Integrated Environmental Assessment and Management, 2025, **21(2),** 384–395
https://doi.org/10.1093/inteam/vjae028
Advance access publication: January 6, 2025
Original Article

Environmental Policy & Regulation

Evaluation process for matching population models to regulatory decisions regarding threatened or endangered species by considering model risk

Julann A. Spromberg[1,*], Scott A. Hecht[2], Cathy A. Laetz[1], Tony Hawkes[3], and David H. Baldwin[3]

[1]Northwest Fisheries Science Center, Environmental and Fisheries Sciences Division, National Marine Fisheries Service, Seattle, WA, United States
[2]NMFS NWFSC Fish Ecology Division, National Marine Fisheries Service, Seattle, WA, United States
[3]Office of Protected Resources, Endangered Species Act Interagency Cooperation Division, National Marine Fisheries Service, Silver Spring, MD, United States

*Corresponding author: Julann A. Spromberg. Email: julann.spromberg@noaa.gov

Abstract

Population models can be an important tool in regulatory decision-making processes regarding natural resources, such as fisheries and rare species. Regulators presented with population models for their use often do not have the specific expertise to gauge the appropriateness of the model to their specific regulatory situation and decline their use in an abundance of caution. In other cases, regulators want to be involved with model development but may lack confidence in the utility of the models and their contribution to model development. The proposed process aims to address these concerns about using population models. The utility of population models depends on the available species data and the alignment of the model structure with regulatory needs. Importantly, the confidence in the available data and the model rigor need to match the types of decisions to be made, the time frame for reassessment, and the level of risk the regulator/agency deems appropriate. Model risk, defined as the possibility that the model is wrong or the output is misapplied, may stem from data limitations, parameter estimation uncertainty, model misspecification, or inappropriate use of a model. Here, we recommend a decision framework for considering the use of population models as a line of evidence in various regulatory contexts. The framework will assist regulators as they either work with modelers to construct new models or as they select from existing models to inform their decisions. Acknowledging and managing model risk increases the confidence of using models in regulatory contexts. As we move forward with utilizing models in regulatory decision-making, use of this process will ensure models fit the regulatory question, reduce model risk, and increase user confidence in applying models.

Keywords: population models, endangered species, decision making, model risk

Introduction

Whether it is forecasting the weather, calculating weight distributions aboard an airliner, estimating returns on investments, or designing the heat shield for a spacecraft, we rely on models for decision-making in many facets of modern life (Kay & King, 2020). In each case, use of the model balances the need for the information the model provides with the chance of that information being misleading. Consequences of relying on erroneous model output could range from a rainy day at the beach to a spacecraft breaking up during launch, so care is needed to select an appropriate model based on the data available and knowledge about the system (Kay & King, 2020). Modeling in the fields of biology and ecology has been growing (Grimm et al., 2014; Raimondo et al., 2018, 2021), but its use related to natural resource decisions has not been adopted as quickly (NAS, 2007, 2013).

Although population models have been used in nonregulatory assessments (Mitchell et al., 2021; Topping & Odderskaer, 2004), formal decision-making processes specifically define the information that can be considered. Regulatory decisions based on natural resource statutes, such as the U.S. Endangered Species Act (ESA), often are required to consider the best available information when determining which management actions will fulfill the policy's conservation or preservation goals. It ultimately is the responsibility of the decision-maker to determine the usefulness and importance of information taken into account when making decisions, including assessing the benefits and risks of information derived from modeling. In the United States, the Magnuson-Stevens Fishery Conservation and Management Act uses management tools, including population models, to identify catch limits that prevent overfishing and ensure sufficient harvests now and in the future (Federal Register, 2009). Population models have not been regularly used in other regulatory contexts, such as the ESA and the Marine Mammal Protection Act (MMPA), which protect rare or declining species and their habitats. The infrequent use of population models for decisions about rare species has been primarily due to a lack of species-specific data needed to develop and parameterize the models. The National Academy of Science panel on Science and the Endangered Species Act (1995) discussed the successful application of spatially explicit population models in a few cases involving rare terrestrial species. In one example, a northern spotted

Received: July 19, 2024. **Revised:** October 16, 2024. **Accepted:** October 31, 2024
Published by Oxford University Press on behalf of the Society of Environmental Toxicology and Chemistry 2025.

owl model helped identify the specific populations critical to the viability of regional metapopulations (NAS, 1995). But they cautioned that, for most endangered species, the data necessary to sufficiently parameterize models were unavailable and would require years of field work to obtain. In the meantime, they recommended continued application of ecological concepts related to population biology and habitat use to inform the conservation and recovery of endangered species (NAS, 1995, 2013). When data become available, the type and scope of the decision being made should influence the willingness and ability to incorporate models.

Model use is accepted and encouraged for decision-making in fields such as finance, meteorology, air quality, and chemical transport/fate (Kay & King, 2020; NAS, 2007). The area of population-level risk assessment outlines model development methodology and application of the output for species that have available data (Forbes et al., 2009, 2010, 2011; Kramer et al., 2011; Raimondo et al., 2018, 2021). When regulatory actions involve rare species, hesitation in relying on models can arise from an incomplete understanding of the underlying biological processes. Fully quantifying these processes is complicated by inherent individual variability, species-specific differences in responses, and scaling across levels of biological organization. To address concerns about quantifying biological scaling, the U.S. National Marine Fisheries Service (NMFS) constructed population models using empirical data and relationships to estimate the effects of pesticide exposure on ESA-listed Pacific salmon to inform biological opinions for pesticide reauthorization in the United States (NMFS, 2008, 2009). These models scaled effects across levels of biological organization from exposure to biochemical interaction, behavior changes, and alterations in individual survival and population growth (Baldwin et al., 2009). Model output informed a line of evidence in the weight of evidence process in NMFS's Biological Opinion (NMFS, 2008, 2009). Consequently, this application of population models was challenged and was one issue examined by a National Academy of Sciences (NAS) review panel. This panel recommended the use of population models when it was possible to parameterize them with species-specific data, including life history and population dynamics (NAS, 2013). Their report cautioned that when species-specific information, such as data quantifying density dependence, was not available, model use carried a high risk of misrepresenting the ESA-listed species and was inappropriate in the decision-making process (NAS, 2013). This prompted our examination of the factors contributing to model risk and how its consideration can increase confidence when selecting models for informing regulatory decisions about rare species.

Model output can be very useful for informing regulatory decisions (Raimondo et al., 2021); however, what is still needed is a process for determining the tradeoffs when considering a population model for a regulatory question. Model output is often viewed as either completely mistrusted and dismissed or overly emphasized and relied on as truth or fact (Kay & King, 2020). To address these divergent points of view, model developers need to openly communicate with users, and both parties need to acknowledge the extent to which models are capturing or representing the situation being assessed (Moon et al., 2017). As the adage goes, all models are wrong, some are useful (Box & Draper, 1987). The responsibility of selecting the useful model for each decision ultimately lies with the regulator and is driven by balancing the benefits of model use with potential losses. Losses refer to the currency of the decision; for financial decisions it is monetary loss, and for rare species it is declining population abundance. Models are beneficial when they reveal and quantify key factors about the system or evaluate alternative scenarios (Kay & King, 2020). Alternatively, the potential for loss if an error is made is high, potentially resulting in species extinction, for decisions involving threatened and endangered species. Model risk is defined as the risk that the model is wrong or the output is misapplied, as developed in the banking and financial fields (OCC, 2011). It has been established that a combination of blind faith in models that were flawed in structure and data and using those models outside of how they were designed were significant contributors to the 2008 financial crisis (Brown et al., 2015; Kay & King, 2020; Salmon, 2009). Therefore, applying the concepts of model risk management to model selection and application for regulatory decision-making will increase confidence in utilizing models and reduce the chance of misuse of model output. Model risk management has been formalized in the financial and banking fields as a direct response to the role of models in the Great Recession (Brown et al., 2015). Although fields outside of finance do not have formal model risk governance policy, regulators and decision-makers are responsible for recognizing and accepting the consequences of decisions made with misapplied or erroneous models. Here, we propose a process for determining when a model matches the regulatory needs by combining weight of evidence assessments of model risk with decision risk tolerance. Following this framework will increase confidence in the application of population models in regulatory decisions for rare species.

Process for evaluating data and models

When addressing a question regarding effects on an endangered species, regulators can develop a model for the situation or, more often, due to time and funding limitations, apply a model that is proposed to them. Proposed models may have been developed for a different purpose from the one being addressed or by an entity with a perspective on implementing the processes involved in the question that differs from the regulator. Assessing the available information about the regulatory question, the target species and potential models along with model risk will help determine whether a model would be beneficial in a regulatory context. We propose an adapted weight of evidence approach for assembling, weighing, and integrating the data and models to inform a regulatory action. The extent of the regulatory action needs to be known, including what species may be involved and how the proposed action directly and indirectly interacts with the species. Effective models will describe how particular actions effecting individuals translate to impacts at a population scale. Models developed to fulfill these requirements can then inform particular lines of evidence in an assessment related to species abundance and productivity.

Below we have outlined a process for determining when the model risk is acceptable for applying a population model to a regulatory question regarding rare or endangered species. The steps include: (1) define the regulatory context, (2) identify available data, (3) establish model objectives, (4) assess model structure, (5) match model complexity/scale, (6) recognize assumptions and uncertainty, and (7) evaluate/weigh model risk. Guiding questions for each step in this process are listed in Box 1. This framework is intended for regulatory contexts that do not have an established model selection and review process (as is prescribed for fisheries stock assessment reviews in accordance with the National Standard 2 of the Magnuson-Stevens Act section 302(g)(1)(E) Federal Register, 2009). Adhering to the proposed

Box 1. Questions for evaluating model risk of models proposed to assess impacts of regulatory actions on threatened or endangered species.

(1) *Regulatory Context:* Is the role of the model in the regulatory context clearly identified?
Do the model inputs align with management options?
Are the model outputs directly related to management objectives/measures?

(2) *Available Data:* Are the available data sufficient to build and make effective use of a model? (Box 2)
If not, what types of data need to be collected to answer the question and develop the model?
Does the model make appropriate use of the available data for the species and stressor?

(3) *Model Objectives:* Are the goals and objectives of the model clear?
Were the model objectives developed with the regulatory question in mind?

(4) *Model Structure:* Are the specific approaches and methods scientifically rigorous and capable of addressing the goals and objectives?
Are there any significant conceptual flaws?
Does the model structure appropriately represent the species biology and life history and influence of the toxicity/stressor?
Has the model structure and data handling undergone review?

(5) *Model Complexity:* Are the level of complexity and scale of the model appropriate to the question?
Does the model output, characterized by metrics on initial conditions, population performance, and population dynamics, allow comparisons across scenarios within populations?
Does the sensitivity analysis of the model reveal any factors with outsized influence that is not expected from the life history?

(6) *Assumptions and Uncertainty:* Are model assumptions and uncertainties clear and appropriate for the modeled species?
Are model strengths and weaknesses identified in terms of the accuracy and precision of model output?

(7) *Model Risk:* Are methods described in sufficient detail and satisfy data, objective, assumption, and uncertainty issues for a regulator to be willing to accept the risks to apply the model output to a line of evidence and accept the consequences of the model being wrong?

framework will provide the decision-maker more confidence in incorporating a model into their regulatory situation.

Define the regulatory context

A primary requirement of applying models to regulatory decisions is clearly defining the regulatory action in terms of the implications for the species of interest. The first step in any decision process should always be finding out "what is going on here?" (Kay & King, 2020). This could include asking to what extent a contaminant exposure will harm a rare or endangered species or slow its recovery. Clarifying the question defines the working context and leads to defining the potential impacts on the species, such as the life stages potentially affected, by how much, and how frequently. Thus, the effects of the action can be properly applied to the population dynamics, reducing the chance for misleading output. It comes down to formulating/using an appropriate model with the correct tuning built in that allows parameter adjustment and investigation of specific impacts representing the question. The model users are responsible for ensuring that the model application niche, defined as the conditions under which the use of the model is scientifically defensible (NAS, 2007), matches the regulatory question. Using a model outside of the context for which it was designed can lead to misuse of output and erroneous conclusions as well as being open to legal challenge (Moon et al., 2017).

Basing decisions on clearly defined and desired population viability endpoints, such as the population growth rate, abundance, or biomass, will focus the role of the decision within the regulatory context, regardless of whether a model is being considered. Once defined, these endpoints can guide the decision-making process and clarify a model's ability to address the regulatory question. Aligning the model inputs with the management options/scenarios and the model outputs with regulatory endpoints will ensure that the model directly relates to the management objectives.

Identify available data

The model appropriately uses the available data for the species (Box 2) and clearly communicates any limitations on how model output can be applied across populations/stocks. Threatened and endangered species often have a paucity of information available regarding basic life history and abundance. Large gaps in basic knowledge about a species may force a modeler to fill in missing information with assumptions based on other species. Models that include the use of surrogate data introduce the potential to produce output that can misrepresent impacts on the target species (NAS, 2013). Slight variations in life history strategies, such as which ages contribute more to population productivity, can result in large differences in the responses of a species to management actions or stressors (Heppell, 1998; Spromberg & Birge, 2005). In light of this, we recommend being aware of the available life history data and focus models within the scope of the data and knowledge. This is particularly important for rare species. For example, if all individuals in a species make up a single population, then all available data can be combined in a model that applies to the population. However, if data are gathered from multiple populations, some of which are not endangered and some are, combining the data in a single model could produce output that is not representative of the rare population. Unfortunately, most rare species do not have sufficient data to conduct a status assessment, much less construct a model (Beissinger & Westphal, 1998; NMFS, 2004). In an effort to improve stock assessments a U.S. National Marine Fisheries Service workshop defined six data categories, with levels 0–4 in each, to use in stock assessments: stock ID, abundance, life history, anthropogenic impacts, assessment quality, and assessment frequency. They found that data for the majority of marine mammals and sea turtles are at level 1 for all six categories, which is inappropriate for use in any sort of quantitative models and therefore, professional judgement

Box 2. Assessing key data needed for development/evaluation of population models:

Life History: Is the life history of the species or population(s) known based on empirical evidence, including: life span, time to reproductive maturity, fecundity, number of reproductive events/year or lifetime, age distribution?

Yes—Proceed to identifying data sources

No—Insufficient data are currently available to proceed with building a reliable model with model risk appropriate to represent this population. Encourage further data collection.

Data sources for parameterizing demographic rates

- Are all individuals in the species part of one single population?
  - All available data can be used together in defining the population model.
- Are the data representative of the whole rare species?
  - If so, model output could be applied to all segments of the species.
- Are data from the same species, but multiple different populations that share a life history strategy and are also listed under the ESA?
  - A generic model of the species can be made and used with caution (similar to the generic salmon models used by NMFS for the pesticide consultations.)

Toxicity/Stressor/Action information requirements:

Link available stressor/action information to demographic rates in the model structure.

Integrate the quantifiable relationships to the demographic rate(s) affected by the action

- Make sure the measurement endpoint can be translated into quantifiable factors for the assessment (e.g., how decreased somatic growth relates to survival).
- When the stressor affects specific ages or life stages, apply the impact directly to those survival or reproductive rate(s) to capture the potential impacts

should inform regulatory questions (NMFS, 2004, 2013). Even if they are not used for decision-making, models developed with surrogate data and extensive assumptions can help identify and prioritize what types of data need to be collected as resources become available. As the amount and quality of knowledge about a species increases, models can be constructed with increasing specificity to the regulatory question and management application, which mitigates model risk.Addressing regulatory questions requires combining information and data on a species with specific details on how the stressor or management action may influence the species (Box 1). Combining the life history knowledge and stressor data into a conceptual model of the system can help define what, where, and how effects may occur. This could range from all individuals benefiting from improved habitat to focused effects of a contaminant targeting a single life stage. The key modeling step is establishing quantitative relationships between the action and the species, based on empirical data. The information allows integration of the effects of the stressor into the population model. Additional data on species ecology and spatial distribution can be integrated to address more detailed questions and allow more complex model development about restoration or consequences of an action.

Establish model objectives

Each model is developed to answer a specific question by combining data with derived or assumed mathematical relationships. The model question defines the model's goals and objectives and lays out the initial conceptual model, providing insights to the purpose for the model's development. The models could fall into three categories: those developed with input from the regulator/manager to examine a specific management action, those developed without input from regulators but designed to address regulatory questions, or those designed for academic pursuits or other management questions that are later proposed for use in other regulatory situations. Ideally, models proposed for decision-making will be developed with the regulatory question in mind and with the regulator assisting in the model development (Raimondo et al., 2021), but when this is not the case, further examination of model objectives is necessary. The model's purpose defines its context and how the model output can be applied. In turn, the output directly applies to and answers the model question. A model solely designed to explore the repercussions of different assumptions about a species' life history may advance scientific knowledge but could be less appropriate for application to a regulatory context. These models inform basic research and hypothesis testing and guide field work. They also investigate management alternatives for restoration projects or future conditions, such as those resulting from projected changes in climate. Although these types of models may influence decisions indirectly, they do not attempt to guide decisions and are clear about their shortcomings. For example, a demographic model was constructed to investigate the intrinsic growth rate and potential recovery times of the endangered smalltooth sawfish (*Pristis pectinata*; Simpfendorfer, 2000). Data gaps for several key life history characteristics (e.g., age at maturity, reproductive frequency, and life span) were substituted with assumed data from the more abundant largetooth sawfish (*Pristis pristis*; Carlson & Simpfendorfer, 2015). Although the model produced some insight into the possible range of population growth rates, the model assumptions also generated large variability in the output. The large variability makes its application to regulatory decisions regarding chemical exposure impacts inappropriate without more species-specific data (John Carlson, NOAA Fisheries, personal communication, January 18, 2017). This results from the large range in assumed input parameter values that produce output that can be interpreted as both having an effect and no effect, increasing the model risk such that the model output is inappropriate for the regulatory decision and cannot provide useful information to managers for making a decision.

Assess model structure

Along with input parameters and output endpoints that align with the regulatory question, models demonstrate that they apply approaches and methods that are scientifically based and capable of addressing the stated goals and objectives with the data available. The model structure must accurately represent the species biology, life history, and responses to stressors based on the best available science and professional judgement of species specialists. In short, the behavior of the model matches and reflects the behavior of the system. This ability to represent specific biological relationships using mechanistic models allows incorporation of specific stressors. For example, Baker et al. (2018) found that a simple mechanistic model that assessed the effects of an exposure to a chemical with a known mode of action can reveal behaviors in output that might be missed by purely statistical or machine learning models. Ensuring that the model represents the system can also be supported through model verification and confirmation (Oreskes et al., 1994). In practice, this can be done by showing that the relationships and computations accurately characterize scientifically accepted theories, empirical data, and algorithms all from peer reviewed or expert sources. How the model handles environmental trends, fluctuations, and density dependence will depend on the model structure, but differences in structure can result in different outcomes (Beissinger & Westphal, 1998). The requirements for the model structure inform the selection of the modeling method, which is beyond the scope of this article. Several excellent reviews examine the attributes of available methods and guidance for model development (Forbes et al., 2009; Grimm et al., 2014; Munns et al., 2008; Raimondo et al., 2018, 2021; Schuwirth et al., 2019). Keep in mind that the model structure and data define the application niche, the set of conditions under which the use of the model is scientifically defensible (Moon et al., 2017). Use of models outside of their application niche can lead to erroneous decisions and undermine confidence in model use. Independent review of both the completed model and its underlying data can confirm the appropriateness of the application niche and increase confidence for the model use in a regulatory context.

Align model complexity/scale

The model complexity and scale match the scale of the regulatory question and the data. Each regulatory question inherently has a scale of focus that decisions will influence. Relevant scale considerations include the biological and geographic. Models can translate effects from biomolecules to cells to organs and up to the organism and population. Regulatory decisions will apply to geographic regions and/or population units, so models also need to address the appropriate scale. A geographic scale can range from a local site that a portion of a population will inhabit to an entire species range. Knowing the population range and portion of the population affected by the decision will help identify the proper model scale needed. The model complexity also needs to follow from the available data to minimize the use of assumptions that introduce unnecessary uncertainty. It is well established that adding complexity for the sake of complexity contributes to increased uncertainty and model error (NAS, 2007; Saltelli et al., 2020). Increasing complexity in an attempt to more accurately represent reality can still fail to capture important features of the system (Saltelli et al., 2020). A model that tracks an individual's daily movement through a territory may be unnecessarily complex for an impact that covers areas much larger than individual territories and/or over seasons or years. Additionally, the model inputs and initial conditions need sufficient detail and complexity to allow for assessing comparison of management alternatives or regulatory scenarios under consideration. If an assessment involves selecting the best option among several scenarios, the model input parameters need to be adjustable at a scale sufficient to differentiate between the scenarios. For example, scenarios that look at differences in contaminant exposure on individuals at a specific life stage needs a model that can modify the survival rate of the target stage separate from the other stages. It can be thought of as having the right set of knobs and sliders to tune the model to the proper settings and explore scenarios and management alternatives.

Recognize model assumptions and uncertainty

All model assumptions and model uncertainty are clear and appropriate for the species. This is particularly key if a model was adapted from a previous application; in this case, there is a need to confirm that the assumptions fit the current situation. The assumptions define the applicability of the model to the question (NAS, 2007), and clearly communicating the major assumptions will ensure appropriate use (Saltelli et al., 2020; Schuwirth et al., 2019). Identifying the strengths and weaknesses stemming from the assumptions in terms of model output precision and accuracy will clarify the consequences on model output of violating the major assumptions. The uncertainties associated with the model assumptions also may bias the model endpoints. Model uncertainties can stem from lack of knowledge of the system or errors in model relationships and assumptions. For example, measurement error may mean that the appropriate value for a potentially stable parameter is uncertain. Alternatively, natural variability may mean that a parameter value may be known for one year but will differ in future years. Conducting rigorous sensitivity and uncertainty analyses will explain the sources of uncertainty in the model output (Saltelli et al., 2020). A sensitivity analysis will identify which model parameters have the strongest influence on model output independent of variability. Uncertainty analysis can determine which parameters account for the most uncertainty in the model output. The primary way to account for uncertainties associated with parameter variability is to integrate measures of variability into models and mathematically propagate the variability from input parameters through to model output.

Acceptance and use of model output that consists of a single value with no estimate of uncertainty is naïve at best or deceptive at worst and jeopardizes the integrity of the decision-making process. Model uncertainty communicated in terms of its magnitude relative to the baseline model output magnitude provides a clear gauge of how the uncertainty influences the output. Model output with high uncertainty, even due to natural variability in the system, may be less useful in a regulatory setting and lead to erroneous decisions. For example, natural variability can mask an effect based on the factors included in the model (Spromberg & Scholz, 2011). With additional information, it may be possible to determine whether the high uncertainty can be attributed to natural variability or errors in model structure or assumptions (NAS, 2007). In either case, using a model with high uncertainty can be done after carefully weighing and communicating how the uncertainty will be incorporated into the decision-making process (NAS, 2007) and the consequences, in terms of model risk, if the uncertainty results from model error.

Evaluate/weigh model risk

The methods and data are described in sufficient detail for the regulator to confidently apply model output and accept the risk that the model is inaccurate. As discussed in the *Introduction* section, in the field of finance, the responsibility of model accuracy or inaccuracy falls on the user applying the model (Black et al., 2017; OCC, 2011). Therefore, the user needs to understand the model in terms of the steps discussed in the *Process for evaluating data and models* section and accepts the risk and consequences associated with decisions based on model output. In contrast, statutes and agency procedures may prescribe how to deal with uncertainty in regulatory contexts. For example, U.S. ESA Section 7 consultations require federal agencies to ensure that their actions do not jeopardize the continued existence of any ESA-listed species. In dealing with uncertainty, federal agencies must adopt procedures to minimize error that could result in management decisions that will protect the species less than is necessary (NAS, 1995). In a statistical context, it is equivalent to avoidance of type II error, saying no effect exists when one actually does. Model risk may stem from three basic sources: data limitations (in terms of both availability and quality); estimation uncertainty or methodological flaws in the model's design (simplifications, approximations, wrong assumptions, incorrect model design, etc.); and inappropriate use of the model (using the model outside its intended use or application niche). Addressing the issues raised in the process outlined in the *Process for evaluating data and models* section will clarify the model risk. We recommend a conceptual approach be applied to determine if the model fits the regulatory question. We propose an adapted weight of evidence approach to integrate the categories outlined in that section into lines of evidence to qualitatively evaluate model risk (Figure 1). The weight of evidence evaluation considers each question regarding model appropriateness as an individual line of evidence, which is assessed independently (Linkov et al., 2009). Each line of evidence is gauged from low to high and an overall model risk is assigned. This can be visualized graphically as choosing a risk factor level along a risk bar (Figure 1). A model that fully satisfies all of the questions in Box 1 would have a low model risk. The further the model deviates from alignment with the questions, the higher the model risk.

The level of model risk deemed acceptable to a decision-maker depends on the consequences of being wrong (Table 1), and can be considered the risk tolerance or risk appetite (Black et al., 2017). It will depend on areas of concern, such as the species status and frequency of reassessment, assuming it can occur. The difference between decisions that will stand for 2 vs. 50 years is similar to the financial investment strategy of short- vs. long-term risk. How much can one afford to lose in the time period before reassessment and a chance for a course correction, if reassessment ever happens? Another qualitative weight of evidence approach can define the risk tolerance specific to each regulatory context (Figure 2). To visualize for each area of concern, the region of risk tolerance will extend from the low end to the assigned risk tolerance (Figure 2). Less risk is acceptable when the consequences are high, such as a large potential for species decline or extinction before the question will be reexamined and management adjustments can be made. Higher model risk may be acceptable when frequent reassessment will occur. The two weight of evidence assessments are then combined to determine whether the model risk and risk tolerance match up (Figure 3). If the model risk is higher than the risk tolerance, then the model would be a poor match for the regulatory situation and its use could be misleading.

The process outlined in the *Process for evaluating data and models* section addresses concerns about model risk related to effects on endangered species and proposes an approach to evaluate and minimize model risk, similar to financial institutions establishing model risk management policies to address the costs associated with model use (Bennett, 2017; OCC, 2011). As discussed, ways to lower the model risk include strict review/oversight in model development, tailoring models to specific regulatory

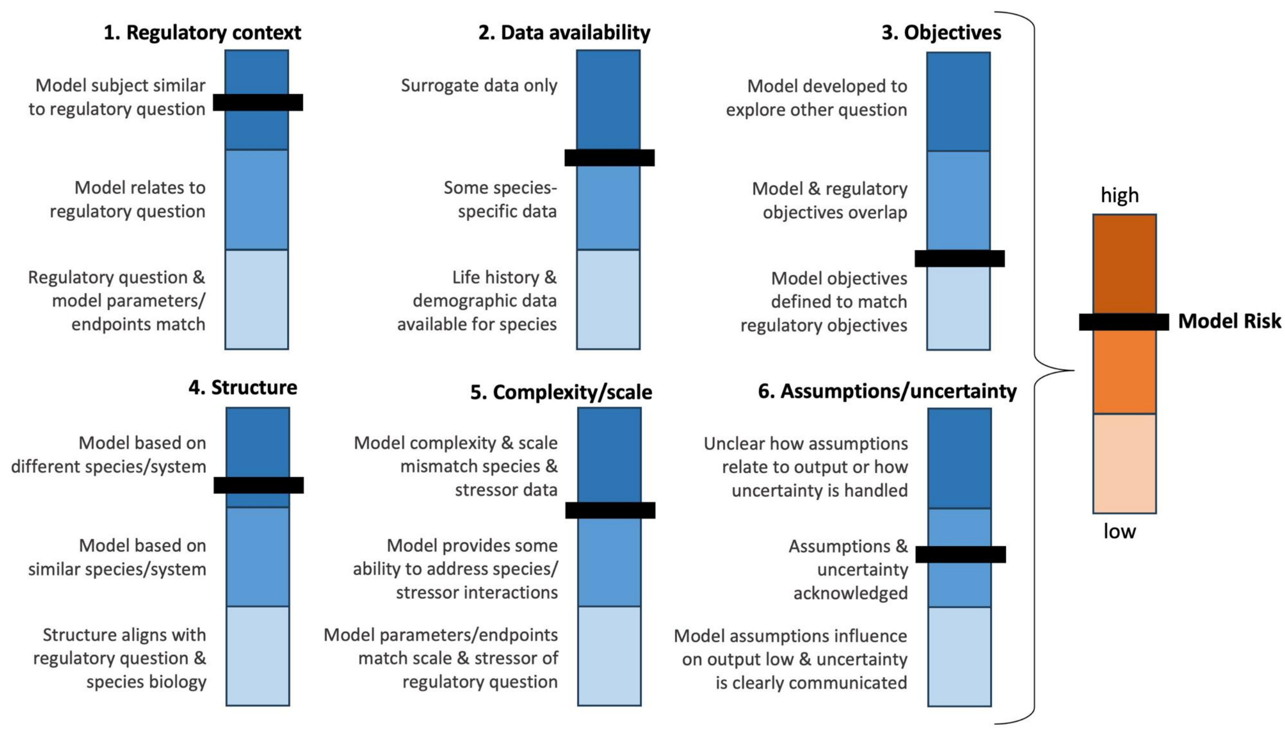

Figure 1. Six factors that contribute to a weight of evidence assessment for evaluating model risk. Each factor (black horizontal bar) is considered when determining the overall model risk. The horizontal black bar placements are for visual example, and do not represent a specific case.

Table 1. Risk tolerance for various types of regulatory decisions.

Reassessment time frame	Decision	Species/population	Stake: What could you lose if the model is inaccurate or incorrectly applied?	Risk tolerance: How comfortable with being wrong? (level of conservatism)
Short (1–5 years)	Stock assessment and harvest limits	Harvested or bycatch rare species/stocks	Abundance/biomass decline/overharvest of fishery. Economic loss from catch limits set too low.	Low to moderately conservative = medium to high risk tolerance
Medium (4–10 years)	ESA (U.S. Endangered Species Act) Consultation: Columbia River Hydropower	Threatened or endangered species	Decline or delayed recovery of ESA-listed salmon species. Loss of hydropower production.	Highly conservative = low risk tolerance
Medium (varies based on endpoint)	Restoration efficacy	Threatened or endangered species	Delay in recovery of threatened or endangered species.	Moderately conservative = medium risk tolerance
Long (15+ years)	ESA Consultation: Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) pesticide registration	Threatened or endangered species	Decline/delayed recovery/extinction of an ESA-listed species.	Highly conservative = low risk tolerance

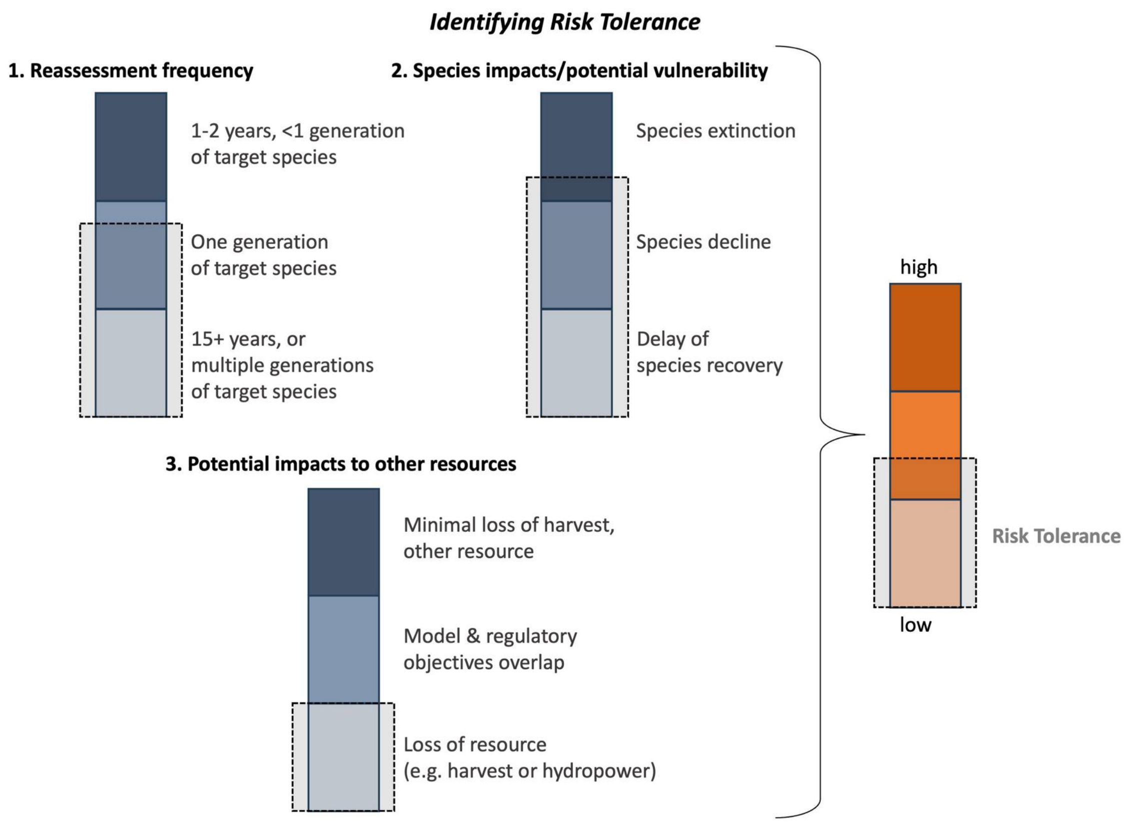

Figure 2. Factors regarding areas of concern that contribute to a weight of evidence assessment of risk tolerance/appetite (represented by the shaded dashed box) to determine whether a model is a match for the regulatory question. The dashed boxes here are visual examples, and do not represent a specific case.

applications, and limiting complexity to the scope and scale necessary (Black et al., 2017). In cases when there are concerns regarding data availability, surrogacy, or model assumptions, there are still options for using models to inform decisions. One option for mitigating model risk is by making decisions using a weight of evidence approach, with model output being one consideration among several informing a line of evidence. This will assure that decisions are not relying only on a model's output. Another option is to use ensembles of models to address the same question. The power of ensembles is that when multiple models that have different structure or input data provide consistent output, it increases the confidence in those output (Anderson et al., 2017; Jardim et al., 2021). Management strategy evaluation is a decision support framework that evaluates management scenarios using a closed feedback loop of models and observations under a range of uncertainties (Kaplan et al., 2021; Punt et al., 2016; Walter

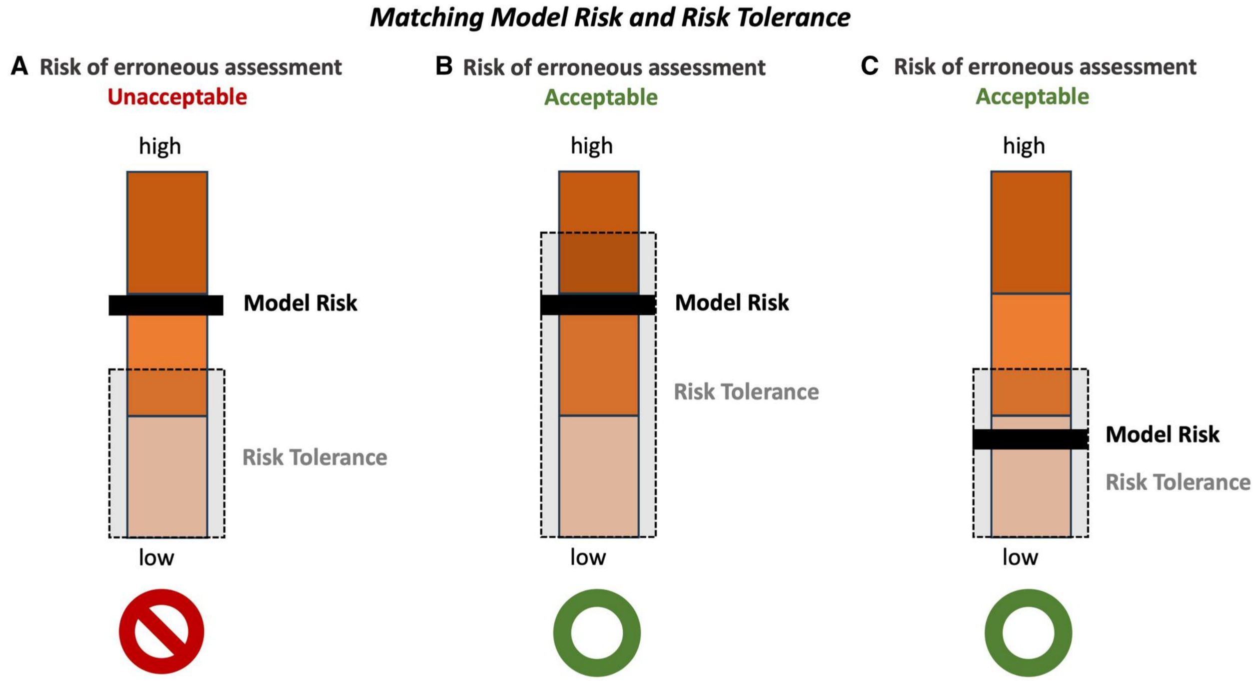

Figure 3. Model risk (horizontal black bar) and risk tolerance (shaded dashed box) are combined to determine whether a model is a match for the regulatory question. (A) Moderately high model risk and low risk tolerance mismatch for this regulatory question. (B) Model risk is moderately high, but risk tolerance is high, so model and question match. (C) Low model risk and low risk tolerance, so model and question match.

et al., 2023). A set of models make up the operating model, which represents the biological components of the system, the influence of the regulatory question on the system, and the implementation options of the management regulation (Punt et al., 2016). The cyclic nature of the process provides inherent reexamination of model uncertainty and output and minimizes model risk.

In addition to using sound modeling techniques and model validation, the Office of the Comptroller of the Currency (OCC) suggests that the model risk management process rely on the expert judgement of model users to ensure that models continue to perform properly and the users do not blindly accept model output as truth (OCC, 2011). This is similar to recommendations from environmental modeling (NAS, 2007). Models are effective tools when used appropriately but should not be treated as fact to the exclusion of other information. An extreme case of overreliance on and misapplication was the financial markets' reliance on David Li's Gaussian copula function, which led directly to the Great Recession (Black et al., 2017; Salmon, 2009). End users misunderstood or overlooked the model assumptions and limitations, failing to recognize the mismatch between model risk and risk tolerance, and that oversight had catastrophic repercussions (Figure 3A; Black et al., 2017; Kay & King, 2020). This example highlights that users must understand the models they are using and that model output informs one line of evidence in an assessment and does not automatically supersede the rest of the assessment or expert judgement. Expert judgement fulfills an essential role in scientific investigation and decision-making (Brownstein et al., 2019). In an exercise to assess ESA species listing decisions based on expert judgement, a Bayesian model incorporating stock abundance, number of populations, abundance trend, and generation time examined the listing decisions of 14 marine species assessed between 1997 and 2011. Contrary to expectation, expert judgement performed equally to the Bayesian models and resulted in the same decisions for categorizing each species as endangered, threatened, or not warranted for listing (Boyd et al., 2017).

Population models can add value to an assessment because appropriately applying them to regulatory decisions regarding endangered species can increase transparency and reproducibility (Forbes et al., 2009; Raimondo et al., 2018, 2021). Frameworks have been proposed for how to develop models and apply the output to match specific questions when the time and resources are available (Raimondo et al., 2018, 2021). In addition to this, having the separate process specifically focused on reducing model risk will increase credibility and confidence in the models selected, as has occurred in other fields (Black et al., 2017; Townsend et al., 2019). One process for assessing existing models is outlined in this article. An alternative process could be formal independent review boards, as are found in fisheries stock assessments and the Federal Columbia River Power System. Adopting any type of formal review process will require an investment of staff time. Similarly, regulatory staff time is required for developing models for specific questions, as it requires coordination with modelers and species specialists, but this is the preferred process because it ensures alignment of objectives and assumptions. Examples of these approaches are discussed as case studies.

Case studies of models and ESA regulatory decisions

Under Section 7 of the U.S. ESA, each federal agency is required to ensure that any action they authorize, fund, or carry out is not likely to jeopardize the continued existence of any endangered or threatened species (ESA-listed) or result in the destruction or adverse modification of critical habitat of listed threatened or endangered species (50 CFR §402.14(a)). The National Marine Fisheries Service (NMFS) and the U.S. Fish and Wildlife Service (USFWS) are responsible for implementing the ESA and making decisions that affect ESA-listed species under their jurisdiction. One example of ESA implementation is consultation with the relevant regulatory agencies on proposed actions to determine the risk an action poses to a species' survival and recovery. The use of population models has been growing and following are some examples of when models were considered in the consultation process.

Northern spotted owl conservation planning

After a 90% decline in available habitat, the northern spotted owl (*Strix occidentalis caurina*) was listed as threatened under the ESA in 1990 (Federal Register 55(123): 26114–26194, 26 June 1990). In support of developing the species conservation plan, a series of models were developed to guide decisions regarding size and location of habitat reserves of old growth forest necessary for spotted owl viability. Models evolved from simple to spatially complex as the questions evolved from a status assessment to considering the best size and placement of habitat reserves. Although the demographic models revealed that adult survival had the greatest influence on population growth rate, the spatially explicit models revealed that if juveniles dispersing from their natal habitat were unable to find a mate in a suitable territory, such as in a highly fragmented habitat, the populations would be at greater risk of decline and extinction (McKelvey et al., 1993). Data from targeted research on the listed populations informed the models, which decreased model risk. The risk tolerance was low due to the low population abundance and spatial distribution when the management actions were being considered (McKelvey et al., 1993, Figure 4A).

Cape Wind proposed offshore wind facility

The U.S. Fish and Wildlife Service consulted with the Minerals Management Service on their proposed lease for the construction, operation, and decommissioning of 130 wind turbine generators in Nantucket Sound. The 20-year project duration was assessed to determine the potential impact to endangered roseate terns (*Sterna dougallii dougallii*) that utilize the habitat (USFWS, 2008). Collision-induced mortality during foraging and migration flights was estimated at 4–5 roseate terns per year. This and related dependent offspring mortality were included in a consultant-proposed population viability analysis (PVA) to assess the implications to extinction risk of the Cape Wind Project. The PVA aimed to assess extinction risk, but the reliable estimates of extinction risk were short-term and did not extend through the project duration, decreasing the applicability of the model output. Models showed that key demographic parameters of juvenile survival and recruitment had dramatic effects on population growth rate but were also highly uncertain becuase they were not well studied, increasing model risk. The USFWS species specialists determined that the PVA model had limited value in assessing the population level effects of the project on roseate terns, a medium to high model risk, and did not rely on the PVA for determining whether the project would jeopardize the continued existence of the species (USFWS, 2008, Figure 4B).

Federal Columbia River power system consultations

A suite of over 20 Pacific salmon life-cycle models (LCMs) have been developed for ESA-listed interior Columbia River populations by NMFS and have been used to inform the Federal Columbia River Power System Biological Opinions. The LCMs have been documented and undergone scientific review prior to being applied in a regulatory context (e.g., Crozier et al., 2021; Pess & Jordan, 2019; Zabel & Jordan, 2020). The LCMs have been designed with the capability to evaluate population responses to hydropower operational alternatives and were also used to estimate habitat restoration actions implemented to mitigate effects of the hydropower system on ESA-listed salmonids. Proposed operation and mitigation alternatives are evaluated using the LCMs to assess potential changes in adult abundance, extinction risk, and productivity (Pess & Jordan, 2019). The LCMs are parameterized and calibrated using long-term datasets (Pess & Jordan, 2019; Zabel & Jordan, 2020). The models were informed by input and feedback from regional scientists and stakeholders, including: federal, state, and tribal resource managers, consultants, and hydropower districts. This process aligns the conceptual models, data sources, model assumptions, structure, and output to local knowledge with the intent that they be appropriate for informing the regulatory decisions. The LCMs have been scrutinized by NMFS scientists and by the Independent Scientific Advisory Board to the Northwest Power and Conservation Council. The LCMs can be updated with new data and rerun if

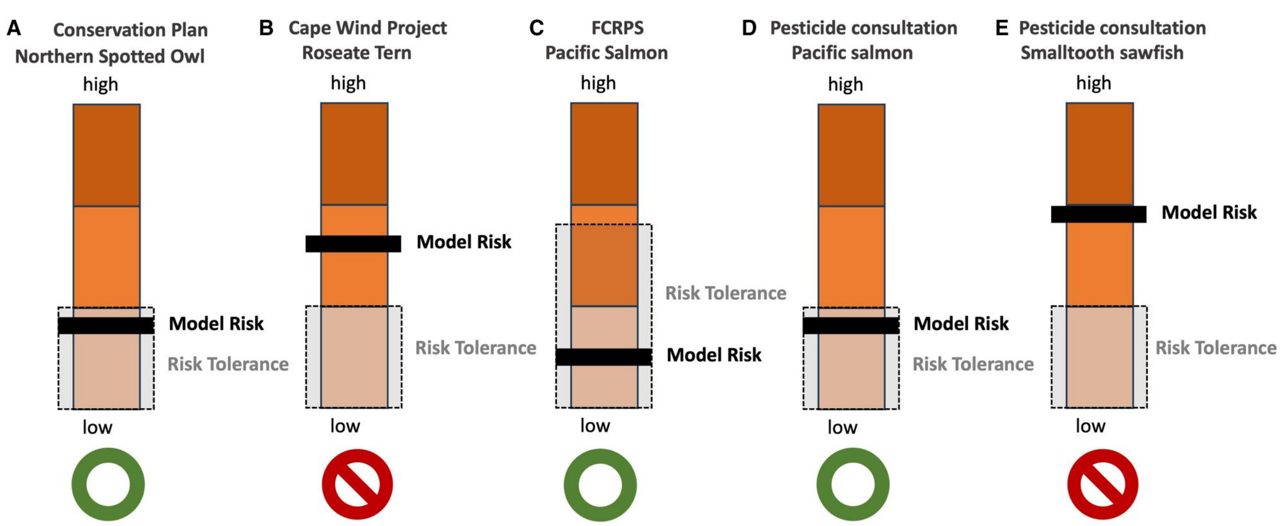

Figure 4. Visual representation of the alignment of risk tolerance and model risk when considering model use in the case examples of the conservation plan, northern spotted owl (A), Cape Wind Project, roseate tern (B) Federal Columbia River Power System(FCRPS), Pacific salmon (C), and pesticide consultations on Pacific salmon (D) and smalltooth sawfish (E).

there is a need to evaluate actions under a biological opinion (typically every four years) or to estimate effects of additional proposed actions, which lowers model risk. The short reassessment frequency also allows for a moderate tolerance of risk. Overall, these models have low risk for these regulatory questions because they are specifically designed to match the questions, are parameterized with data specifically collected to inform the models, match the complexity of the question, and directly communicate their uncertainty and assumptions (Figure 4C).

National pesticide consultations

The NMFS entered into formal consultation with the U.S. Environmental Protection Agency (USEPA) to determine whether USEPA's reauthorization of the use of specific anti-acetylcholinesterase pesticides posed a risk to ESA-listed Pacific salmonids (NMFS, 2008). Regulators and modelers worked together to construct models to determine the relative change in population growth rate resulting from anti-acetylcholinesterase (AChE) pesticide exposure during the freshwater growth phase of subyearling Pacific salmon. The model output directly informed the juvenile exposure line of evidence in the assessment (NMFS, 2022). The teamwork ensured the exposure and effects assumptions and the model output would match the needs of the assessment.

The ESA-listed species of Pacific salmon are generally considered data rich and the life histories are well established. Although data about survival and reproduction were available for some of the ESA-listed stocks and ESUs, there was not enough to parameterize a model unique for each ESA-listed ESU. Rather, data gathered from ESA-listed populations of each species were used to construct generic models for four life history strategies exhibited by ESA-listed Chinook (*Oncorhynchus tshawytscha*), coho (*O. kisutch*), and sockeye (*O. nerka*) salmon. Modeling the effects of pesticide exposure on individual salmon relied on empirical data linking pesticide-induced declines in AChE activity to decreased feeding ability, resulting in reduced somatic growth of juvenile salmon (Sandahl et al., 2005). Published data linked juvenile salmon size to age-specific survival (Zabel & Achord, 2004) and this relationship allowed the development of models investigating changes in population productivity (Baldwin et al., 2009; Macneale et al., 2014). The somatic growth portion of the model tracked changes in body size of juvenile salmon during their first year and linked to a population model through a size-dependent first-year survival rate. First-year survival was the only demographic parameter affected by the pesticide exposure and the one with direct empirical data, simplifying the modeling and reducing complexity. This assumption means that the effects seen in the model output could be an underestimate of actual impacts, but because these chemicals are not bioaccumulative, the physiological impairments are short term, and the exposures are primarily limited to the juvenile freshwater stage, the risk of misrepresenting the effects was determined to be low. The pesticide growth model integrated effects on juvenile body size with effects of the insecticides on prey availability and linked them with the population model. Model components represented different biological scales and used empirical data to inform key links between the scales. Model assumptions were supported by experimental data (Baldwin et al., 2009; Macneale et al., 2014). Sensitivity analyses ensured no parameters had undue influence from what would be expected on model output. Natural variability in the data was included by selecting every parameter from a distribution of its mean and standard deviation each iteration in the model runs. The models initially underwent iterative peer review by NMFS scientists with expertise in modeling and Pacific salmon population dynamics. These scientists ensured that the model structure and assumptions did not misrepresent, misinterpret, or overinterpret the available data. Additionally, the models were reviewed by anonymous reviewers when published in peer-reviewed journals. These multiple layers of review served to reduce the model risk. As the USEPA is officially tasked with reauthorizing pesticide use every 15 years, decisions using erroneous model output would be not be reassessed for several generations of the salmon populations. Therefore, the tolerance for model risk was low (Figure 4D). The model output informed one line of evidence in the assessment and was not relied on for the entire assessment. In all cases, model output concurred with the other lines of evidence (NMFS, 2008).

Once the Pacific salmon pesticide models were established, it was proposed that a published smalltooth sawfish model developed to explore life history characteristics (Carlson & Simpfendorfer, 2015; Simpfendorfer, 2000) be adapted to incorporate the effects of anti-AChE pesticides exposures. The life-history and demographic data gaps in the proposed model utilized data from nonlisted species of largetooth sawfish. Toxicity data would be derived from surrogate species and other taxa (e.g., fathead minnow or rainbow trout). The nonspecific data may not sufficiently represent the response of the smalltooth sawfish and makes the model risk high. After consultation with the model developers and species specialists, NMFS regulators agreed that the lack of data on the smalltooth sawfish life-history, demographics, and toxicity, and the resulting uncertainties made the model risk too high to pursue (John Carlson, NOAA Fisheries, personal communication, January 18, 2017; Figure 4E). When species-specific data become available, it may be appropriate to reconsider developing this model.

Next steps

Here, we have outlined a process that combines weight of evidence assessments of model risk with decision risk tolerance to increase confidence in the application of population models in regulatory decisions for rare species with varying amounts of available data. The examples demonstrate effective ways that independent review and close coordination during model development minimize model risk and provide support for decision-making. Acknowledging and managing model risk will open the door for applying population models to decisions regarding rare species with less data. As models are increasingly used in regulatory decision-making, processes implemented to ensure models fit the regulatory question and reduce model risk will improve each model's applicability and increase confidence in their use (Townsend et al., 2019). Ideally, we recommend that regulators and modelers work collaboratively to develop population models using the presented framework to best support decision-making processes (Raimondo et al., 2021; Saltelli et al., 2020; Townsend et al., 2019).

Data availability

No new data were generated or analyzed in support of this research.

Author contributions

Julann Spromberg (Conceptualization, Visualization, Writing—original draft, Writing—review & editing), Scott Hecht (Conceptualization, Funding acquisition, Writing—original draft, Writing—review & editing), Cathy Laetz (Conceptualization, Visualization, Writing—review & editing), Tony Hawkes (Conceptualization, Funding acquisition, Writing—review &

editing), and David Baldwin (Conceptualization, Validation, Writing—original draft, Writing—review & editing)

Funding

Funding was provided by the National Marine Fisheries Service Office of Protected Resources.

Conflicts of interest

The authors declare no conflicts of interest.

Disclaimer

The peer review for this article was managed by the Editorial Board without the involvement of J.A.S.

Acknowledgments

The authors want to acknowledge the following individuals for their insightful discussions about regulatory decision making and population modeling: George Noguchi (US Fish & Wildlife Service ret.), Andre Punt (University of Washington), Elizabeth Holmes (NOAA Fisheries, Northwest Fisheries Science Center), Jeff Jorgenson (Northwest Fisheries Science Center), and Thomas Hooper (NOAA Office of Protected Resources, ret.).

References

Anderson, S. C., Cooper, A. B., Jensen, O. P., Minto, C., Thorson, J. T., Walsh, J. C., Afflerbach, J., Dickey-Collas, M., Kleisner, K. M., Longo, C., Osio, G. C., Ovando, D., Mosqueira, I., Rosenberg, A. A., & Selig, E. R. (2017). Improving estimates of population status and trend with superensemble models. *Fish and Fisheries*, *18*, 732–741. https://doi.org/10.1111/faf.12200

Baker, R. E., Peña, J.-M., Jayamohan, J., & Jérusalem, A. (2018). Mechanistic models versus machine learning, a fight worth fighting for the biological community? *Biology Letters*, *14*, 20170660.

Baldwin, D. B., Spromberg, J. A., Collier, T. K., & Scholz, N. L. (2009). A fish of many scales: Extrapolating sublethal pesticide exposures to the productivity of wild salmon populations. *Ecological Applications*, 19, 2004–2015.

Beissinger, S. R., & Westphal, M. I. (1998). On the use of demographic models of population viability in endangered species management. *The Journal of Wildlife Management*, 62, 821–841.

Bennett, D. E. (2017). Governance and organizational requirements for effective model risk management. *Journal of Risk Model Validation*, *11*, 97–116. https://doi.org/10.21314/JRMV.2017.188

Black, R., Tsanakas, A., Smith, A. D., Beck, M. B., Maclugash, I. D., Grewal, J., Witts, L., Morjaria, N., Green, R. J., & Lim, Z. (2017). Model risk: Illuminating the black box. *British Actuarial Journal*, 23, 1–58.

Box, G. E. P., & Draper, N. R. (1987). *Empirical model-building and response surfaces*. John Wiley & Sons.

Boyd, C., DeMaster, D. P., Waples, R. S., Ward, E. J., & Taylor, B. L. (2017). Consistent extinction risk assessment under the U.S. Endangered Species Act. *Conservation Letters*, 10, 328–336.

Brown, J. A., McGourty, B., & Schuermann, T. (2015). Model risk and the great financial crisis: The rise of modern model risk management. *World Scientific Studies in International Economics: The New International Financial System*, 15, 339–353.

Brownstein, N. C., Louis, T. A., O'Hagan, A., & Pendergast, J. (2019). The role of expert judgment in statistical inference and evidence-based decision-making. *The American Statistician*, *73*, 56–68.

Carlson, J. K., & Simpfendorfer, C. A. (2015). Recovery potential of smalltooth sawfish, *Pristis pectinate*, in the United States determined using population viability models. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 25, 187–200.

Crozier, L. G., Burke, B. J., Chasco, B. E., Widener, D. L., & Zabel, R. W. (2021). Climate change threatens Chinook salmon throughout their life cycle. *Communications Biology*, 4, 222. https://doi.org/10.1038/s42003-021-01734-w

Federal Register. (2009). *Magnuson-Stevens Act Provisions; Annual Catch Limits; National Standard Guidelines*. US Government Printing Office. https://www.federalregister.gov/documents/2009/01/16/E9-636/magnuson-stevens-act-provisions-annual-catch-limits-national-standard-guidelines

Forbes, V., Calow, P., Grimm, V., Hayashi, T. I., Jager, T., Katholm, A., Palmqvist, A., Pastorok, R., Salvito, D., Sibly, R., Spromberg, J. A., Stark, J., & Stillman, R. A. (2010). Integrating population modeling into ecological risk assessment. *Integrated Environmental Assessment and Management*, 6, 191–193.

Forbes, V., Calow, P., Grimm, V., Hayashi, T. I., Jager, T., Katholm, A., Palmqvist, A., Pastorok, R., Salvito, D., Sibly, R., Spromberg, J. A., Stark, J., & Stillman, R. A. (2011). Adding value to ecological risk assessment with population modeling. *Human and Ecological Risk Assessment*, *17*, 287–299.

Forbes, V. E., Hommen, U., Thorbek, P., Heimbach, F., Van den Brink, P. J., Wogram, J., Thulke, H.-H., & Grimm, V. (2009). Ecological models in support of regulatory risk assessments of pesticides: Developing a strategy for the future. *Integrated Environmental Assessment and Management*, 5, 167–172.

Grimm, V., Augusiak, J., Focks, A., Frank, B. M., Gabsi, F., Johnston, A. S. A., Liu, C., Martin, B. T., Meli, M., Radchuk, V., Thorbek, P., & Railsback, S. F. (2014). Towards better modelling and decision support: Documenting model development, testing, and analysis using TRACE. *Ecological Modelling*, *280*, 129–139.

Heppell, S. S. (1998). Application of life-history theory and population model analysis to turtle conservation. *Copeia*, 1998, 367–375.

Jardim, E., Azevedo, M., Brodziak, J., Brooks, E. N., Johnson, K. F., Klibansky, N., Millar, C. P., Minto, C., Mosqueira, I., Nash, R. D. M., Vasilakopoulos, P., & Wells, B. K. (2021). Operationalizing ensemble models for scientific advice to fisheries management. *ICES Journal of Marine Science*, *78*, 1209–1216.

Kaplan, I. C., Gaichas, S. K., Stawitz, C. C., Lynch, P. D., Marshall, K. N., Deroba, J. J., Masi, M., Brodziak, J. K. T., Aydin, K. Y., Holsman, K., Townsend, H., Tommasi, D., Smith, J. A., Koenigstein, S., Weijerman, M., & Link, J. (2021). Management strategy evaluation: Allowing the light on the hill to illuminate more than one species. *Frontiers in Marine Science*, *8*, 624355. https://doi.org/10.3389/fmars.2021.624355

Kay, J., & King, M. (2020). *Radical uncertainty: Decision-making beyond the numbers* (528 p.). WW Norton and Company.

Kramer, V. J., Etterson, M. A., Hecker, M., Murphy, C. A., Roesijadi, G., Spade, D. J., Spromberg, J. A., Wang, M., & Ankley, G. T. (2011). Adverse outcome pathways and ecological risk assessment: Bridging to population level effects. *Environmental Toxicology and Chemistry*, 30, 64–76.

Linkov, I., Loney, D., Cormier, S., Satterstrom, F. K., & Bridges, T. (2009). Weight-of-evidence evaluation in environmental assessment: Review of qualitative and quantitative approaches. *The Science of the Total Environment*, *407*, 5199–5205.

Macneale, K. H., Spromberg, J. A., Baldwin, D. H., & Scholz, N. L. (2014). A modeled comparison of direct and food web-mediated impacts of common pesticides on Pacific salmon. *PloS One*, 9, e92436.

McKelvey, K., Noon, B. R., & Lamberson, R. H. (1993). Conservation planning for species occupying fragmented landscapes: The case of the northern spotted owl. In P. M. Kareiva, J. G. Kingsolver, & R. B. Huey (Eds.), *Biotic interactions and global change* (pp. 424–450). Sinauer Associates.

Mitchell, C. J., Lawrence, E., Chu, V. R., Harris, M. J., Landis, W. G., von Stackelberg, K. E., & Stark, J. D. (2021). Integrating metapopulation dynamics into a Bayesian network relative risk model: Assessing risk of pesticides to Chinook salmon (*Oncorhynchus tshawytscha*) in an ecological context. *Integrated Environmental Assessment and Management*, *17*, 95–109.

Moon, J. B., DeWitt, T. H., Errend, M. N., Bruins, R. J. F., Kentula, M. E., Chamberlain, S. J., Fennessy, M. S., & Naithani, K. J. (2017). Model application niche analysis: Assessing the transferability and generalizability of ecological models. *Ecosphere (Washington, D.C)*, *8*. https://doi.org/10.1002/ecs2.1974

Munns, W. R., Gervais, J., Hoffman, A. A., Hommen, U., Nacci, D. E., Nakamaru, M., Sibly, R., & Topping, C. J. (2008). Modeling approaches to population-level ecological risk assessment. In L. W. Barnthouse, W. R. Munns, & M. T. Sorenson (Eds.), *Population-level ecological risk assessment* (pp. 179–210). CRC.

National Academies of Sciences (NAS). (1995). *Science and the Endangered Species Act*. National Research Council of the National Academy of Sciences, The National Academies Press.

National Academies of Sciences (NAS). (2007). *Models in environmental regulatory decision making*. National Research Council of the National Academy of Sciences, The National Academies Press.

National Academies of Sciences (NAS). (2013). *Assessing risks to endangered and threatened species from pesticides*. National Research Council of the National Academy of Sciences, The National Academies Press.

National Marine Fisheries Service (NMFS). (2004). *A requirements plan for improving the understanding of the status of U.S. protected marine species. Report of the National Task Force for Improving Marine Mammal and Turtle Stock Assessments*. U.S. Department of Commerce, NOAA Technical Memorandum NMFS-F/SPO=63.

National Marine Fisheries Service (NMFS). (2008). *Endangered Species Act Section 7 consultation, biological opinion: Environmental Protection Agency registration of pesticides containing chlorpyrifos, diazinon, and malathion*. Biological Opinion.

National Marine Fisheries Service (NMFS). (2009). *Endangered Species Act section 7 consultation: Environmental Protection Agency registration of pesticides containing carbaryl, carbofuran and methomyl*. Biological Opinion.

National Marine Fisheries Service (NMFS). (2013). *Sea turtle assessment status and research needs*. U.S. Department of Commerce, NOAA Technical Memorandum NMFS-F/SPO-131.

National Marine Fisheries Service (NMFS). (2022). *Revised conference and biological opinion on the Environmental Protection Agency's registration review of pesticide products containing chlorpyrifos, malathion, and diazinon*. ESA Section 7 Consultation. https://doi.org/10.25923/mqyt-xh03

Office of the Comptroller of the Currency (OCC). (2011). *Supervisory guidance on model risk management*. OCC.

Oreskes, N., Shrader-Frechette, K., & Belitz, K. (1994). Verification, validation and confirmation of numerical models in the earth sciences. *Science (New York, N.Y.)*, *263*, 641–646.

Pess, G., & Jordan, C. E. (Eds.). (2019). *Characterizing watershed-scale effects of habitat restoration actions to inform life cycle models: Case studies using data-rich vs. data-poor approaches*. U.S. Department of Commerce, NOAA Technical Memorandum NMFS-NWFSC-151.

Punt, A. E., Butterworth, D. S., Moor, C. L., De Oliveira, J. A. A., & Haddon, M. (2016). Management strategy evaluation: Best practices. *Fish and Fisheries*, *17*, 303–334.

Raimondo, S., Etterson, M., Pollesch, N., Garber, K., Kanarek, A., Lehmann, W., & Awkerman, J. (2018). A framework for linking population model development with ecological risk assessment objectives. *Integrated Environmental Assessment and Management*, *14*, 369–380.

Raimondo, S., Schmolke, A., Pollesch, N., Accolla, C., Galic, N., Moore, A., Vaugeois, M., Rueda-Cediel, P., Kanarek, A., Awkerman, J., & Forbes, V. (2021). Pop-guide: Population modeling guidance, use, interpretation, and development for ecological risk assessment. *Integrated Environmental Assessment and Management*, *17*, 767–784.

Salmon, F. (February 23, 2009). *Recipe for disaster: The formula that killed Wall Street*. Wired Magazine 17(3):17-03. https://www.wired.com/2009/02/wp-quant/.

Saltelli, A., Bammer, G., Bruno, I., Charters, E., Di Fiore, M., Didier, E., Nelson Espeland, W., Kay, J., Lo Piano, S., Mayo, D., Pielke, R., Jr, Portaluri, T., Porter, T. M., Puy, A., Rafols, I., Ravetz, J. R., Reinert, E., Sarewitz, D., Stark, P. B., Stirling, A., van der Sluijs, J., … Vineis, P. (2020). Five ways to ensure that models serve society: A manifesto. *Nature*, *582*, 482–484.

Sandahl, J. F., Baldwin, D. H., Jenkins, J. J., & Scholz, N. L. (2005). Comparative thresholds for acetylcholinesterase inhibition and behavioral impairment in coho salmon exposed to chlorpyrifos. *Environmental Toxicology and Chemistry*, *24*, 136–145.

Schuwirth, N., Borgwardt, F., Domisch, S., Friedrichs, M., Kattwinkel, M., Kneis, D., Kuemmerlen, M., Langhans, S. D., Martínez-López, J., & Vermeiren, P. (2019). How to make ecological models useful for environmental management. *Ecological Modelling*, *411*, 108784.

Simpfendorfer, C. A. (2000). Predicting population recovery for endangered western Atlantic sawfishes using demographic analysis. *Environmental Biology of Fishes*, *58*, 371–377.

Spromberg, J. A., & Birge, W. J. (2005). Modeling the effects of chronic toxicity on fish populations: The influence of life-history strategies. *Environmental Toxicology and Chemistry*, *24*, 1532–1540.

Spromberg, J. A., & Scholz, N. L. (2011). Estimating the future decline of wild coho salmon populations resulting from early spawner die-offs in urbanizing watersheds of the Pacific Northwest, USA. *Integrated Environmental Assessment and Management*, *7*, 648–656.

Topping, C. J., & Odderskaer, P. (2004). Modeling the influence of temporal and spatial factors on the assessment of impacts of pesticides on skylarks. *Environmental Toxicology and Chemistry*, *23*, 509–520.

Townsend, H., Harvey, C. J., deReynier, Y., Davis, D., Zador, S. G., Gaichas, S., Weijerman, M., Hazen, E. L., & Kaplan, I. C. (2019). Progress on implementing ecosystem-based fisheries management in the United States through the use of ecosystem models and analysis. *Frontiers in Marine Science*, *6*, 641. https://doi.org/10.3389/fmars.2019.00641

U.S. Fish and Wildlife Service (USFWS). (2008). *Endangered Species Act Section 7 consultation, biological opinion: Cape Wind Energy Project Nantucket Sound, Massachusetts*. Formal Consultation # 08-F-0323, U.S. Department of Interior.

Walter, J. F., Peterson, C. D., Marshall, K., Deroba, J. J., Gaichas, S., Williams, B. C., Stohs, S., Tommasi, D., & Ahrens, R. (2023). When to conduct, and when not to conduct, management strategy evaluations. *ICES Journal of Marine Science*, *80*, 719–727. https://doi.org/10.1093/icesjms/fsad031

Zabel, R. W., & Achord, S. (2004). Relating size of juveniles to survival within and among populations of Chinook salmon. *Ecology*, *85*, 795–806.

Zabel, R. W., & Jordan, C. E. (Eds.). (2020). *Life cycle models of interior Columbia River Basin spring/summer-run Chinook salmon populations*. U.S. Department of Commerce, NOAA Technical Memorandum NMFS-NWFSC-156.

 Integrated Environmental Assessment and Management, 2025, 21, 384–395 https://doi.org/10.1093/inteam/vjae028 Original Article