Technical Memo: Incorporating Mixture Toxicity into Bayesian Networks to calculate risk to pesticides in the Upper San Francisco Estuary.

Eric J. Lawrence, Skyler R. Elmstrom, Emma E. Sharpe, Wayne G. Landis*
Institute of Environmental Toxicology and Chemistry

*landis@wwu.edu.

Abstract. This memo presents the methods we have developed to calculate risk of mixtures of pesticides for the Upper San Francisco Estuary (USFE). We used curve fitting to estimate the exposure-response curves for each individual chemical and then the mixture. For the mixture the models were normalized for specific ECx values. In that way the curve fitting was optimized for effects that are similar to most threshold values. A Bayesian network was then built that incorporated four different pesticides and a specific mode of action. The input distributions of the pesticides were measured amounts from each of the six risk regions. Sensitivity analysis identified the components of the Bayesian network most important in determining the toxicity. We did demonstrate that curve fitting using additive models for mixtures can be used to estimate fish toxicity in this proof-of-concept model. Bifenthrin and the specific risk region were the two variables that were most important to the risk calculation. These techniques appear applicable to estimating risk due to the variety of chemicals and other stressors in the USFE and to the multiple endpoints under management.

Introduction
The Sacramento-San Joaquin River Delta Watershed (Delta) drains the entirety of the Central Valley of California with many different contaminants ending up in Suisun Bay and the Delta. Agricultural and urban land use practices are the primary sources for these contaminants. Contaminants have long been considered a threat to fish, as well as other aquatic organisms in the Suisun/Delta region of the upper San Francisco Estuary (USFE). The USFE contains key species and ecosystem services. The Delta smelt, a key forage fish endemic to California and only present in the San Francisco Estuary. Chinook salmon are an iconic species and many runs pass through the USFE to spawning grounds upstream. The macroinvertebrate community is a food resource to multiple fish and other species. The habitats in the region support these and numerous other birds, mammal, amphibian, and insect species, as well as provide recreational opportunities and water for irrigation, drinking, transportation.

Multiple contaminants can be found in the region along with issues of temperature and other water quality and water availability issues. The purpose of this study is to demonstrate how mixture toxicity can be estimated and risk calculated using a Bayesian network (BN) relative risk model (BN-RRM). Exposure-response models have been used to describe toxicity in BNs since Landis et al (2017a, 2020) incorporated the equations and the derived conditional probability tables as part of the process. Since 2013 we have published a series of papers using the BN-RRM (Landis 2021) and Bayesian networks have become widely used in environmental assessments (Kaikkonen et al. 2021).
Bayesian networks, or influence diagrams, are acyclic graphical models based on conditional probability distributions to describe cause-effect relationships between model variables. Bayesian networks link cause and effect relationships through a web of nodes using conditional probability to estimate the likely outcome. As summarized by Tighe et al. (2013), a BN contains the following components (Figure 1):

**Node**: A variable that can be divided into a number of states.

**State**: Conditions of the variable depicted as numerical ranges or ranks.

**Parent or Input Node**: A node that provides information to another node.

**Child or Conditional Node**: The node that receives information from a parent node.

**Link**: A graphical representation of the causal pathway between parent node(s) and child node(s).

**Conditional Probability Table (CPT)**: This table describes the conditional probabilities between the occurrence of states in the parent nodes and the resulting probabilities of states in the child nodes. In this memo the conditional probability table is computed from equations describing the interactions of chemical mixtures and the resulting toxicity.

![Figure 1. Components of a Bayesian network.](image-url)

Since the mid-2000s, Bayesian networks have been applied to environmental management, risk assessment, and guiding research and monitoring to support decision-making and resource management (Marcot et al. 2006, Pollino et al. 2007, Uusitalo et al. 2007, Barton et al. 2012, Landis et al. 2017a, Nyberg et al. 2006, Carriger and Newman 2011, Carriger et al. 2016).

Ayre and Landis (2012) demonstrated that the causal framework of the RRM translates directly into the tiered node structure of a BN. Since 2012, the utility of the integrated BN-RRM has been applied in numerous contexts including contaminated sites (Hines and Landis 2014, Landis et al 2017a), emergent disease (Ayre et al. 2014), nonindigenous species (Herring et al 2014), and forestry management (Ayre and Landis 2012). A series of papers estimating risk due to mercury in the South River in Virginia demonstrate the applicability of the BN-RRM to estimate risk to organisms and water quality (Landis et al. 2017a), human health and well-being (Harris et al. 2017) the evaluation of management alternatives, and adaptive management (Johns et al. 2017, Landis et al 2017b).
Lawrence (2020) and Mitchell et al. (2021) have demonstrated that mixtures of organophosphate pesticides can be incorporated into the BN-RRM framework. These studies took measured concentrations of OP pesticides to estimate acetylcholinesterase inhibition and translated it into changes in population dynamics for Chinook salmon. In those same models the water quality parameters dissolved oxygen and temperature were incorporated. Mitchell et al. (2021) examined the interactions within different sections of the Yakima River and estimated the risks to different segments of the Chinook salmon metapopulation.

The rest of this memo describes the development of exposure-response curve fitting models and how these tools are applied to a risk assessment of pesticides for the USFE.

**Methods**

**Study Area.** The study area is located in the Central Valley of California and encompasses an area of approximately 3,441 square kilometers. It is delineated by the Legal Delta Boundary established under the Delta Protection Act (Section 12220 of the Water Code) (CDWR 2020a) and the Suisun Boundary, Conservation Zone 11, as defined by the Bay Delta Conservation Plan (Figure 2). To encompass the entire Suisun Bay channel, the Suisun Bay boundary was extended to border the Suisun Bay Estuaries California Small Watershed, HUC12 identification 180500010401. In total, the area includes the southern half of the Sacramento River watershed, the northern half of the San Joaquin River watershed, the Delta, and Suisun Bay, Suisun Marsh and its watershed.

**Risk Regions.** As part of the BN-RRM methodology, the study area was then divided into six smaller sub (risk) regions based on hydrological delineations and land use similarities. Boundary lines follow those delineations. The resulting risk regions, from north to south, are: North Delta, Sacramento River, Central Delta, and South Delta, and from east to west: Confluence and Suisun Bay.

The **North Delta** risk region is delineated by the Legal Delta Boundary on its north and west border. Its east border includes the Sacramento Deep Water Ship Canal and is adjacent to the western border of the Sacramento River risk region. The risk region encompasses the southwest portion of Yolo County and the eastern portion of Solano County.

The **Sacramento** risk region is directly east and adjacent to the North Delta region, sharing its western border, the Sacramento Deep Water Ship Canal, with it. Its east border extends south along the Legal Delta Boundary and terminates at the northern boundary of the Central Delta risk region. This risk region encompasses the southeastern portion of Yolo County, the southwestern portion of Sacramento County and the southeastern portion of Solano County.
The Central Delta risk region borders the Confluence to the west and the study site boundary to the east. Its southern boundary includes the Clifton Court Forebay, Union Island, and Robert’s Island-Trapper Slough watersheds that delineate the northern border of the South Delta risk region. The Central subregion northern border is delineated by the Threemile Slough, South Mokelumne River, and Hog Slough watershed that forms the southern border of the Sacramento risk region. The risk region encompasses the southwestern portion of Sacramento County, the northeastern portion of Contra Costa County and the eastern portion of San Joaquin County.

The South Delta risk region shares its northern border with the Central Delta region, whereas its east, south, and western borders are delineated by the Legal Delta Boundary’s southeastern, south, and southwestern boundaries. The risk region encompasses the southwestern portion of San Joaquin County and the northeastern portion of Alameda County.

The Confluence is bordered west by the Suisun Bay risk region, on the north and south by the Legal Delta Boundary, and east by the Central Delta risk region. The eastern border originates in the south at the Lower Marsh Creek watershed border and extends north to the beginning of the Sacramento Deep Water Ship Canal. The region encompasses the southwestern portion of Sacramento County and the northeastern section of Contra Costa County.

The Suisun Bay risk region was delineated on its north, south and west borders by the Suisun Boundary. It shares its eastern border with the Confluence risk region that originates south near Shore Acres and extends northeast to the south edge of the Lucol-Hollow watershed near Montezuma Hills. Most of the region is in the southeastern section of Solano County with the Contra Costa County along its southern border.

The Bayesian network model is designed to allow a calculation for each of these risk regions or for an overarching estimate.
Project Data. The data for the Bayesian networks for the USFE were developed using an integration of water quality and chemical data from California Environmental Data Exchange Network (CEDEN)\(^1\) and DPR Surface Water Database (SURF)\(^2\), land use data from USGS NLCD\(^3\), and precipitation data from Oregon State University PRISM Climate Database. IETC-modified data and R code are accessible through GitHub\(^4\). The core data components for this project are currently stored within 6 GitHub repositories. Each repository contains R markdown documentation on how to fully reproduce data processing and analyses from a data source to its modified output. Output data is then incorporated into Bayesian networks in Netica.

IETC GitHub repositories contain the data, code, and documentation for modifying CEDEN data, modifying and integrating SURF data with CEDEN data, creating dose-response models, analysis of macroinvertebrate data, land use tabulations for our project area, and the preparation of source and stressor data for Bayesian network conditional probability tables.

The USFE project also utilizes spatial project data compiled and maintained within a local geodatabase for map production and spatial analysis. Essential shapefiles—such as project boundaries and sampling station locations—are derived from this geodatabase and stored within our GitHub repositories.

Mixture Models-Dose Response Model Averaging. The concentration addition (CA) model normalizes concentrations within a mixture by an EC\(x\) value, or a concentration that corresponds to a level of toxic effect. These normalized concentrations, also called toxic units, represent the relative potencies of the mixture that can then be added together.

\[
\sum_{i=1}^{n} \left( \frac{c_i}{EC_{x_i}} \right) = 1
\]

\(c_i\) = Concentration of chemical \(i\) in a mixture.

\(EC_{x_i}\) = Effective concentration for \(x\) level of effect for chemical \(i\).

Model Averaging Approach. The model averaging approach (MAA) builds on the CA model in a way that allows us to use the entire dose-response relationship of a mixture while only having toxicity data for individual components, assuming additivity. We fit the individual dose-response curves for the components of the mixture to calculate the EC\(x\) values. We then normalize the concentrations by a common EC\(x\), for example EC20. We normalize the concentrations for each mixture component and then fit a new curve to the EC20 normalized concentrations. Then we take the geometric mean of the model parameters to calculate an averaged curve that has

---

\(^1\) [http://www.ceden.org/](http://www.ceden.org/); California Environmental Data Exchange Network
\(^2\) [https://www.cdpr.ca.gov/docs/emon/surfwtr/surfwtr.shtml](https://www.cdpr.ca.gov/docs/emon/surfwtr/surfwtr.shtml); CA DPR Surface Water Database
\(^3\) [https://www.usgs.gov/centers/eros/science/national-land-cover-database](https://www.usgs.gov/centers/eros/science/national-land-cover-database); now accessible from the MRLC at [https://www.mrlc.gov/](https://www.mrlc.gov/)
\(^4\) [https://github.com/WWU-IETC-R-Collab](https://github.com/WWU-IETC-R-Collab). Access to most repositories for IETC projects are private, but will be made publicly accessible once the project is published.
representative shape of all the individual curves that we can put the sum of EC20 normalized concentrations as the x value to calculate the mixture toxicity. Since the MAA mixture is represented by an equation, we can input that equation into the Bayesian network to produce a CPT for mixture toxicity. Below is an example of the MAA using the log-logistic 3 parameter model.

Log logistic 3 parameter model:

\[
f(x) = \frac{d}{1 + \exp(b(\log(x) - \log(e)))}
\]

Model averaged log logistic 3 parameter model:

\[
f_{MA}(x) = \frac{\overline{d_{ECx}}}{1 + \exp(\overline{b_{ECx}}(\log(x) - \log(\overline{e_{ECx}})))}
\]

\[f_{MA}(x)\] = model averaged function of x where the x axis is the total sum of ECx normalized concentrations of components in the mixture.

\[\overline{d_{ECx}}\] = the geometric mean of all the d parameters calculated from the model fitting of ECx normalized individual component.

\[\overline{b_{ECx}}\] = the geometric mean of all the d parameters calculated from the model fitting of ECx normalized individual components.

\[\overline{e_{ECx}}\] = the geometric mean of all the e parameters calculated from the model fitting of ECx normalized individual components.

\[x = \sum_{i=1}^{n} \left( \frac{c_i}{ECx_i} \right)\]

\[c_i\] = Concentration of chemical i in a mixture.

\[ECx_i\] = Effective concentration for x level of effect for chemical i.

Building the BN. The basic format of the RRM has been published (Landis and Wiegers 1997; Wiegers et al. 1998) (Figure 3). The RRM was invented to incorporate multiple stressors and link them to multiple endpoints within a landscape. In each risk region there are sources of stressors that exist within the location or habitat that, upon release of those stressors, result in endpoints being exposed and adverse effects being generated. Ranks are used so that the combinations of sources, stressors, habitats, and effects all using different metrics can be combined into a final distribution of risk rankings (Landis and Wiegers 1997; Wiegers et al. 1998; Colnar and Landis 2007, Landis et al. 2020, Lawrence 2020, Mitchell et al. 2021).
A. The relative risk model

B. The chemical pathway

C. The Bayesian network

**Figure 3.** The steps for building a Bayesian network. Figure 3A is the fundamental structure of a cause-effect pathway for determining risk at large scales. This network is focused on the probability of toxicity to fish due to mixtures of pesticides (Figure 3B). The Bayesian network is the last step (Fig 3C). The final risk assessment for the USFE will incorporate multiple pathways for pesticides and other stressors.

Figure 3A is an illustration of the source-stressor-habitat-effect-impact pathway that describes the cause-effect structure of the RRM. In this instance the sources are the various inputs into the risk region. The stressors of interest are the pesticides. Other chemical and physical stressors can be added later. The habitat is in this case fish habitat-the waterbodies in each region. The effects are fish mortality, and fish in general are the endpoint. The impact is the risk distribution calculated. That portion of the distribution above a regulatory threshold constitutes the risk. Figure 3B summarizes the specific portions of the pathway that we are modeling. Finally Figure 3C is the structure of the BN described in more detail in the next paragraph.

Figure 4 is a portrayal of the BN used in this study. The first node designates the risk region that is used to supply the concentration distributions of each chemical to be used for the risk calculation. These concentrations are derived from the applicable CEDEN and SURF datasets. In the proof of concept there are two pathways. The top pathway is for acetylcholinesterase (AChE) inhibitors and their effect on rainbow trout fish AChE. The inhibition of AChE is translated into the probability of fish mortality as in Mitchell et al. (2021). The lower pathway uses data from conventional toxicity tests. Details of the nodes and the methods to estimate exposure responses are specific to the pathways.
Figure 4. Mixture BN model for Fish Mortality. Four different pesticides are incorporated. The mixture mortality nodes incorporate the mixture additive equations to estimate the toxicity. The concentration distributions are taken from measured values for each of the risk regions.

Pesticide Concentration Node States. For Bifenthrin and Chlorpyrifos, node states were chosen based on the EC5, EC10, EC20, EC50, and the highest record concentration for each pesticide in the field (Table 1).

For Malathion and Diazinon, node states were chosen based on the EC5, EC10, EC20 and EC50 values. For both pesticides the highest record concentration in the field was below the EC10 value and therefore it was not included as a state for these nodes.

<table>
<thead>
<tr>
<th>Values (µg/L)</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0165</td>
<td>EC5</td>
</tr>
<tr>
<td>0.0682</td>
<td>EC10</td>
</tr>
<tr>
<td>0.317</td>
<td>EC20</td>
</tr>
<tr>
<td>4.397</td>
<td>EC50</td>
</tr>
<tr>
<td>5.06</td>
<td>Highest record concentration from field data</td>
</tr>
</tbody>
</table>

Table 1. Chlorpyrifos concentrations that corresponds with state discretization in the Chlorpyrifos node
Toxicity Node States. All the toxicity nodes included in this Bayesian Network use percentage of organisms effected to indicate level of toxicity. This is illustrated with states that are broken down by 0-5, 5-10, 10-20, 20-50, and 50-100. The only exceptions to this are the AChE Activity node and the Total Fish Mortality node. The AChE Activity concentration-response curve (Figure 7) has values above 100. The Total Fish Mortality Node is additive between Mixture Toxicity 1 node and Mixture Toxicity 2 node so in some cases the result will be above 100.

Estimating Mixture Toxicity. The model averaged approach follows the following steps:

- For each mixture component, fit a log logistic 3 parameter model to the available toxicity data.
- For each mixture component, calculate the ECx.
- For each mixture component, normalize the concentrations of the toxicity data by the ECx.
- For each mixture component, fit a log logistic 3 parameter model to the ECx normalized data.
- Take the geometric mean of the three-log logistic 3 parameter model parameters for the ECx normalized models.
- Use the geometric means in the log logistic 3 parameter model to create the mixture equation.

The following figures are examples of implementing the MAA. Figure 5 shows the application of the MAA to data from Hutton et al. (2020) who conducted pesticide toxicity tests to silversides using seven pesticides with mortality as the endpoint. There are three figures (Figures 3A to 3C) with the test concentrations normalized to the respective EC10, EC20, and EC50 for each pesticide. Choosing which ECx value to normalize the data by depends on where you expect the environmentally relevant range of concentrations to be. The models will converge the most around the area of the ECx (where the ECx normalized concentration is = 1).
Figure 5. Comparison of the exposure-response model curves for several chemicals using EC10 (A), EC20 (B) and EC50 (C) normalized concentrations. The data are from Hutton et al. (2021).

Figure 6 shows the application of the MAA to data from Laetz et al. (2009) who conducted organophosphate pesticide toxicity tests to Coho salmon for individual pesticides and binary mixtures. Since we have the actual mixture toxicity results, these figures show a comparison of the MAA approach and the actual mixture results for different ECx fractions. The actual mixture results are more toxic than the MAA predicted based on the single chemical models but that is expected given that the MAA assumes additivity and chlorpyrifos and diazinon are known to be a synergistic mixture.
Figure 6. Comparison of the exposure-response model curves for several binary mixtures of organophosphates with AChE inhibition as the endpoint. The data are derived from the studies of Laetz et al. (2009) via personal communication to the investigators. The ECx values used were EC5 (A), EC10 (B), EC20 (C) and an EC50 (D).

The conditional probability tables for estimating the effects of the pesticides are derived using these two approaches.

Mixture Mortality 1 calculates AChE toxicity into mortality after Mitchell et al. (2021). The diazinon and malathion mixture uses a dose-response model constructed from a binary mixture experiment by Laetz et al. (2009) where they measured AChE inhibition in Coho salmon. Diazinon and malathion are a synergistic mixture, exhibiting more AChE inhibition than would be expected from additivity (Figure 7). The AChE inhibition node is linked to mortality using another dose-response equation (Figure 8). The equations used to construct the CPTs for the malathion and diazinon mixture pathway are below.

Malathion and diazinon mixture equation:

\[
AChE(Diazinon, Malathion) = 0 + \frac{99.24952}{(1 + \exp(10.18484 - \log(Diazinon/53.3358 + Malathion/83.29) - \log(0.05025)))}
\]

Mortality caused by AChE equation:

\[
MortalMixture\{AChE\} = (-21.97 \times \log(AChE)) + 100.49
\]
The Mixture Mortality 2 node incorporates the MAA equation for the chlorpyrifos and bifenthrin mixture to determine the CPT and calculate mixture effects (Figure 9). The MAA for Mixture Mortality 2 uses the same structure as the previous example but the MAA was calculated using the geometric means of the model parameters for bifenthrin and chlorpyrifos only, using the EC20 normalized concentrations. The MAA equation used to construct the Mixture Mortality 1 node is below. This equation uses the log logistic 3 parameter model where the (Bifenthrin, Chlorpyrifos) concentrations from the chlorpyrifos and bifenthrin concentration nodes are converted from $\mu$g/L to mg/L, divided by the respective EC20s, and summed.

Bifenthrin and Chlorpyrifos equation:

\[
\text{Mortality}(\text{Bifenthrin, Chlorpyrifos}) = 0 + \frac{1-0}{1+\exp(\log(\text{Bifenthrin}\cdot 0.000000027412+\text{Chlorpyrifos}\cdot 0.00003173264) - \log(0.337577))}
\]

**Figure 7.** Concentration-response model for AChE inhibition in Coho salmon exposed to a malathion and diazinon mixture with 95% confidence interval and 95% prediction interval. The concentration is in toxic units or EC50 normalized concentration. Data from Laetz et al. (2009).
Figure 8. Dose-response relationship between AChE inhibition and mortality for rainbow trout exposed to organophosphate pesticides. Data are from Duangsawasdi (1977).

Figure 9. Model averaging approach for EC20 normalized mixture of Bifenthrin and Chlorpyrifos. Individual curves are color coded. The points are average concentrations, and the shaded areas are 95% confidence intervals. Data are from Hutton et al. (2021).
Results

Fish Mortality. The calculation for the probability of fish mortality to the Confluence risk region is represented by Figure 10. Note that the Risk Regions node is set to the Confluence, and the concentrations of the four pesticides are set to those measured in the region. It is then possible to examine the effects of those mixtures by following each pathway to the Mixture Mortality 1 (MM1) and the Mixture mortality 2 (MM2) nodes. It is apparent that the mixture in MM2 is more toxic that MM1. The overall Total Fish Mortality percent probability is displayed in the final node. Although broken down by states a straightforward means of interpreting this output is by picking a percent threshold and counting the probability of the concentrations above that. In this instance choosing 20% mortality as the threshold results in a 51.4 probability of being above that point. If 10% mortality is the goal, then the probability is 71.4 probability of an exceedance.

Figure 10. The calculation of the probability of Fish Mortality for the Confluence risk region.

A similar process for performed for all of the risk regions. Figure 11 illustrates the summary results based on an exceedance of an 20% mortality. The Confluence risk region has the highest probability, followed by the South Delta with the Suisun Bay and Central Delta almost identical. The Sacramento River and the North Delta are the lowest and with very similar scores. Such a graph provides a straightforward picture of the relative risk to the Fish Mortality endpoint to stakeholders and decision makers.
Sensitivity analysis. The next step is to rank the most important nodes in determining the estimated Fish Mortality for each of the risk regions. We performed a sensitivity analysis on each BN that was created for the study (7 BNs total). The “Sensitivity to Findings” tool within Netica was used to run this analysis. “Sensitivity to Findings” measures mutual information between each of the input nodes and the endpoint node (Pollino et al. 2007; Norsys Software 2014). A high value of mutual information for an input indicates a greater degree of influence on the endpoint node (Marcot 2012). Mutual information is a function of both the findings in the node (input frequency) and the relationship described in the CPT (Marcot 2012; Norsys Software 2014). The outputs for each Risk Region are illustrated in Figure 12.

As would be expected the nodes closest to the final node are the most important in each of the risk regions. The pesticide Bifenthrin is the most important of the four pesticides in the final risk estimation. It is about the same influence as the AChE activity node that combines the effects of malathion and diazinon.

Uncertainty analysis. This process is a proof of concept and output should not be used to make policy decisions. The measurement of the pesticides within each of the risk regions is the best available from the databases. The estimates of toxicity using curve fitting are those for which we have the laboratory data, and we have conducted those analyses ourselves. The Fish Mortality node is that for Chinook salmon, a species that we have experience with in the past and the connection between AChE inhibition and mortality are well known. In the final risk assessment, we will be using species more similar to those in our endpoint list.

Figure 11. The calculation of the probability of exceeding 20% mortality for each of the risk regions.
Figure 12. Sensitivity analysis for each of the risk regions for each node.

Discussion

As a proof-of-concept study we do demonstrate that the approach provides estimates of fish mortality of a four chemical mixture using direct measurements from the field. The exposure-response curves are generated from laboratory data, and the total toxicity is estimated to effects levels relevant to decision making. The tools are now available to calculate risk to an individual species such as the Delta Smelt, and a similar process should be able to estimate pesticide effects to macroinvertebrates as well.

Research Support. CA Department of Pesticide Regulation, the Metropolitan Water District and the State Water Contractors.


Hines EE, Landis WG. 2014. Regional risk assessment of the Puyallup River Watershed and the evaluation of low impact development in meeting management goals. Integr Environ Assess Manag. 10:269-278.


