Improving Access to Fish Consumption Advisories and Maintaining Confidence in California's Healthy Seafood Products

Technical Documentation submitted in partial satisfaction of the requirements
for the degree of Master of Environmental Data Science
for the
Bren School of Environmental Science & Management

By:
Benjamin Versteeg
Hope Hahn
Kate Becker
Luna Herschenfeld-Catalán

Committee in charge:
Bruce Kendall
Carmen Galaz García

Clients:
Scripps Institution of Oceanography
California Cooperative Oceanic Fisheries Investigations (CalCOFI)

May 2024

SaferSeafood

Improving Access to Fish Consumption Advisories and Maintaining Confidence in California's Healthy Seafood Practice

As developers of this Capstone Project documentation, we archive this documentation on the Bren School's website such that the results of our research are available for all to read. Our signatures on the document signify our joint responsibility to fulfill the archiving standards set by the Bren School of Environmental Science & Management.

____ BENJAMIN VERSTEEG

____ HOPE HAHN

____ KATE BECKER

____ LUNA HERSCHENFELD-CATALÁN

The Bren School of Environmental Science & Management produces professionals with unrivaled training in environmental science and management who will devote their unique skills to the diagnosis, assessment, mitigation, prevention, and remedy of the environmental problems of today and the future. A guiding principle of the School is that the analysis of environmental problems requires quantitative training in more than one discipline and an awareness of the physical, biological, social, political, and economic consequences that arise from scientific or technological decisions.

The Capstone Project is required of all students in the Master of Environmental Data Science (MEDS) Program. The project is a six-month-long activity in which small groups of students contribute to data science practices, products or analyses that address a challenge or need related to a specific environmental issue. This MEDS Capstone Project Technical Documentation is authored by MEDS students and has been reviewed and approved by:

____

DR. BRUCE KENDALL, FACULTY ADVISOR

____

DATE

____

DR. CARMEN GALAZ GARCÍA, CAPSTONE ADVISOR

____

DATE

Acknowledgments

This project could not have been completed without the support and guidance of this project.

Faculty Advisor:	Dr. Bruce Kendall, Bren School of Environmental Science & Management
Capstone Advisor:	Dr. Carmen Galaz García, Bren School of Environmental Science & Management
Clients:	Dr. Erin Satterthwaite, Sustainability Researcher, Sea Grant & CalCOFI, Scripps Institution of Oceanography, UCSD. Dr. Lillian McGill, Scripps Institution of Oceanography, UCSD. Dr. Brice X. Semmens, Director, California Cooperative Oceanic Fisheries Investigations (CalCOFI), Professor, Scripps Institution of Oceanography, UCSD
Special Thanks:	Allie Caughman Ph.D. Candidate, Bren School of Environmental Science & Management Dr. Amalia Aruda Almada, Extension Program Leader, Science, Research, and Policy Specialist, *University of Southern California Sea Grant Program* Maria Madrigal, M.S., Education Manager, University of Southern California Sea Grant Program Ian Brunjes, SCCOOS Dr. Wes Smith, Chief, Fish, Ecotoxicology & Water Section, OEHHA/CalEPA All the amazing people in the Masters of Environmental Data Science Cohort and Capstone Committee

Abstract

Dichlorodiphenyltrichloroethane (DDT) is an insecticide that is resistant to degradation and can cause increased risks of cancer, premature births, developmental abnormalities, and neurological diseases in humans and animals. Although banned in 1972, DDT has been dumped into the ocean off the coast of Southern California. Recent documentation of the extent of this contamination has captured public attention and raised concerns regarding the consumption of contaminated seafood. The state of California currently issues statewide consumption advisories for coastal communities, but these advisories do not address DDT. To improve consumption advisory accessibility, the Bren School of Environmental Science & Management MEDS Capstone team, SaferSeafood, has partnered with the Scripps Institution of Oceanography and the California Cooperative Oceanic Fisheries Investigations (CalCOFI), who compiled previously collected data on the incidence of DDT in sediment and fish in nearshore Southern California waters. The team has updated and validated a spatiotemporal statistical model to predict DDT concentrations in fish based on species and location. The model was then integrated into an interactive web application that allows anglers to input information about their catch and receive predicted DDT concentrations as well as consumption advisories. The results of this project allow individuals to make well-informed decisions about their seafood choices in the face of environmental challenges and health risks associated with DDT contamination.

Authors

This project was completed by a group of students pursuing a Master's degree in Environmental Data Science at the Bren School of Environmental Science & Management, UC Santa Barbara. Each team member was assigned roles and assigned work accordingly. The Project Manager, Ben Versteeg, was responsible for ensuring project deliverables were being met, scheduling meetings, writing agendas, and team cohesion. The Communications manager, Luna Herschenfeld-Catalan, served as the point person for communication between the team, client, advisors, and any external resources. Hope Hahn, the Data Manager, was responsible for the tidying of data, organization of code and files, and archiving of all data and metadata utilized for the deployment of the final deliverables in order to ultimately ensure transparency and reproducibility. Finally, Kate Becker, the product lead, was responsible for managing the team’s application creation, deployment, and debugging.

Executive Summary

The harmful effects of Dichlorodiphenyltrichloroethane (DDT), and its breakdown products (DDX), have triggered widespread concerns, particularly due to the recent rediscovery of a barrel field containing DDT-laced sludge off the Southern California coast (Los Angeles Times). This alarming find has not only captured the public's attention but has also highlighted its potential threats to human and environmental health. DDT is a synthetic organic compound that was commonly used in insecticides for decades prior to its ban in 1972. It can accumulate in the fatty tissues of fish through the intake of toxic sediment due to bioaccumulation and magnification up the food chain. The negative side effects of DDT, including heightened cancer risks, premature births, developmental abnormalities, and neurological diseases in both humans and animals, have raised concerns of consuming seafood from the contaminated area. The consequences go beyond immediate health worries also affecting the local economy and the well-being of recreational fishing communities.

DDT dumping in the Southern California coast is an important current issue due to there not being enough information about the DDT concentration in the fish caught there. The California Environmental Protection Agency Office of Environmental Health Hazard Assessment (OEHHA) currently issues statewide consumption advisories for coastal communities. However, these advisories are severely limited as they are site and species-specific, covering only two chemicals: mercury and Polychlorinated biphenyls (PCBs). This gap in offering comprehensive guidance on safe seafood choices aggravates the challenge for consumers seeking reliable information to make informed decisions about their seafood consumption.

Figure 1: Advisories provided by OEHHA

In response to these critical issues, scientists from the Scripps Institution of Oceanography and California Cooperative Oceanic Fisheries Investigations (CalCOFI) have collected and analyzed fish and sediment monitoring data to understand the extensive human and ecological impacts resulting from legacy DDT dumping. Their current model accurately predicts the localized risk of DDT in sport fish off the Southern California coast. This risk encompasses the potential adverse effects on human health due to exposure to these contaminants (Marjadi et al. 2021). The goal of this project was to expand on this by updating their model and predicting fish DDT levels using factors like sediment DDT, capture year, and fish characteristics. These were used to develop a spatiotemporal statistical model to predict DDT concentrations for species and locations included in the sample collected by Scripps.

Taking transparency and accessibility into account, the team created an interactive web application. This application allows anglers to input specific details of their catch such as species and location, and receive personalized predictions of DDT concentrations with corresponding advisories. Ultimately, this project aspired to inform and educate, allowing individuals to make well-informed decisions about their seafood choices in the face of environmental challenges and health risks associated with DDT contamination.

The effects of this problem have traditional, cultural, and socioeconomic impacts. A survey conducted by Heal the Bay, an environmental advocacy group, surveyed 3030 pier anglers over 12 months and found that 78% only fish for subsistence from piers and from shore (never from boats); 84.6% are of non-White/Euro-American ethnicity and speak English as a second language (Stevenson, Charlotte et al., 2011). Through conversations with Southern California anglers, the team found that there was a common theme throughout conversations. There has always been a gap between public policy/scientific research and the actual wants and needs of those most affected, such as anglers. Heal the Bay has also stated that engaging non-traditional stakeholders in public policy is critical for decision-makers to gauge all views from those standing to be affected by a policy and not just those that regularly attend policy meetings. The creation of the SaferSeafood application creates an easily and openly accessible platform for community members, of all demographics, the tools and resources to stay informed on informational updates on regulations and policies.

This application was created as a result of intensive data collection, statistical analysis, user testing, and performance predictions. After initial data analysis, performed by the clients, it was found that there is a strong linear relationship between DDT in fish and sediment DDT. This finding prompted the Scripps team to develop a series of candidate models for predicting DDT based on life history characteristics and SaferSeafood confirmed the selection of the model that performed best.

After the best-performing model was selected for DDT predictions, the model was used in the SaferSeafood online application for the public to make their own DDT inquiries based on their catch; ultimately attempting to bridge the gap between public knowledge and scientific research as requested by Southern California anglers.

Table of Contents

Problem Statement

The introduction of pollutants into coastal ecosystems threatens the quality and quantity of seafood yields through accumulation in sediment and biomagnification in marine food webs. From as early as 1940, the deep ocean basins off the coast of Los Angeles have been dumpsites for hazardous industrial waste including pesticides, petrochemicals, and other harmful contaminants (Venkatesan et al. 1996). As a result, legacy pesticides such as DDT have accumulated in the tissues of fish through various exposure pathways, such as the uptake of toxic sediment (Klasing, S., and R. Brodberg., 2008). Uncertainty in the quantity, location, and contents of dumped materials poses an ongoing risk to marine ecosystems and human health. This also has a profound effect on Southern California's economy. NOAA has estimated that anglers spend approx $156 million dollars on recreational fishing in California generating more than $200 million in annual economic output and supporting nearly 1,400 jobs (Marquis, Sarah., 2015). Similarly, there remains a large socioeconomic importance in subsistence fishing, such as pier fishing within rural, indigenous, and native communities as it supports social, nutritional, and cultural benefits as well as tradition. The accessibility and low cost of pier fishing may make it an especially attractive option for poor, undocumented, and underprivileged members of urban communities (Quimby et al. 2020). That being said, the influx of seafood contamination does not only affect the economy but the communities that rely on it for their livelihood.

On a broader scale, due to the long-term impacts of anthropogenic changes in the Southern California Bight, contaminant monitoring for sediment fish in this region has been ongoing since 1998. Sediment data has been collected by the clients via grab samples at the top 5 cm of embayment sites and fish data has been collected via fish tissue samples off piers and boats and saved for analysis as muscle tissue filet. The findings of this initial study confirmed that there was a strong linear relationship between DDT in fish and sediment DDT varying both by diet and by habitat. It was also found that the highest DDT in sediment occurred on the Palos Verdes shelf which supported the theory that contaminants persist in their environment even decades after initial exposure. Unfortunately, there is no public platform for consumers and fishermen to access information and advisories regarding DDT concentrations in their caught seafood.

After about 25 years of targeted data collection, the clients performed model validation and measured predictive performance on lipid-normalized data. This initiated a modified study that would employ non-lipid normalized data: data that was not statistically transformed or collected via tissue samples. This study targets those who don't have the resources or tools to make predictions on data derived and modified by research scientists. As a result, the final model uses species and location, primarily, to measure DDT concentrations for catch with relevant California consumption advisories using spatial predictions.

While this project hopes to inform and educate, the main goal is to limit consumption of contaminated seafood but there have been previous barriers to reaching the target audience. The option to eliminate or even limit consumption of contaminated seafood may be too great of a hardship or disruption in lifestyle for many anglers and therefore may not be accepted. This remains a much larger problem than health concerns but ultimately relies on the proximity of other safe fishing sources, easy access to other inexpensive food sources, and programs that allow local people to maintain community and traditional activities, for this to be most effective. Although the group is trying to promote community health, this project is a small part of a much larger environmental injustice and socioeconomic issue. As a prime example, the New York Department of Environmental Conservation has found that both the general public and non-commercial anglers do not widely support fishing bans as a means to protect public health because it's largely self-regulated (Office of Science and Technology Office of Water U.S. Environmental Protection Agency Washington, DC., 1990). The team hopes that the resulting application will provide the public with knowledge giving them the confidence to act accordingly in terms of their health.

Specific Objectives

To address these issues, there were two main objectives:

1. Update the client's existing spatiotemporal statistical model to:
    a. Predict non lipid-normalized DDT concentrations (instead of lipid-normalized DDT concentrations) in sportfish across Southern California
    b. Improve the predictive accuracy
2. Create a publicly accessible and transparent interactive web application that implements the updated model and allows anglers to input catch and location to receive predicted DDT concentration and consumption advisories

Summary of Solution Design

1. Data

The data used in the dashboard, with the exception of Fishbase, was compiled by Scripps and CalCOFI from stores of previously collected data. The data used for this project was publicly available to us on the CalCOFI database and within Dr. Lillian McGill's manuscript (McGill, L. 2024). The data includes three comprehensive databases: Sediment Rasters, Fish DDT Monitoring Data, and Species Life History Characteristics. All data was collected from nine primary sources between 1998 and 2021. The metadata can be found in the team's GitHub repository. All of the data used for this project was collected from public data files, and all code and future data/modeling will be available publicly through the team's GitHub organization as well as publicly on Data Dryad. All statistical and web application coding will be conducted in R within RStudio, so any interested parties will be able to reproduce any work in R.

The Sediment Raster data is modeled sediment DDT concentrations based on sediment samples across the Southern California Bight in 2003, 2008, 2013, and 2018. This includes test site specifics and fishing zone averaged DDT concentrations present in marine sediment. Fishing zone averages were generated by the client to be used as covariates in a model of fish bioaccumulation and continuous spatial estimates of sediment DDT were calculated by fitting spatiotemporal regression models to DDT measurements. The 'pelagic_nearshore_fish_zones.rds' consists of spatial arrays, referred to as fishing zones, used by the client to approximate fish contaminant exposure when the capture location was unknown. Uncertainty in extrapolation results increased the further the data collected was from the study area.

The DDT Monitoring Data is a database of fish DDT monitoring data collected with the Southern California Bight. Contaminated fish tissue data was aggregated from eight different surveys across Southern California collected between 1998 and 2021 and are reported in lipid-normalized as well as non-normalized wet weights. Normalizing the lipid data, performed by the client, was due to the positive relationship between lipid content and organic contaminant concentration in fish tissue. Comparatively, the team assumes the audience of this web application to be primarily fishermen who may not have the necessary tools and resources to measure tissue wet weights in their catch, so the values remain non-transformed.

Finally, the 'fish_life_history.csv' table was largely used to assign 61 species diet and habitat classifications according to adult life history characteristics. These variables were used to test relationships between different covariates in the model. Fishbase was also used via the *rfishbase* package in R to obtain fish life history characteristics and to test for prediction accuracy for out of sample species.

2. Analysis

The main analysis for this project was based on a pre-existing workflow developed by the client. The original study normalized DDT concentrations which they standardized based on lipid concentration in fish (lipid-normalization) due to the previously documented positive relationship between organic contaminant concentration and lipid content in fish tissue. It is anticipated that the intended audience, anglers, will not be measuring the amount of lipid in their catch, so the clients preferred a model with DDT concentrations that are not lipid-normalized. Excluding lipid DDT concentrations in the model prompted the change in model formula and the need for additional model validation.

All of the datasets were joined into a large tabular dataset that included all necessary parameters as well as additional columns with appropriately transformed parameters. All data cleaning code can be found in the Models repository under the pathway "/code/data_cleaning/data_cleaning.R/".

After data was collected and wrangled, a series of models were tested for predictive performance for measuring DDT. Using the client's model performance criteria, we found that the best model for predicting DDT concentration in fish contained the same variables as the one the client had previously found to best predict DDT concentration in lipids. This model had similar accuracy and performance to the client's initial model which used lipid-normalized data.

To further test the predictive performance the model was run through a machine learning approach and it was found that the best performing model was the one that used all covariates (sediment DDT concentrations, diet, habitat, species, and year). The process of model validation is described in more detail below in this section.

After the candidate model was selected it was implemented into the Shiny Dashboard that accepts location and species as inputs (two predictors in the selected candidate model) and outputs DDT concentrations in a user's catch. All data used in the shiny dashboard can be found in the Shiny-Dashboard repository under the pathway "/shinydashboard/data/".

Validation and Analysis:

brms Models (brms_workflow.rmd)

The model was created by taking the client's exact model structure but switching out the DDT concentration variable to the one that was appropriately transformed for this project (non lipid-normalized as opposed to lipid-normalized). As the clients had used the *brms* package to create their models and used the loo package to perform LOO cross-validation, the initial approach was to use the *brms* format to create the models and use LOO cross-validation to determine the best performing model (choose model with highest elpd). Each model was created using the brm() function, and the formulas for the models that were tested were the same as the client's original model testing (found in client_models.R), however, the DDT concentrations were non lipid-normalized as opposed to their lipid-normalized DDT concentrations. The LOO() function was used to test model performance. The best model was chosen using LOOIC (leave one out information criterion) which showed that the best updated model formula used the same predictors as the client's best-performing model. After testing the updated model with the best performance, this model was compared against the client's original model using LOO cross-validation scores. The updated model had a slightly higher LOOIC value than the original model but was still the highest performing compared to all of the other updated models.

Because this model was used to predict DDT concentration in fish for data points, and potentially species, that were not included in the original dataset, the model was tested again with species removed as a predictor. This model performance was also tested using LOO cross-validation, and the results were compared against the best original and updated models. Removing species as a predictor caused a decrease in model performance, so it was added back into the model. However, to further test the generalizability of out-of-sample data, the model was also tested by: updating the formula to include family instead of species, running LOO cross-validation, and comparing the results to the original and updated models. This model also did not perform as well as the original model, but it performed better than the model that completely removed species. The final model used in this project included all original parameters used by the client: trophic level, feeding position, species, year, and sediment DDT concentration. This model was saved in the object called 'brm.diet.habitat.species.year.non' and can be found in the 'brms_workflow.Rmd file' and was also saved to 'brm_species_model.rda' in the 'data_outputs' folder within the 'data' folder.

Machine Learning Analysis (tidymodels_workflow.rmd)

To evaluate the robustness and predictive power from the models using *brms* and *LOO*, the model was run through a machine learning approach in the *tidymodels* package, in which the data was split into a training set and a testing set. The training set is 70% of the data while the testing set is 30% of the data and is left out of all of the model fitting processes. To apply Bayesian statistics to the model, the extension package *multilevelmod* was used in the *tidymodels* workflow. The same formula and priors as the *brms* models were used in the machine learning testing. With *tidymodels*, a model specification and workflow were set, and this workflow was used to fit the training data to the model. The model fit with the training data was then used to make predictions on the testing data. The performance of the model was determined using the R-squared and RMSE values because LOO CV was not used in this process. The R-squared value was used to compare predictive power to the models that were fit using *brms*. The R-squared value was similar for both methods using *brms* and *tidymodels*, confirming the model performance.

Feature Selection (feature_selection.rmd)

To add to the machine learning process, the *projpred* package was used to understand whether all predictors in the final model were useful and the model was not overfitted. The results of this process suggested that all the variables in the model were important in making model predictions, so the original model was left as is.

html outputs of all code files will be included in the repository for more detailed outline of the model processes.

3. Web Application

(1) A description of the purpose of the app and how it supports the clients goals

The primary deliverable for the client is a user-friendly web application designed to assist fishermen by predicting DDT concentrations in various fish species based on their catch location and species. The *SaferSeafood Web Application* is created for recreational anglers in Southern California in order to inform them of California seafood safety and advisories. The user is first shown an interactive map to select the location of their catch and specify the species. These variables are used to fill the parameters to run the Bayesian model and calculate predictions. The rendered output displays a range of values that were calculated using the `predict()` function in R. This is the 95% credible interval of predicted DDT concentrations in the fish species at the specified location. This measurement is shown in ng/g units, which reflects the DDT levels typically found in the tissue of the species based on the entered parameters. Along with the DDT concentration, a recommended number of servings and other health advisories from California EPA's OEHHA are returned.

The website highlights that the resulting outputs should be regarded as an educational resource to make informed decisions rather than being a definitive guide to whether the fish is safe to eat. The app includes links to other seafood advisories (regarding contaminants other than DDT) provided by the California Office of Environmental Health Hazard Assessment (OEHHA), the Food and Drug Administration (FDA), and the Fish Contamination Education Collaborative (FCEC).

(2) How the analysis feeds into the app.

When the user selects a species within the collected sample, the species name and relevant life history characteristics are inputted to the model. When the user selects a location, the coordinates are assigned to the nearest fishing zone within 500m. If the selected location is outside of the study area defined by the fishing zones, an "invalid location" error is displayed that guides the user to select a different location. Time-varying fishing zone averages of DDT concentration in sediment were calculated for each fishing zone defined in the paper. Using the coordinates, the associated concentration of DDT in sediment is inputted into the regression model. The model generates a predictive range of DDT concentrations in the species at the specified location, and these concentrations are used to assign a range of recommended number of servings based on the health advisory thresholds for DDT consumption. Additionally, the coordinates are assigned to the nearest OEHHA advisory blocks within 500m to communicate existing health advisory recommendations that are also based on concentration of mercury and PCBs. Users are encouraged to follow the more conservative estimate of the number of recommended servings a week between the team’s DDT advisories and OEHHA’s advisories.

Products and Deliverables

1. Web Application

A web application will be provided to the public that will allow users (eg. anglers, researchers, and the general public) to input species and location to receive catch-specific DDT concentration predictions, recommended number of servings, and additional information on California consumption advisories provided by OEHHA. This application will be hosted and maintained by CalCOFI on their organization's independent infrastructure.

2. Repository

A GitHub organization holds three GitHub repositories (Models, Shiny-Dashboard, and .github) that contain all files and data used in completing the deliverables of this project will be published and openly available. All code, data, and metadata are thoroughly commented and workflows are cleaned to ensure reproducibility.

Within the Models repository, there are two main folders which are named 'code' and 'data'. The data folder contains three folders, two of which ('fish_data' and 'sediment_data') contain the data that was received from the client, and the third 'data_outputs' folder contains cleaned data and tables that were produced in the process of completing this project. All code files exist in the folder called 'code', which has two folders nested within it: data_cleaning and models. The 'data_cleaning' folder contains a script that includes all necessary data cleaning, and all cleaned data was saved as a data output in the 'data_outputs' folder within the main 'data' folder to streamline data usage within the modeling files. The 'models' folder contains files of model analysis processes. All of the model testing processes are in .Rmd files that contain all the code including a written description of each step, which is described in detail in the *Analysis* section. The session_info.txt file documents all packages and versions used in the processing of code.

The Shiny-Dashboard repository includes all code used for creating the interactive web application. The main files for building the shinydashboard are the ui.R, server.R, and global.R. This repository also includes a 'data' folder with all the data outputs from the analysis in the 'Models' repository that are needed, a 'sediment_data' folder to build the maps of fishing zones, 'polygons' folder with polygons of OEHHA advisory zones, and 'OEHHA' folder with the .csv files of advisories associated with each polygon, and 'fish_image' with images of fish rendered at each prediction. The 'www' folder includes any logos and other images used.

Packages used are included in the appendix.

Structure

The structure of the repo is as follows:

```
Models
│   README.md
│   Models.Rproj
│  .gitignore
│   session_info.txt
└───data
      │ fish_data
      │ sediment_data
      │ data_outputs
└───code
      │ data_cleaning
      │ models
```

The structure of the Models repository.

3. Application User Manual and Technical Guide

The User Manual and Technical Guide are resources for how to use and interpret the information on the application, and outline the functionalities and motivations behind the application. The user manual (section 1.5) includes detailed instructions on getting started, using the application to get predictions, and additional information on uncertainty.

The Technical Guide (section 1.6) is intended for software professionals and the clients to aid in the maintenance of the application after completion of the project. This includes guides on fixing bugs, adding new functionality, or updating the model with new data.

A PDF copy of the user manual will be located in the project's repository and an abbreviated version will be included within the app.

Summary of Testing

1. User Testing

App usability was tested with two rounds of user testing to gain feedback on the accessibility, aesthetics, and comprehensibility. The first round occurred in collaboration with the Fish Contamination Education Collaborative (FCEC) and the clients. Their feedback led to a series of updates to increase accessibility and understanding. This included adding a boundary around the study area to guide users to the valid locations; adding a bar representing risk was added with a gradient from red (unsafe) to green (safe) to make it easier to understand the output of the recommended number of servings per week; adding more information from the FDA and OEHHA about their recommended number of servings; including an image of the selected fish when prediction is rendered to allow self-check for fish identification; updating the language for communicating health recommendations.

Additional suggestions were to add multiple versions of common species names since species may be known by different names in different areas. The current application guides users to a fish identification page if a user cannot identify their catch.

The second round of user testing occurred at the *Ocean Observing in California: Celebrate the Past, Showcase the Present, and Envision the Future* conference hosted by CalCOFI, SCCOOS, and CeNCOOS. Evaluation of the application occurred using a 10-question survey that was given to the client team at Scripps and CalCOFI and the FCEC. Survey questions were designed to receive feedback on user experience and ease of use (eg. "How clear are the instructions on the *Toxin Tracker* page?") as well as how well it informs users (eg. "Were you able to understand the implications of the output DDT concentrations? "). The full list of questions can be found in the appendix. This feedback led to a series of changes to effectively communicate information to a wide range of users. In response to feedback, the team added explorable layers to the map to see historical dumpsites and piers; added a click function to select map location instead of dragging it since it was more intuitive for users; included more clarity for the number of servings a week based on demographic information (eg. women and children); added a reset zoom was added to help navigate the map; and when printing the predictions, a range of predicted DDT values is provided.

Further recommendations were provided by members of the California Environmental Protection Agency and Heal the Bay. These recommendations included a change in aesthetics and labels, as well as expanding the OEHHA advisories to the Santa Catalina Islands. This feedback will be incorporated into the final version of the application, and any other updates that were suggested but not appropriate for the timeline of this project will be communicated to the clients.

User Documentation

1. User Manual

1.1. Access

The application can be accessed on the CalCOFI website accessible here, and here (expires 9/24). The code and data are housed in the SaferSeafood GitHub organization for reproducibility and further project expansion.

1.2. Purpose of the application

The core motivation behind the development of this application is to provide species and location-specific information to seafood consumers and southern California recreational anglers regarding the DDT contamination of sportfish along the California coast. The application prompts the public to input the species of their catch and the location in which they made that catch. These inputs will prompt an output that displays a threshold of DDT accompanied by measures of uncertainty such as a 95% credible interval that the actual calculated value falls within. These values should be used in conjunction with sanctioned organizations such as the Environmental Protection Agency (EPA) and the California Office of Environmental Health Hazard Assessment (OEHHA). Along with a threshold of DDT values, a colored bar is displayed that depicts how unsafe it would be to eat the fish and suggests a range of recommended number of servings to consume each week. Additionally, the application displays OEHHA advisories consumption advisories for their catch. These additional advisories are meant to bring awareness to other contaminants that may impact seafood being deemed as safe or unsafe. Additionally, this application will provide a platform to highlight other work and collaborations in the DDT research space.

1.3. Background

Dichlorodiphenyltrichloroethane (DDT) is an insecticide that is resistant to degradation and can cause increased risks of cancer, premature births, developmental abnormalities, and neurological diseases in humans and animals. A recent rediscovery of a vast barrel field of DDT-laced sludge off the coast of southern California has captured the attention of the public and raised concerns regarding the consumption of contaminated seafood. Alongside direct public health impacts, a decrease in seafood consumers poses a threat to the regional economy and recreational fishing communities. This project helps inform the public and gives users the autonomy to understand the risk and make informed decisions on their seafood consumption. The interactive element of this application will allow users to access predicted concentrations of total DDT in seafood catches based on their location and the specific species of their catch.

1.3.1. Authors

This application was developed for The Scripps Institution of Oceanography and the California Cooperative Oceanic Fisheries Investigation, CalCOFI, as part of a Masters of Environmental Data Science capstone project at the Bren School of Environmental Science & Management, UC Santa Barbara.

The team of graduate students who completed the project are Ben Versteeg, Hope Hahn, Kate Becker, Luna Herschenfeld-Catalán with guidance from the project's Faculty Advisor, Bruce Kendall, and Capstone Advisor, Carmen Galaz García.

1.3.2. Data

The data was compiled from researchers at the Scripps Institution of Oceanography from previously collected fish and sediment contaminant monitoring data from 1998 through 2021. The data came from the following sources: Southern California Bight Regional Monitoring Program, SWAMP Statewide Coastal Screening Survey, SWAMP Coastal Fish Contamination Program, Jarvis et al. 2007, McLaughlin et al. 2021, Southern California Coastal Marine Fish Contaminants Survey, LA County Sanitation District Local Trends Assessment, LA County Sanitation District Seafood Safety Assessment, and City of San Diego POTW Monitoring. This data focuses on the Southern California Bight. The sediment samples were collected via grab samples of top 5 cm at embayment sites and top 2 cm at offshore sites in 2003, 2008, 2013, and 2018. Fish tissue samples were collected off piers and boats, and composites consisted of 5-10 specimens and included only single species per composite. The data was subset to include only sediment and fish samples that explicitly measured DDX (2,4′-DDE, 4,4′-DDE, 2,4′-DDD, 4,4′-DDD, 2,4′-DDT, and 4,4′-DDT). Each species was assigned diet and habitat, and composites were assigned to fishing zones. The DDT concentrations, both lipid-normalized and non lipid-normalized were normalized, and the sediment DDT concentrations were also normalized.

All of the data used for this project has been collected from public data files, and all code and future data and modeling will be available publicly through the team's GitHub organization and repositories. Data outputs of this project have also been published here.

1.3.3. How to use “Toxin Tracker”

These instructions are available under the *Help* tab on the application, and are elaborated in more greater detail below:

Step 1: Navigate to the *'Toxin Tracker'* tab.

Step 2: Use the interactive map to click on a location of your fish catch

On the landing page, users must select the location of their catch using the marker, within the bounds of the study area, indicated by the dotted blue line. Two layers can be toggled on or off that will indicate the location of historical dumpsites and piers of high popularity to provide more context.

Step 3: Select a fish species from the dropdown menu.

Next, users are prompted to input the species of their catch. A list of common names is provided alphabetically in a dropdown menu that supports the search for species name by manually typing the name or scrolling through the list. To select a different species, press the delete button to clear selection, and then type or scroll again.

Step 4: Click the 'Predict DDT' button to receive the forecast.

After filling in a valid location and selecting a species, the user clicks the predict button. A DDT concentration prediction is outputted, as well as a suggested number of servings. Based on the species, an image is generated to allow users to confirm whether they have selected the correct species. The DDT prediction is provided by a range indicating a 95% credible interval showcasing uncertainty.

1.3.4. Interpreting DDT predictions

Based on the DDT prediction, the recommended number of servings are also provided with a color scale making it easy to communicate whether the concentration prediction is in a safe to eat (green), moderately safe to eat (orange), or unsafe / do not eat (red) range. The information next to *DDT Prediction Results* provides further information on the definition of a serving size: ‘A serving size is defined by the OEHHA as an 8 oz skinless filet.’ Additionally, health advisories from the OEHHA, that also include mercury and PCB advisories, are included based on the specified location and species. The public should refer to the advisories with the lowest recommended number of servings and proceed with caution. These recommendations are also accompanied by a note that the FDA only recommends eating 2 servings of fish per week. The generated prediction results are intended for educational and informational purposes and are not intended to replace official health advisory information. Users can access additional information about ongoing DDT research under the *Resources* tab.

1.4. Sections

1.4.1. Toxin Tracker

Once the application is open you are prompted with the Toxin Tracker for an interactive experience that yields on-the-spot and accurate contaminant calculations and advisories. The tracker provided predicted DDT concentrations, health advisory and a recommended number of servings, as well as information regarding the sources of DDT.

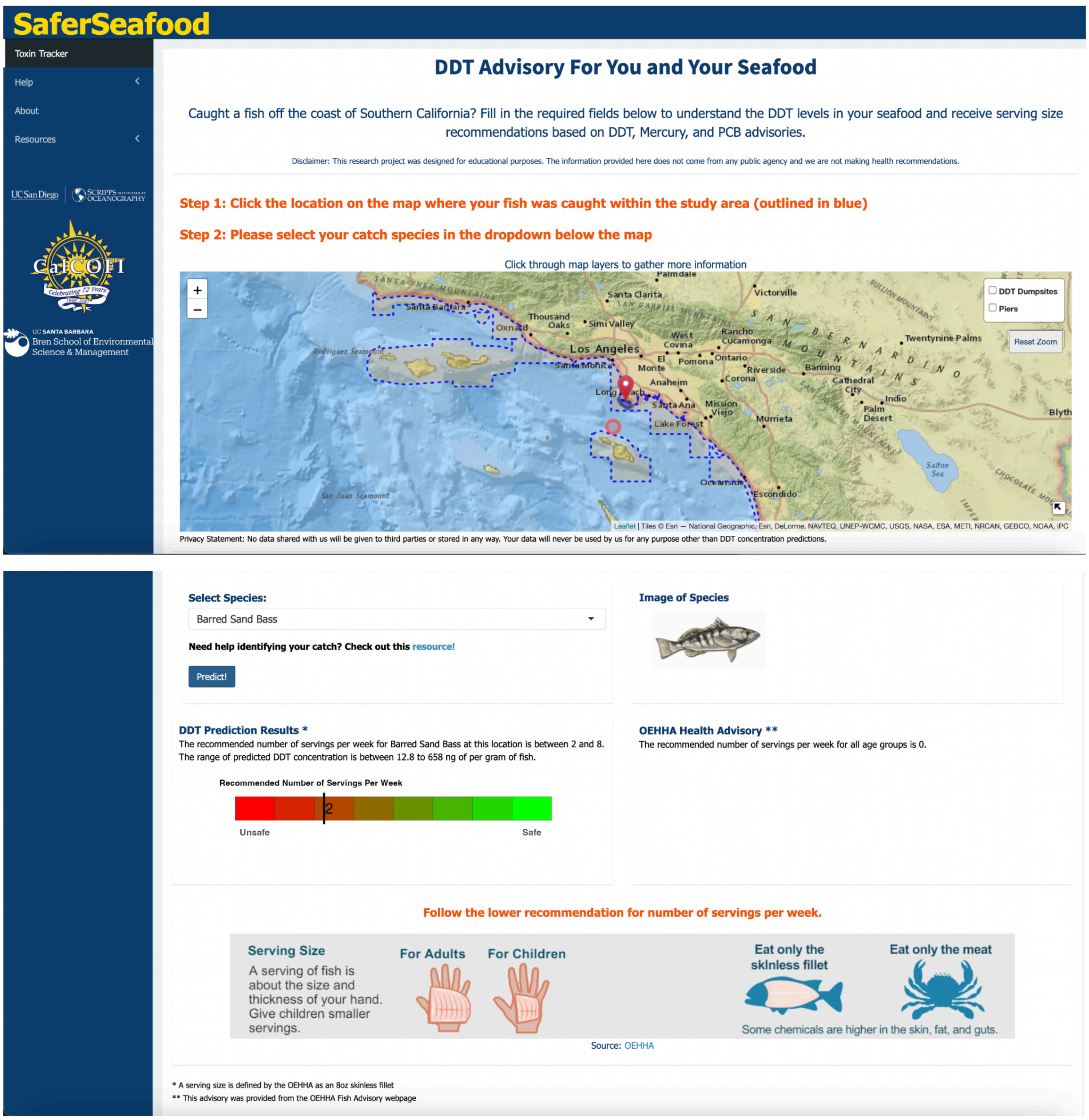

Figure: Toxin Tracker, Main Page

1.4.2. Help

This page includes steps for how to use and interpret the *Toxin Tracker* tool, and troubleshooting any error messages.

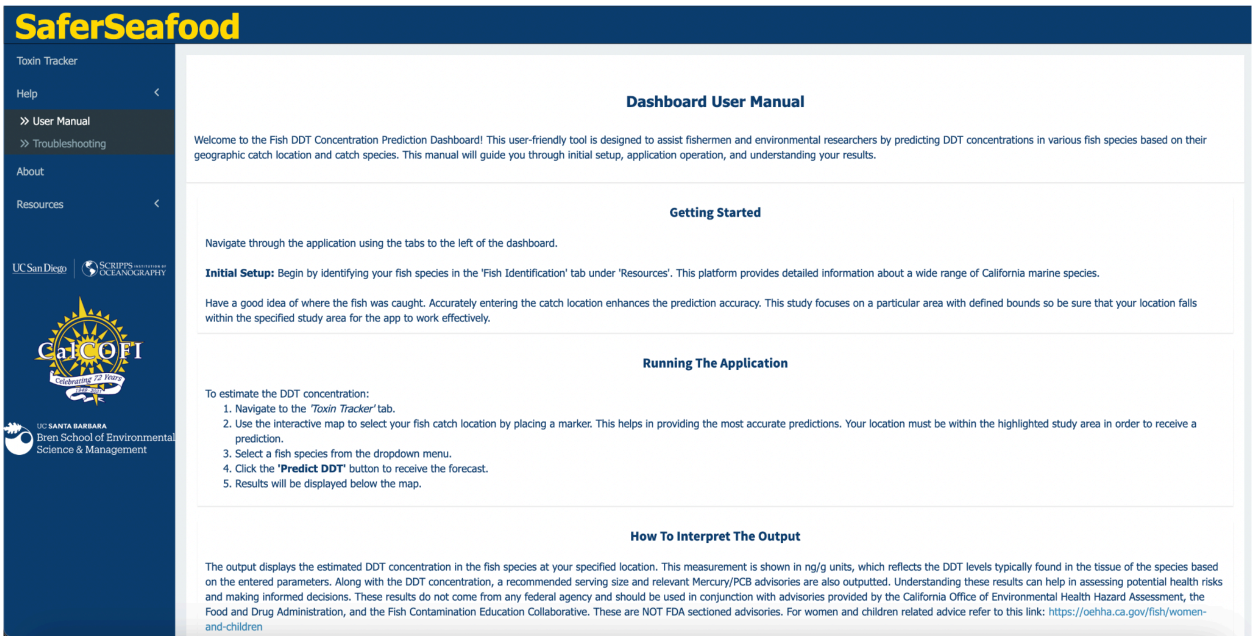

Figure: User Manual

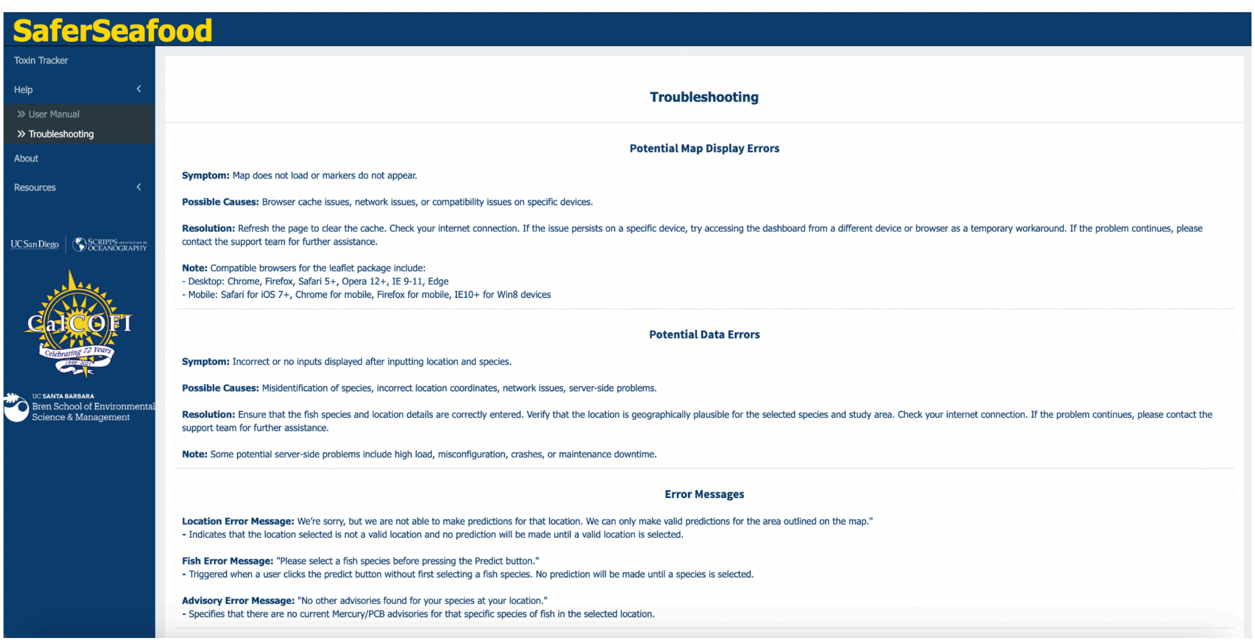

Figure: Troubleshooting

1.4.3. About

The following page includes two tabs: *Project Background* and *Data*. The project background highlights the motivation and the background of this project, and further information on the authors. The following tab elaborates on the data employed in this project such as data background, data accessibility, citations, and disclaimers.

Figure: Project Background Page

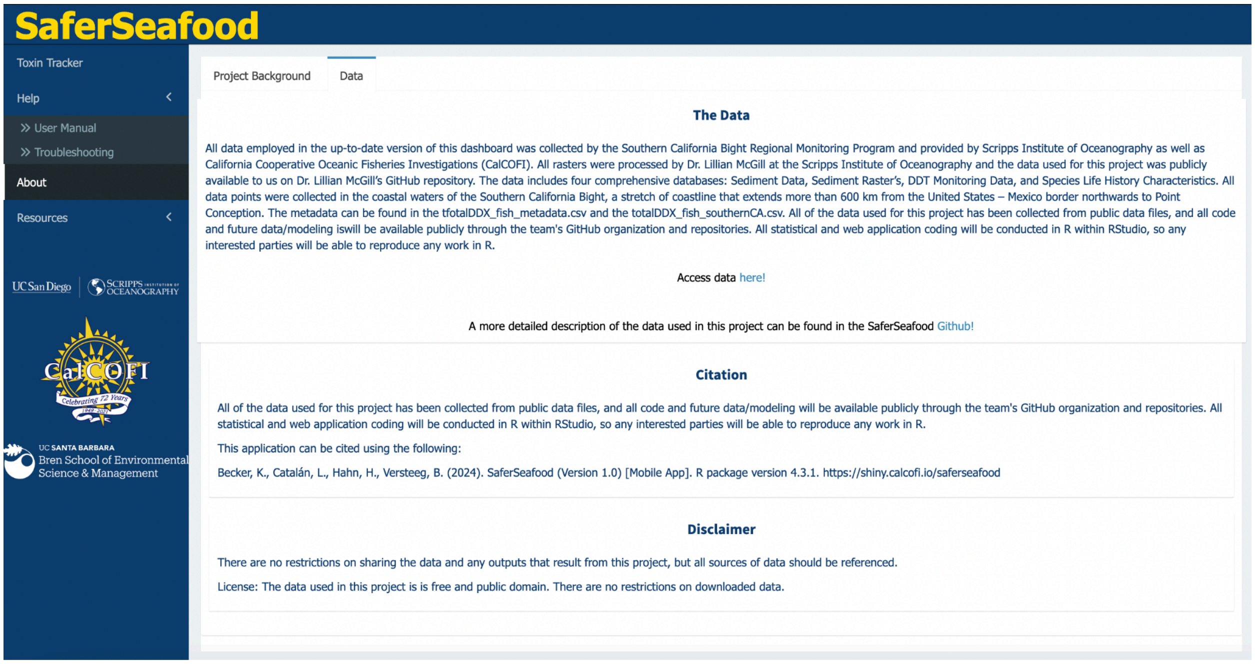

Figure: Data Page

1.4.4. Resources

The final tab of this dashboard includes two additional tabs: *Fish Identification* and *DDT research.* The Fish Identification tab aids in species lookup. If you're still unsure of your catch it provides a link to an external site California Marine Species Portal, where you can find detailed information about various fish species. This resource may assist you in identifying the fish species by their common names, scientific names, and visual characteristics. Further research and educational collaborators working on understanding the human and ecological impacts of the recently discovered DDT dumpsite will be highlighted here. The web application provides a platform that enables researchers to connect with communities in a meaningful way.

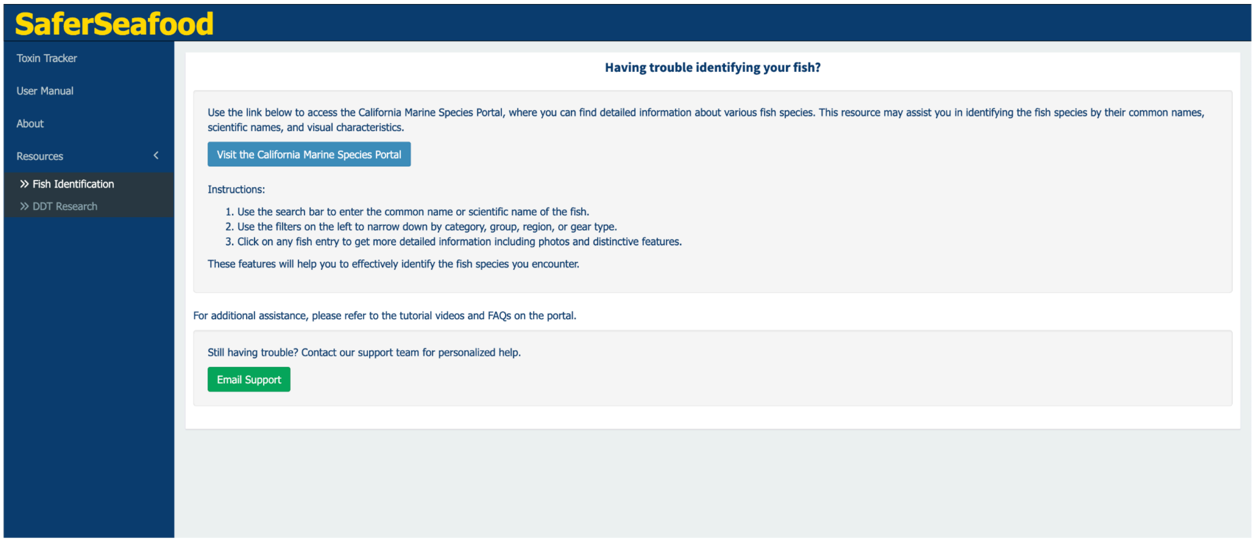

Figure: Fish Identification Tab

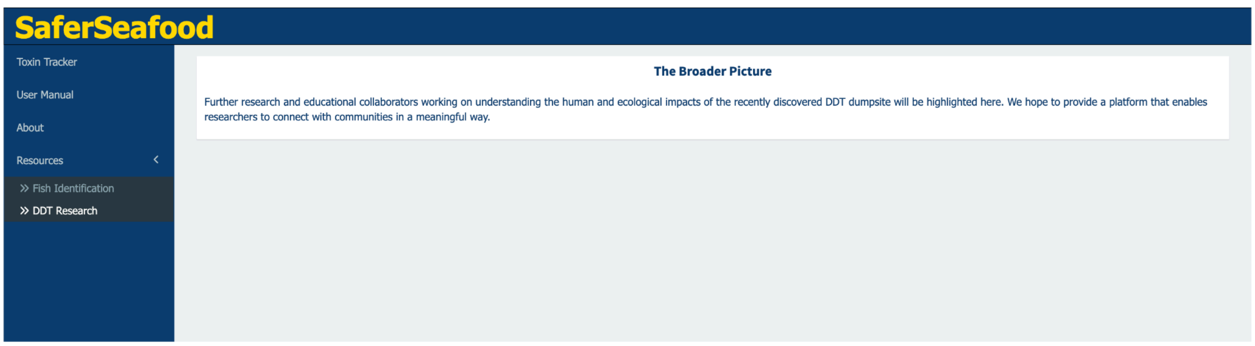

Figure: DDT Research Page

2. Technical Guide for Shiny Dashboard

This section of the report serves as a comprehensive technical guide for the Shiny dashboard designed to predict DDT concentrations in fish based on user input. It is intended for software professionals maintaining the application and for clients who need a deeper understanding of its functionalities and troubleshooting procedures.

2.1. Troubleshooting

Common Issues and Solutions

1. **Map Display Errors**:

   Symptom: Map does not load or markers do not appear.

   Possible Causes: Browser cache issues or compatibility issues on specific devices.

   Resolution: Refresh the page to clear the cache. If the issue persists on a specific device, try accessing the dashboard from a different device or browser as a temporary workaround. If the problem continues, please contact the support team for further assistance.

   Note: Compatible browsers for the leaflet package include

   - Desktop: Chrome, Firefox, Safari 5+, Opera 12+, IE 9-11, Edge
   - Mobile: Safari for iOS 7+, Chrome for mobile, Firefox for mobile, IE10+ for Win8 devices

2. **Data Errors**:

   Symptom: Incorrect or no inputs displayed after inputting location and species.

   Possible Causes: Misidentification of species, incorrect location coordinates, network issues, server-side problems.

   Resolution: Ensure the species lookup table is accurate and up-to-date and ensure coordinates are entered correctly. Regular server maintenance, load balancing, and proper configuration management can help mitigate server-side problems.

   Note: Some potential server-side problems include high load, misconfiguration, crashes, or maintenance downtime.

3. **Performance Issues**:

   Symptom: Slow response times or dashboard freezes.

   Possible Causes: High server load, or large data handling.

   Resolution: Optimize data queries and consider server scaling options if consistently high traffic is observed. Think about maybe storing predictions in a lookup table if the prediction function is too computationally intensive for the server (Will need to update every time new data is imputed into the model).

4. **Server Downtime**:

   Symptom: Dashboard is inaccessible.

   Possible Causes: Scheduled maintenance, unexpected server crashes.

Resolution: Check server status and logs. Restart the server if necessary. Ensure there is a downtime notification system in place to inform users.

User-Friendly Features for Error Minimization

To minimize user errors and enhance the user experience, the following features are integrated:

- **Interactive Map**: Has users precisely pinpoint their fishing location which, in turn, reduces geographical input errors.
- **Species Dropdown Table**: Has users more easily input their specific fish species without worrying about typos or mislabeling the species.
- **Outline of Valid Locations**: A border of the valid locations is shown on the map in order to show users what locations work for the prediction tool.
- **Fish Species Photo Output**: Acts as a check for users to make sure that they inputted the correct fish species.

Error Messages

"We're sorry, but we are not able to make predictions for that location. We can only make valid predictions for the area outlined on the map.": Indicates that the location selected is not a valid location and no prediction will be made until a valid location is selected.

"Please select a fish species before pressing the Predict button.": Triggered when a user clicks the predict button without first selecting a fish species. No prediction will be made until a species is selected.

"No other advisories found for your species at your location.": Specifies that there are no current OEHHA advisories for that specific species of fish in the selected location.

2.1.1. How to maintain the application

Maintenance and Updates

Data Standards:

All tabular data follows tidy data standards (Wickham, 2017); each .csv file has each variable as a single column, each observation has its own row, and every cell has a single value.

This structured format ensures consistency and ease of data manipulation within the application.
Additionally, raster data are stored as .gri and .grd files, maintaining uniform resolution and dimensions across datasets. Adhering to these standards facilitates seamless integration and analysis of spatial data within the dashboard.

The following variables must be present in the dataset to update the models and feed the updated models into the app:

- Fish Species: The common and scientific names.
- Location Coordinates: Latitude and longitude.
- DDT Concentration in Sediment: Measured in ng/fishing zone.
- DDT Concentration in Fish Tissue: Measured in ng/g.
- Trophic Category: Herbivore, Primary Carnivore, Secondary Carnivore, Tertiary Carnivore.
- Feeding Position: Benthopelagic, Midwater, Benthic, Pelagic.
- Date of Sampling: The date when the sample was collected.
- Optional Variables: Length, weight, and other relevant biological data of the fish if available.

GitHub Repository:

- Our dashboard’s source code and documentation are maintained on GitHub. This repository includes detailed commit histories, implementation strategies for new features or fixes, and a README that explains the model selection and data integration process.

Routine Maintenance Checklist:

- Update Dependencies: Regularly check and update R packages to their latest versions to ensure compatibility and security. Note: Updating packages can sometimes cause compatibility issues or break the app, so always test updates in a development environment before applying them to the production server.
- Backup Data: Schedule regular backups of the data and configuration files.
- Check Server Logs: Monitor server logs for any unusual activity or errors.
- Security Audits: Perform routine security audits to protect against vulnerabilities.
- Performance Testing: Periodically perform load testing to ensure the app can handle expected traffic volumes.

3. Expansion and Collaboration

Data Updates and Enhancements:

- Species Expansion: There are plans to expand the database to include more fish species and updated DDT data as new research becomes available.
    - Workflow to Ingest New Data and Update the App:
        1. Data Collection: Gather new data from reliable sources.
        2. Data Validation: Verify the accuracy and completeness of the new data.
        3. Data Formatting: Format the data to match the existing data structure (e.g., tidy data standards).
        4. Model Update: Integrate the new data into the model training pipeline. Retrain the model if necessary.
        5. Testing: Validate the updated model with a subset of the new data to ensure it performs as expected.
        6. Deployment: Deploy the updated model to the production environment.
        7. Monitor: Continuously monitor the performance of the updated model.

- Collaborative Efforts: This application is open to collaborating with environmental agencies and research institutions to enrich the dataset and analytical capabilities.
- Language Translations: Many anglers speak Spanish, so having a translation of the application and advisories in Spanish or other languages increases accessibility.
- Outreach Efforts: Some ideas include:
    - Posting QR codes to the application on piers or near the shore (bait shops and/or charter boats).
    - Collaborating with angler outreach programs to make this application more known (heal the bay, surfrider, aquariums).

Potential Future Challenges:

- Scalability: As user base and data volume grow, the dashboard may require more robust server infrastructure or cloud services to handle increased loads.
- Data Accuracy: Continuous verification of the model's predictions which involves regularly comparing the model's predictions with actual measured data to ensure the predictions remain accurate over time. This is different from simply updating the models with new data; it requires ongoing validation against real-world outcomes to detect and correct any drifts or inaccuracies that may arise due to changes in underlying patterns or data quality.

Additional Information

1. Data limitations

When updating the model, length and weight seemed like suitable predictors to add to the model. However, the data provided by the client does not include the length and weight of most of the observations, so it was not feasible to predict DDT concentrations with these predictors. Additionally, the data is limited to specific latitude and longitudes, so predictions will only exist within these ranges. Sediment DDT concentrations are also predicted using a model due to the lack of sampling in every location within the ranges, so there may be a level of uncertainty in using sediment DDT as a predictor in the model.

2. Representing uncertainty

Using statistical models to make predictions will always result in some level of uncertainty as it is impossible to make correct predictions 100% of the time. However, when creating predictions, the level of error is also calculated, which allows an understanding of how accurate the model is at predicting the targeted variable; which in this case is DDT concentration in fish. The results of this project will influence recreational fishermen deciding whether a piece of seafood is safe to eat or not. As this directly influences the health of the consumers, it is important to make sure that the predictions are as accurate as possible. Additionally, to enhance transparency and allow people to make the best-educated decision regarding seafood, a level of uncertainty and risk will also be conveyed to the web application users by providing a range of values that the DDT concentration is likely to fall between as well as a range of recommended number of servings that is likely safe. Depending on different situations - whether an individual is pregnant, elderly, or a child for example - there will be different OEHHA advisories of how risky a piece of seafood is.

Archive Access

The original data and metadata is published within the client's public GitHub repository. This data as well as code and data outputs from this project will be published within the capstone group's GitHub repository, which will be transferred to the client, and available for public access. The tabular data outputs were also archived in Dryad.

References

California Cooperative Oceanic Fisheries Investigations. (n.d.). *CalCOFI: California cooperative oceanic fisheries investigations*. Retrieved June 3, 2024, from https://calcofi.org/.

California Department of Fish and Wildlife. (n.d.). *CA Marine Species Portal*. Retrieved June 3, 2024, from https://marinespecies.wildlife.ca.gov/.

Environmental Protection Agency. (1988). *Guidance for Assessing Chemical Contamination for Use in Fish Advisories* (Volume III: Overview of Risk Management). Retrieved June 3, 2024, from https://semspub.epa.gov/work/HQ/189675.pdf.

Leeworthy, V., & Schwarzmann, D. (2015). Economic Impact of the Recreational Fisheries on Local County Economies in California's National Marine Sanctuary 2010, 2011 and 2012. *Marine Sanctuaries Conservation Series ONMS-2015-07. U.S. Department of Commerce, National Oceanic and Atmospheric Administration, Office of National Marine Sanctuaries, Silver Spring, MD. 27 pp.*

McGill, L., Sleugh, T., Petrik, C., Schiff, K., McLaughlin, K., Aluwihare, L., & Semmens, B. (2024). *The persistent DDT footprint of ocean disposal and ecological controls of bioaccumulation in fishes*. Unpublished manuscript, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA, USA, and Southern California Coastal Water Research Project, Costa Mesa, CA, USA.

National Oceanic and Atmospheric Administration. (2015). *New NOAA report shows significant economic benefits of recreational fishing in California's national marine sanctuaries*. Retrieved June 3, 2024, from https://sanctuaries.noaa.gov/news/press/2015/rec-fishing-california.html.

Office of Environmental Health Hazard Assessment. (n.d.). *How to follow advisories*. Retrieved June 3, 2024, from https://oehha.ca.gov/fish/how-follow-advisories.

Quimby, B., Crook, S. E., Miller, K. M., Ruiz, J., Lopez-Carr, D. (2020). "Identifying, defining and exploring angling as urban subsistence: Pier Fishing in Santa Barbara, California." *Marine Policy*, vol. 121, Nov. 2020, p. 104197, https://doi.org/10.1016/j.marpol.2020.104197.

Scripps Institution of Oceanography. (2024, June 6). https://scripps.ucsd.edu/

Stevenson, C., Sikich, S.A., Gold, M. (2012). “Engaging Los Angeles County subsistence anglers in the California Marine Protected Area Planning Process.” *Marine Policy*, vol. 36, no. 2, Mar. 2012, pp. 559–563, https://doi.org/10.1016/j.marpol.2011.08.001.

Xia, R. “Here’s What We Know about the Legacy of DDT Dumping off L.A.’s Coast.” *Los Angeles Times,* Los Angeles Time. Retrieved 7, June 2024, from https://www.latimes.com/environment/story/2022-05-18/heres-what-we-know-about-the-legacy-of-ddt-dumping-near-catalina .

Appendix

1. Packages

Model Script Packages:
Tidyverse, Tidymodels, Brms, Rfishbase, Projpred, Multilevelmod, Rstanarm, Posterior, Bayesplot, Pacman, GeojsonR, Factoextra, Sf, Dplyr, Ggplot2, Maps, Fields, Raster, MuMIn, Lubridate, Tidyr, Ggh4x, Lme4, SdmTMB, Inlabru, Cowplot, INLAutils, Marmap, SjPlot, Rgeos, Plyr, Tidybayes, Brms, Loo, Ggeffects, DHARMa, DHARMa.helpers

Packages Used in Dashboard Script:
Shiny, Shinydashboard, Shinyjsm, ShinyWidgets, ShinyBS, Leaflet, Leaflet.extras, Shinycssloaders, Extrafont, Showtext, Markdown, Rfishbase, Sf, Skimr, Tidymodel, Caret, Corrplot, sjPlot, Rstanarm, Terra, Reactlog, Geos, Tidyverse, Raster, Sp, Brms

2. Survey Questions

   a. How clear are the instructions on the "Toxin Tracker" page? (1-5)
   b. If unclear, please explain or provide suggestions on how to improve clarity.
   c. Were you able to navigate to your desired location easily? (1-5)
   d. Are there particular landmarks on the map that would be helpful? (eg. piers, dumpsite, harbors)
   e. Were you able to understand the implication of the output DDT concentrations?
   f. Is there any other information you would like to have presented when outputting DDT concentrations?
   g. Imagine you are trying to figure out what the fish in front of you is. Navigate to the fish identification tab, and follow the instructions. Was this helpful?
   h. Please comment on your experience with the error messages. Were they clear and informative? If not, how could they be improved?
   i. Was the separation of descriptions and information on the User Manual page intuitive or would it be more helpful to have that information on the main page with the prediction results?
   j. Do you have any additional comments, suggestions, or ideas?

3. Contact List

Here is a list of contacts that provided guidance and support throughout this process. They have expressed interest in continuing to be updated on the progress of the application, and could help with distribution in the next phases.

Organization	Names & Emails
EPA	Renee Jordan Ward JordanWard.Renee@epa.gov
Scripps	Ian Brunjes ibrunjes@ucsd.edu
OEHHA	Wesley Smith Wesley.Smith@oehha.ca.gov Huyen Tran Pham Huyen.Tran.Pham@oehha.ca.gov
FCEC	Amalia Aruda Almada amaliaal@usc.edu Maria Madrigal mdmadrig@usc.edu
Heal the Bay	Katherine Pease kpease@healthebay.org