10 establish high (H), medium (M), and low (L) weights for these criteria, and connects these criteria to the confidence to be placed in the risk estimate, weights for each criterion do not appear in Appendices 3-6 or 4-3. Instead, the entire exposure analysis appears to be only a lengthy narrative, with no explicit weights given for individual exposure criteria, and no clear connection between a given narration and a weight. Having established criteria for and presumably weights to the exposure LOE, it seems incumbent on the analyst to provide a summary of these weights by criterion by attribute somewhere in the BE – in Chapter 3, for example – as was done for risk in Chapter 4. Ideally, each of the 4 or 5 criteria associated with exposure would be weighted individually and then those weights combined to establish the overall weight for the exposure LOE for a given assessment attribute. This would produce an objective, defensible and repeatable WOE analysis. This would also enable readers to determine the strengths and weaknesses for a given criterion for a given attribute. If these 4 or 5 criterion are, in fact, weighted somewhere within the BE where those weights might appear it is not at all obvious.

10.9.2.6 Unclear Ranking of Effect LOE Criteria

The only LOE for effect appears to be the results from toxicity studies (EC25, EC50, LC50, LOAEC, SSD, etc.). This is not evident from the discussion of the effect LOE criteria but becomes so once the criteria



for ranking risk (page A9 (PF)-9) are examined. Within this context (and also looking at Attachment 4-3), there appear to be five criteria.

- 1) Established relationship between measure of effect (a toxicity endpoint) and assessment attribute (not assessment endpoint as stated in Attachment 1-9) [Potential attributes include not only survival, growth, reproduction, behavior, and sensory function, but also AOP, brain AChE activity, acceleration, stamina, and olfactory recordings. Note that this list includes attributes found here and not elsewhere in the text.];
- 2) Representativeness of test organisms (surrogates) to listed species or critical habitat, based on phylogenetic closeness (same order) or similar life history or physiology;
- 3) Number of toxicity studies potentially applicable to a given listed species or critical habitat;
- 4) Consistency within toxicity studies potentially applicable to a given listed species or critical habitat;
- 5) Availability of field studies.

Here, again, weights for each of these criteria do not appear in Appendices 3-6 or 4-3. Instead, the entire effects analysis appears to be only a lengthy narrative, with no explicit rankings of these effect criteria, and no clear connection between a given narration and a weight. Having established criteria and presumably weights for the effect LOE, it seems incumbent on the analyst to provide a summary of these weights by criterion by attribute somewhere in the BE – in Chapter 4, for example. Again, ideally, each of the 5 criteria associated with effects would be weighted individually and then those weights combined to establish the overall weight for the effect LOE for a given assessment attribute.

10.9.2.7 Risk Estimate Based on Just One LOE

On page A9 (PF)-6, it is stated that risk is established by comparing the overlap of exposure and effect levels – with “consideration” given to the degree of overlap. Overlap could imply that distributions of exposure and effect are being compared to derive a probabilistic risk estimate (there are SSDs in use here but is unclear whether the probabilities this method is capable of generating were used in the WOE analysis). However, on page A9 (PF)-9, it becomes clear that here “risk” is determined simply by where exposure is relative to a variety of toxicity thresholds, with consideration of the overlap limited to: L if below the lowest threshold, H if above a threshold where effects were observed, and M in between. So, in short, this entire WOE analysis comes down to just one LOE (for toxicity), with its “risk” estimate derived solely as a point estimate of exposure relative to a point estimate of effect (i.e., with a quotient methodology). The primary concern here, which has been a long acknowledged weakness in the quotient methodology, is that a threshold exceedance (i.e., a quotient >1) should not be automatically interpreted as a “high” probability of adversity. This is based on the lack of any fixed relationship between a quotient value and the probability of an adverse effect. Inferring such a probability with just a threshold is difficult at best, if not impossible. This is one key feature that argues against the use of a quotient methodology to generate a “risk” estimate in a WOE analysis. Although thresholds (quotients)



have become an established and comfortable approach to regulatory decision making, they are not risk and are not informative when one's desire is to base a decision on the probability of a future outcome – or to even know how large or small that probability might be. Decision makers with no, or only heavily biased, insights into such chances may be encouraged or required to incur or impose significant direct and indirect costs to avoid an outcome that has no real chance of ever materializing. They may also choose a risk management alternative that limits or precludes the possibility of beneficial outcomes. Having said this, however, a numerical relationship (“quotient”) as a flagging criterion – but not as a basis for final decisions on risk – is a reasonable way for policy to guide a tiered risk assessment process as long as a quotient of >1 is defined as the need to apply certain additional data and processes to a refined risk assessment exercise.

The explanatory text in Attachment 4-1 for the summary sheets (e.g., Figure A 4-1.20) incorrectly identifies assessment attributes (e.g., mortality, etc.) as LOEs, thus giving the impression that multiple, different LOEs were used to assess risk. As noted above, there is just one LOE (for toxicity) being applied to several different attributes for a given assessment entity. “Chemical stressors” is a variation on toxicity involving environmental mixtures or the parent a.i. alone. “Abiotic stressors” are measures of characteristics (such as temperature, bacterial/viral prevalence, or environmental pH) that could enhance an assessment entity’s susceptibility to the toxicity of the parent compound. Thus this WOE analysis comes down to just the basic exposure vs. toxicity comparison approach expanded to include a larger number of assessment entities and their attributes. While there may be useful information about toxicity to be gleaned from such an expansion, it does not constitute a WOE analysis that weighs a number of truly different types of evidence (i.e., that beyond just exposure/toxicity) to support conclusions about whether a pesticide may or may not pose a threat to a given assessment entity.

It is also not clear whether these summary sheets (e.g., Figure A 4-1.20) were intended to be the risk vs. confidence summary graphs like the one shown in Figure A 1-9.2, because what is shown in this figure could not be found in the main text.

10.9.2.8 Unclear Derivation of Confidence in the Risk Estimate

A table on page A9 (PF)-10 shows how confidence in the risk estimate is assessed on the basis of weights assigned to the LOEs for exposure and effect. Here, all of the criteria applicable to either exposure or effect are reduced to a single H, M, or L weight. Exactly how this was accomplished on a consistent basis is unclear. And (as was commented on above) while there are criteria and overall weights (ranks) for both the exposure and effects LOEs, there appears to be no summary of these rankings by individual criterion and attribute anywhere in the BE. Ideally, each of the criteria associated with both exposure and effect would be weighted individually and then those weights combined to establish the overall weight (rank) for the toxicity LOE for a given assessment attribute for a given assessment entity. The long narratives in Attachment 4-3 could serve as backup for these individual weights rather than (as they do now) the sole basis for these weights. What is also not clear here is exactly how this determination



of confidence “...may also consider other factors...” While considerations evaluated subjectively or with best professional judgment are acceptable, the introduction of a superficial, subjective conclusion called a “WOE determination” is incorrect. There needs to be greater clarity as to when and how this was done.

10.9.2.9 Effects Determination

There are policy considerations in how a risk/confidence estimate relates to an effects determination (per page A9 (PF)-8) that are beyond the scope of these comments. However, there are three technical aspects of the Effects Determination table (Table A 1-9.2) that merit comment:

1. It confusingly suggests that the risk estimate applies to a LOE. Instead, what is presented is a risk estimate and the assessor’s confidence in it (both based only on a LOE for toxicity (exposure and effect)) for an attribute of an entity.
2. Per Figure A 4-1.20 in Attachment 4-1, there could be as many as 10 attributes (if chemical and abiotic stressors are included as attributes rather than as factors) for any given assessment entity. An effects determination is being made on the basis of just one of these. If, to posit an extreme example, the risk/confidence for one attribute were weighted H/H and that for all the rest were weighted L/L, you’d still conclude LAA. Thus you’d reach an LAA determination by ignoring 90% of the evidence. Policy considerations aside, ignoring evidence seems to contravene the purpose of a WOE analysis, even one based on one LOE (for toxicity). Since 90% of the evidence seems to be ignored, or subsumed by aggregation, and the conclusion that over 90% of the species are at risk from the use of OPs is obviously an indication of how the assessment has been adversely impacted by the methods used.
3. As noted before, it is probably not safe to assume that each attribute is of equal weight with respect to an LAA determination. For example, it may be possible to assess a mortality attribute with greater certainty or less equivocation than a sensory or behavioral attribute. So to continue the example in (2) above, if the attribute whose risk/confidence is weighted H/H is the weakest (lowest weighted) of the 10 attributes, then not only is the assessment making an LAA determination with 10% of the available evidence but also with the weakest part of that evidence. This too seems to contravene the intent of a WOE analysis.

10.9.3 Improvements Needed

CLA provides extensive comments on the risk characterization section (“effects determination”) of the assessment. We preface our comments, however, by pointing out that there is a serious lack of transparency in the section, and it is very difficult to piece together what was done. For example, there is lack of information on how “uncertainty” was defined and no references to the literature on weight of evidence approaches. These and other examples are discussed in more detail below.



The main deficiency in the lines of evidence used in the BEs is the reality that, in effect, only a single line of evidence is brought into the analysis: toxicity. As pointed out above, this is made clear on page A9 (PF)-9, where risk is determined only by comparison of a point estimate of exposure to a variety of toxicity thresholds. The analysis is therefore confounded by identifying these toxicity thresholds with attribute changes relevant to risk hypotheses for which lines of evidence are considered and evaluated (1998 Guidelines for Ecological Risk Assessment, EPA/630/R-95/002F [1998 Guidelines]). These include (1) mortality, (2) reduced growth, and (3) reduced reproduction (see, for example, chlorpyrifos Chapter 1 Table 1-4 and the related Figures 1-3 and 1-4). CLA points out that in prior published examples following a generally accepted process, mortality from direct acute (short-term) exposure is evaluated by bringing into the analysis plan individual lines of evidence of exposure (modeled, monitored, evaluated in field studies) paired with exposure/response information such as dose-response results (from laboratory, semi-field, and field studies) to determine the probability of mortality. The BEs confuse attributes with lines of evidence. The lines of evidence are estimates of whether the attribute(s) will change, not the other way around. Such confusion distorts the risk assessment process and represents a radical departure from the recommendations of the 1998 Guidelines. Moreover, the remaining "lines of evidence" given in Table 1-4 represent either: a) attributes that relate to (1), mortality, (2), reduced growth, or (3), reduced reproduction (which includes EFED's item (4), impaired behavior and (5), impaired sensory function); b) independent risk hypotheses (which includes EFED's item (6), indirect effects and (7), mixture toxicity); or c) ecosystem characteristics (which includes EFED's item (8), factors independent of an OP insecticide). Of these, only item (8) can be considered amenable to evaluation with a toxicity line of evidence (exposure + effect). For this case, any example of (8) should be evaluated in a specific pairing with a line of evidence including exposure and exposure/response to determine how either is mediated by the relevant ecosystem characteristic. Due to these radical departures from the recommended risk assessment process it is difficult for a reviewer to follow the logic of the analysis plan and judge the validity of the overall BE risk characterization.

Temporal scale is particularly important for interpreting ecological adverse outcome, and since the OP insecticides assessed in the BEs have been used for more than 50 years, the entire approach of the predictive analysis (that assumes this Agency action involves release of a new toxic chemical into the environment) is conceptually unsound. If one considers a listed species to be challenged by multiple stressors limiting recovery, and a specific chemical stressor (for example, chlorpyrifos) is evaluated in a predictive assessment for its marginal contribution to possible extirpation and found to greatly impact the fitness of individuals of populations for 99.9% of the species, then it is reasonable to ask the question why this predicted mass extirpation has not occurred after decades of potential exposure. This question is not found in the BE analysis but should be considered.

Thus, the entire analysis is undermined by an unrecognized conflation of causal and predictive assessment that fails to acknowledge an environmental baseline incorporating many decades of stressor presence in the environment. Consequently, there are lines of evidence not pointing toward the same conclusion such as water quality monitoring for chemical stressors vs. exposure modeling. In addition,



whole lines of evidence are ignored: population recovery surveys taking into account natural fluctuations can be compared to predictions from effects and population models. The case of Kirtland's warbler provides such a recovery survey, which, if included in the analysis, would not point to the same conclusion as was supported by the predictions of flawed effects and population modeling.

Review of the BEs is greatly hampered by lack of understanding of the methods used for reviewing, analyzing, and summarizing the uncertainty in the "lines of evidence" supporting the risk assessment. What is missing is clarity as to the methods used to pick exposure and effect (components of a toxicity line of evidence) criteria, weight them, and combine them. A further deficiency is non-transparency in evaluation of data quality and relevance. The documents do not appear to provide an explicit systematic assignment of quality or relevance to data sources crucial to the risk characterization. Therefore, in many cases it is not possible to judge the validity of lines of evidence contributing to risk conclusions.

The line of evidence related to the possibility of exposure in specific locations suffers from a large overstatement in spatial scale of potential use sites, and this leads to overprediction of ecological adversity for both direct and indirect effects. It is difficult to accept the premise that crops not on product labels contribute to potential exposure in specific growing regions. Likewise, it is highly unlikely that wide-area uses are actually used everywhere. Instead of assuming this is the case, additional lines of evidence relevant to species and use setting – such as historical mosquito adulticide application areas – should have been considered.

10.9.4 High, Medium and Low Weighting Assignments

The **weight of evidence** approach is incomplete. The assessment identifies multiple lines of evidence, each of which has an associated qualitatively defined "high", "medium", and "low" risk value and a similar qualitative "certainty" value. However, each of the lines of evidence is then considered on its own, and there is no aggregate weighting of ALL the lines to get a final "weight of evidence" for the species/crop/chemical combination (see: Hope BK, Clarkson JR. 2014. A strategy for using weight-of-evidence methods in ecological risk assessments. HERA 20:290-315). In fact, if at least one of the lines of evidence has a "high risk" or "medium risk" estimation, regardless of the confidence in this result, then the species is given a "LAA" determination to be carried forward to Step 3.

The graphic that shows the relative placement of all the lines of evidence in the risk-uncertainty space is a good start. One should then give scores on a scale of 1-5 (or 1-10; or whatever is preferred) for each line. The scores are then added up and divided by the total number of lines to give an average score. This average of all the scores determines the overall conclusion for that species/crop/chemical combination. A qualitative decision is made about how high a score is required to move a species/crop/chemical to the next Step (i.e., to assume a LAA).



Example 1:

- Line 1: High Risk/Low confidence = 3
- Line 2: Low Risk/High confidence = 4
- Line 3: High Risk/High confidence = 5
- Line 4: Low Risk/Low confidence = 1
- Line 5: High Risk/Low confidence = 3
- Line 6: High Risk/High confidence = 5
- Average Score: 3.5

Example 2:

- Line 1: High Risk/Low confidence = 3
- Line 2: High Risk/Low confidence = 3
- Line 3: Low Risk/Low confidence = 1
- Line 4: Low Risk/Low confidence = 1
- Line 5: High Risk/Low confidence = 3
- Line 6: High Risk/High confidence = 5

Average Score: 2.7

An average score of 3.5 or higher would result in LAA. So Example 1 would be LAA and Example 2 would be NLAA. Note that in Example 2, there is one line of evidence that has “high risk/high confidence” but the entire WoE shows low confidence and some aspects of low risk.

10.9.5 Other Observations

The information presented in the WoE tables for relevance, surrogacy, and robustness is very generic and it is not clear how it was used to justify decisions. For example, the evaluation of robustness of effects data for terrestrial plants does not mention anything about the number of studies supporting the assessment or the quality of the studies, let alone the degree to which the studies agree. These are the factors mentioned in the Problem Formulation chapter, Attachment 1-9, that should be considered when evaluating robustness.

In the Problem Formulation, Chapter 1, Attachment 1-9 Supplemental Information 1, the following is stated regarding establishing conclusions for high, medium and low risk, for assessing direct risks to animals and plants: “If exposure exceeds the threshold or lowest endpoint but not an endpoint where effects were observed, risk is MEDIUM.” If exposure does not exceed a threshold of effect, why is there a conclusion of any risk?



For indirect effects to plants, the risk conclusion is MEDIUM “if exposure is below the lowest EC₂₅ (terrestrial) or EC₅₀ (aquatic) for some plant types but above for other plant types.” How many are “some” plant types? Would just one below the lowest threshold qualify? How is “plant types” used here? Is it herbaceous dicot, woody dicot, etc.? Then wouldn’t the question be “is exposure below the EC₂₅ for some but not all species within the plant type”? It would not matter if there were effects on monocots, for example, if the assessment was for herbaceous dicots. Does the amount of exceedance matter, i.e., is a risk quotient of 2 considered the same as one of 20 or 200?

These three categories also ignore any type of probability analysis. Something that may be named as in the “high” category but composed of a combination of circumstances that is virtually impossible in the environment is an improbable event – and certainly, improbable events should fall into a low category for potential risk. For example, in addressing concentrations where no effects were observed in fish and aquatic-phase amphibian studies, a wide range of values across species where effects were not noted in studies with constant exposures (unlike field conditions) were presented. However, the nature of these effects data are not considered or applied to the risk assessment with respect to the probability of an actual “risk.”

Valid data is not being used in a way that supports the development of lines of evidence. For example, acute mortality (LC₅₀) data for chlorpyrifos are available for 40 fish species and 8 amphibian species based on studies submitted by the registrant or identified in the ECOTOX database. However, the value of this data is lost by the manner in which these studies are combined together, built into an “array” without meaningful and necessary analytical thought. To understand toxicity, the sensitivity range within a taxonomic group must be understood, but is lost entirely by the BE method. This, then, does not allow for appropriate association between a laboratory study and how it might be interpreted for a listed species. The ultimate flaw is that the level of resolution required for risk characterization and possible risk mitigation is entirely missing. Additionally, Species effects data and endpoints from studies ranging in exposure durations (e.g. 96 hour up to 10 days) were compiled, but it is not clear if or how the differing exposure durations were translated into simulated exercises in exposure modeling and risk analysis.

The exposure estimates are not specified in the WoE spreadsheets for the risk conclusion. The WoE spreadsheets indicate a comparison of threshold values to potential exposure values, however, the calculation of exposure needs to be clarified. Exposure values are calculated in screening exposure modeling derived from TerrPlant and AgDrift modeling, which is compiled in TED. The default assumptions used in these models need to be defined, as potential risk is overestimated due to these conservative parameters. For example, the exposure portion of the WoE spreadsheet for listed plant species is derived from the TED tool, where the input values for terrestrial plants are the threshold values from chapter 2, identified as “indirect and direct effects, for “growth” related endpoints. The TED spreadsheet uses these “growth” based threshold values to define the distance from the edge of field



that risk extends, however, this information is not considered in the WoE risk conclusion, as the WoE often concludes that critical habitat overlaps 100% in most cases. Each of these conservative assumptions should be included in the risk conclusion before a “likely to adversely affect” call is made.

In each summary table in the Appendix of Chapter 4, virtually all of the plant species in the list have "High Risk" findings listed for mortality, growth, and indirect/pollinator. This is difficult to reconcile with the findings in individual tables where it is frequently stated that NOEC or LOEC were not expected to be exceeded.

The jump from pollinator effects to indirect plant effects is poorly elucidated. Given the range in diversity in potential pollinators, it is not obvious that effects on pollinators on a field will significantly reduce pollination services on plants that are off-field and not directly exposed. Overlap with mosquito control use areas is reported in some species, but the application rates exhibiting risk may far exceed the label rate for mosquito use.

In some cases, the statement in the appendix table is made that "Indirect effects do not extend off field based on impacts to biotic diaspore dispersal vectors." [See *Kanaloa kahoolawensis*, for example]. However, indirect risks to NTP are still given as HIGH. Only off-field effects are relevant for virtually all listed plants.

The presentation of both the NOAECs and LOAECs in the WoE tables for plants makes it unclear how they were used. According to the Problem Formulation, the NOAECs are to be used in Step 1 and the LOAECs used in Step 2.

In multiple cases for the analysis of terrestrial plant risk, for example (chlorpyrifos Appendix 4-3aa, "*Pritchardia remota*"), a growth direct effect was identified as a “high risk”, with “low confidence”. The “high risk” direct effect to growth is determined based on a threshold value NOEC being exceeded for the minimum and maximum application rates at the site of a pre-emergent application, however, for post-emergence application, the LOEC value is not exceeded for the upper bound rate (4 lb a.i./A). Based on *Pritchardia remota* habitat being located in scattered, small groves in two valleys with elevations of 50 to 800 feet (information from WoE Appendix 4-3aa), actual in field or edge of field exposure would be rare at most. It is not evident that this WoE included overlap in critical habitat from the CDLs, or the TED tool analysis for distance that risk extends from edge of field.

10.10 Qualitative Marine / Cave Species Assessments

The qualitative assessments for sea turtles, whales, deep sea fish, other marine mammals, cave dwelling invertebrate species, lichens, pinnipeds, and otters suffer from lack of available data and great acknowledged uncertainties. It is difficult to accept these risk characterizations as anything other than



professional opinion supported by unvalidated screening and pre-screening models employing simplifying assumptions that do not correspond to actual characteristics of agroecosystem/natural ecosystem boundaries (beaches, estuaries, caves), stressor environmental behavior, or receptor absorption and metabolism. The categories of low-medium-high risk and confidence do not appear to be based on more than non-transparent professional judgment.

10.11 Other Types of Assessments

The comprehensive assessment for cattle ear tags represents a new methodology that has not undergone peer review. It is surprising that any LAA findings can be associated with a slow-release product that is not released into the environment upon application but rather is attached to the animal requiring protection from arthropod pests.