

Marine Sensitivity

Project Documentation

Benjamin D. Best

Timothy P. White

2026-08-10

Table of contents

Preface	3
Interactive Apps	3
Documentation	4
How to Cite	4
Acknowledgments	4
1 Introduction	5
1.1 Overview	6
I Science	8
2 Science	9
2.1 Conceptual Framework	9
2.2 Cell Vulnerability	9
2.3 Sensitivity Components	10
2.4 Ecoregional Rescaling	10
2.4.1 Geographic Scope	11
3 Data Sources	13
3.1 Dataset Overview	13
3.2 Source Datasets	14
3.2.1 AquaMaps SDM (<code>am_0.05</code>)	14
3.2.2 NMFS Core Area (<code>ca_nmfs</code>)	14
3.2.3 NMFS Critical Habitat (<code>ch_nmfs</code>)	14
3.2.4 FWS Critical Habitat (<code>ch_fws</code>)	15
3.2.5 FWS Range (<code>rng_fws</code>)	15
3.2.6 BirdLife Range (<code>bl</code>)	15
3.2.7 IUCN Range (<code>rng_iucn</code>)	15
3.2.8 SWOT+DPS Turtle Range (<code>rng_turtle_swot_dps</code>)	16
3.3 Merged Model (<code>ms_merge</code>)	16
3.4 Standard Grid	17
3.5 Primary Productivity	17
4 Taxonomy	19
4.1 Taxonomic Authorities	19

4.2	ID Resolution Cascade	20
4.3	Species Categories	20
4.4	Data Quality	21
4.4.1	Duplicate Resolution	21
4.4.2	Synonym Handling	21
4.4.3	Valid Species Filter (<code>is_ok</code>)	21
4.4.4	From source taxa to valid species	22
4.5	Key Function	23
5	Extinction Risk	24
5.1	Regulatory Framework	24
5.1.1	Endangered Species Act (ESA)	24
5.1.2	Marine Mammal Protection Act (MMPA)	24
5.1.3	Migratory Bird Treaty Act (MBTA)	25
5.1.4	IUCN Red List	25
5.2	Scoring Methodology	25
5.2.1	US-listed Species	26
5.2.2	Species not federally listed (IUCN)	26
5.2.3	Scale Note	27
5.3	Implementation	27
6	Model Merging	28
6.1	Pipeline Overview	28
6.1.1	Step 1: Gather Source Models	28
6.1.2	Step 2: MAX Across Datasets	28
6.1.3	Step 2b: Turtle Multiplicative Merge	30
6.1.4	Step 3: Spatial Masking	30
6.1.5	Step 4: MMPA Spatial Floor	31
6.1.6	Step 5: MBTA Spatial Floor	32
6.1.7	Step 6: Persist Results	32
6.2	Valid Species Filter	32
7	Scoring	33
7.1	Cell-Level Scoring	33
7.1.1	Two Scoring Approaches	34
7.2	Metric Keys	34
7.3	Ecoregional Rescaling	35
7.3.1	Geographic Scope	35
7.3.2	Cross-Ecoregion Program Areas	35
7.4	Primary Productivity Scoring	36
7.5	Zone Aggregation	37
7.6	Visualization	37
7.6.1	Flower Plot	37

7.6.2	Treemap	39
7.6.3	Interactive Maps	41
8	Scoring Walkthrough	42
9	Stressors	43
9.1	Offshore Wind Energy	43
9.2	Oil & Gas	43
II	Software	45
10	Software	46
10.1	Software Overview	46
10.1.1	Interactive Applications	47
10.1.2	Overcoming Challenges with Large Spatial Data	47
10.1.3	Github Repositories	51
10.1.4	Software Components	54
11	Server	55
11.1	Setup	55
11.1.1	launch instance	55
11.1.2	ssh to server	56
11.1.3	install docker	57
11.1.4	Backup /share with snapshots	58
11.2	Docker compose	59
11.3	DNS	66
11.4	Caddyfile	67
11.5	Services	72
11.5.1	App + data plane	72
11.5.2	Admin	73
11.5.3	Legacy (carried for backward-compat; not in the critical path)	74
12	Database	75
12.1	Table and column naming conventions	75
12.2	Core tables	76
12.2.1	How the core tables feed the serving tier	77
13	Workflows	79
13.1	Ingest workflows (populate <code>sdm.duckdb</code>)	79
13.2	Scoring workflow (<code>cell_metric</code> + <code>zone_metric</code>)	80
13.3	Derived artifacts for the serving tier	80

14 APIs	81
14.1 msens TiTiler factory	81
14.1.1 Endpoints	81
14.1.2 Caching	82
14.1.3 Derived COG: <code>r_cellid.tif</code>	83
14.2 Custom plumber API	83
14.3 Legacy APIs (deprecated)	83
15 Libraries	85
15.1 Map-tile helpers (as of 0.3.4)	85
16 Apps	86
16.1 Current Applications	86
16.1.1 Scores — mapper of aggregate sensitivity scores	86
16.1.2 Species — mapper of individual species distributions	87
16.2 Legacy Applications	87
17 Apps User Guide	88
17.1 Scores app	88
17.1.1 1. Study Area	88
17.1.2 2. Spatial Units	88
17.1.3 3. Layer Selection	91
17.1.4 4. Map tab	91
17.1.5 5. Layers control	91
17.1.6 6. Go to location	94
17.1.7 7. Full screen	94
17.1.8 8. Zoom in / out	94
17.1.9 9. Plot of Scores tab	97
17.1.10 10. Table of Species tab	97
17.1.11 11. Report tab	97
17.2 Species app	99
17.2.1 1. Select a Species	99
17.2.2 2. Mask Selection	99
17.2.3 3. Layer Selector	102
17.2.4 4. Species Map	102
17.2.5 5. Species Info	102
18 Docs	105
19 Summary	106
19.1 Key Findings	106
19.2 Future Directions	106
References	108

Appendices	110
Glossary	110
III Applications (Current)	112
A Scores — mapper of aggregate sensitivity scores	113
A.1 Overview	113
A.2 Features	113
A.2.1 Map Display	113
A.2.2 Sensitivity Metrics	116
A.2.3 Flower Plot	116
A.2.4 Species Table	116
A.2.5 Data Export	116
A.3 Generating Screenshots	116
A.4 Source Code	116
B Species — mapper of individual species distributions	117
B.1 Overview	117
B.2 Features	117
B.2.1 Species Selection	117
B.2.2 Model Display	117
B.2.3 Species Info Panel	119
B.2.4 Deep-Linking	119
B.3 Generating Screenshots	119
B.4 Source Code	119
IV Applications (Legacy)	120
Areas of Interest	121
Bird Hotspots	122
Regional Map	123
C Scores — mapper of aggregate sensitivity scores	124
C.1 Overview	124
C.2 Features	124
C.2.1 Map Display	124
C.2.2 Sensitivity Metrics	127
C.2.3 Flower Plot	127
C.2.4 Species Table	127

C.2.5 Data Export	127
C.3 Generating Screenshots	127
C.4 Source Code	127
Distributions, Vector	128
Distributions, Raster	129
Vulnerability Mapper	130

Preface

The **Marine Sensitivity Toolkit** (MST) is a cloud-native system developed for the Bureau of Ocean Energy Management (BOEM) to assess the sensitivity of marine ecosystems to offshore energy development. It integrates 16,153 species distribution models with extinction risk assessments across US waters (9,230 of which fall within the 20 BOEM Program Areas of the current program cycle).

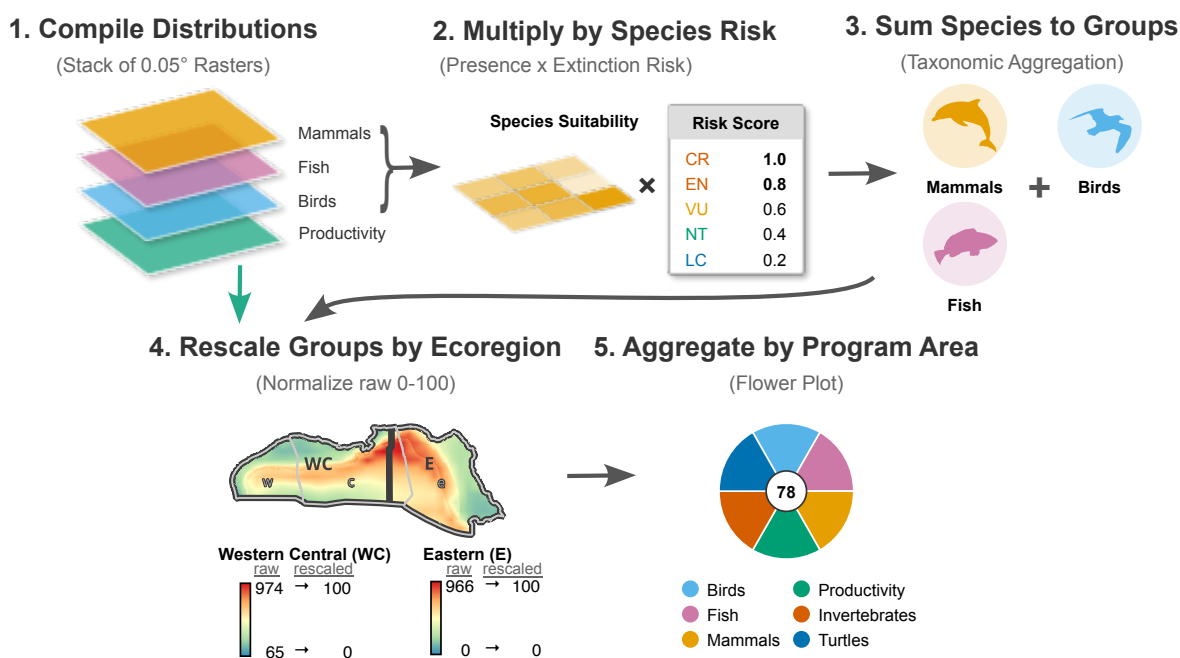


Figure 1: Overview of the MST methodology: species distribution models from 7 source datasets are merged, weighted by extinction risk scores (incorporating ESA, MMPA, MBTA, and IUCN protections), rescaled by ecoregion, and aggregated to management zones for visualization.

Interactive Apps

- **Composite Scores** — explore sensitivity scores across BOEM Program Areas with flower plot summaries

- **Species Distribution** — view individual species distribution models and habitat suitability maps
- **Scoring Walkthrough** — step through the scoring pipeline interactively, from input rasters to final scores

Documentation

This book documents the scientific methodology, software infrastructure, and interactive applications that comprise the MST. It is organized in two parts:

- **Chapter 2** — data sources, taxonomic integration, extinction risk scoring, model merging, and sensitivity scoring methodology
- **Chapter 10** — server architecture, databases, workflows, APIs, R libraries, and interactive applications

How to Cite

Best, B.D., T.P. White (2026). Marine Sensitivity Toolkit: Project Documentation. Bureau of Ocean Energy Management, Environmental Studies Program.
<https://marinesensitivity.org/docs/>

Acknowledgments

This work is funded by the [BOEM Environmental Studies Program](#) (BOEM-ESP). We thank the data providers whose open datasets make this work possible: [AquaMaps](#), [BirdLife International](#), [IUCN Red List](#), [NOAA Fisheries](#), and [U.S. Fish & Wildlife Service](#).

1 Introduction

The Marine Sensitivity Toolkit (MST) is a cloud-native system developed for [BOEM](#) (Bureau of Ocean Energy Management) to assess the sensitivity of marine species to offshore energy development, whether oil & gas or wind. The Outer Continental Shelf Lands Act (OCSLA), Section 18(a)(2)(G), mandates that BOEM consider “the relative environmental sensitivity and marine productivity of different areas of the OCS” when planning the timing and location of offshore energy activities. The MST operationalizes this mandate by combining the best available species distribution data with extinction risk assessments to map areas of the ocean that are most sensitive to human activities.

The MST marks a significant advancement over prior RESA methodologies (Morandi et al. 2018). Earlier approaches often relied on aggregated data from a limited set of broad species groups and surrogate species, lacking spatially explicit information for individual organisms. As a result, previous assessments were typically coarse and areawide, frequently missing critical ecological variation and fine-scale patterns across the OCS. In contrast, the MST uses a high-resolution 0.05° grid (~4 km per cell) on a species level to capture fine-scale conservation concerns and ecological patterns across US waters.

The current system integrates **16,153 valid marine species** across 7 source datasets, mapped at **0.05° resolution** (~4 km cells) across the full US Exclusive Economic Zone (EEZ) study area; of these, **9,230** have modeled distributions that fall within the **20 BOEM Program Areas** of the current program cycle. Species distributions are weighted by extinction risk scores that incorporate protections under the Endangered Species Act (ESA), Marine Mammal Protection Act (MMPA), and Migratory Bird Treaty Act (MBTA), as well as IUCN Red List assessments. The merged distribution of sensitivity per species is masked to expert ranges where available — species whose global range maps fall entirely outside the US EEZ are excluded, ensuring that only species with expert-confirmed presence in US waters contribute to scores.

This is a process, not a product. Information is imperfect, especially given the large expanse of US waters. Distributions and abundance of species change, modified increasingly by climate change and human activities. Knowledge on species sensitivities continues to expand with more research. Methods for both modeling and distributing this information continue to improve. In alignment with [Executive Order 14303](#), we aim to provide a transparent and reproducible process that can be regularly updated as new data and methods become available.

1.1 Overview

The MST analytical pipeline proceeds through the following stages (Figure 1.1):

1. **Data Sources:** 7 source datasets of species distributions (AquaMaps SDM, NMFS Core Area, NMFS Critical Habitat, FWS Critical Habitat, FWS Range, BirdLife Range, IUCN Range)
2. **Taxonomy Matching:** species names reconciled across WoRMS, GBIF, ITIS, IUCN, and BirdLife taxonomic authorities
3. **Model Merging:** per-species models merged by taking the MAX suitability across datasets, constrained by IUCN/Critical Habitat spatial masks, with MMPA/MBTA spatial floors
4. **Cell Scoring:** extinction risk weighted by species presence, summed by species category (bird, coral, fish, invertebrate, mammal, reptile/turtle)
5. **Ecoregional Rescaling:** scores normalized to [0–100%] within each BOEM ecoregion for meaningful cross-region comparison
6. **Zone Aggregation:** area-weighted averages to program areas, subregions, and ecoregions
7. **Visualization:** interactive flower plots, treemaps, and map applications

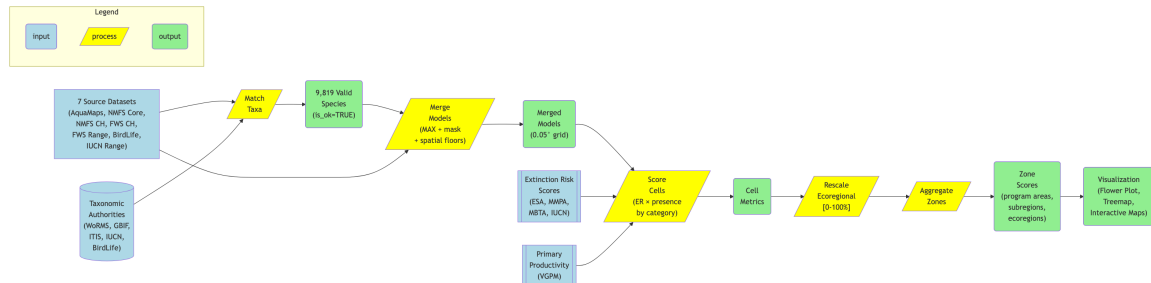


Figure 1.1: Flowchart of the MST analytical pipeline: from source datasets and taxonomic matching through model merging, scoring, and visualization.

Figure 1.2 illustrates the full methodology at a glance. On the left, species distribution data flows in from 7 source datasets and 5 taxonomic authorities. In the center, models are merged per species and weighted by extinction risk scores derived from ESA, MMPA, MBTA, and IUCN regulatory frameworks. On the right, cell-level scores are rescaled within BOEM ecoregions and aggregated to program areas for visualization in the interactive mapping applications.

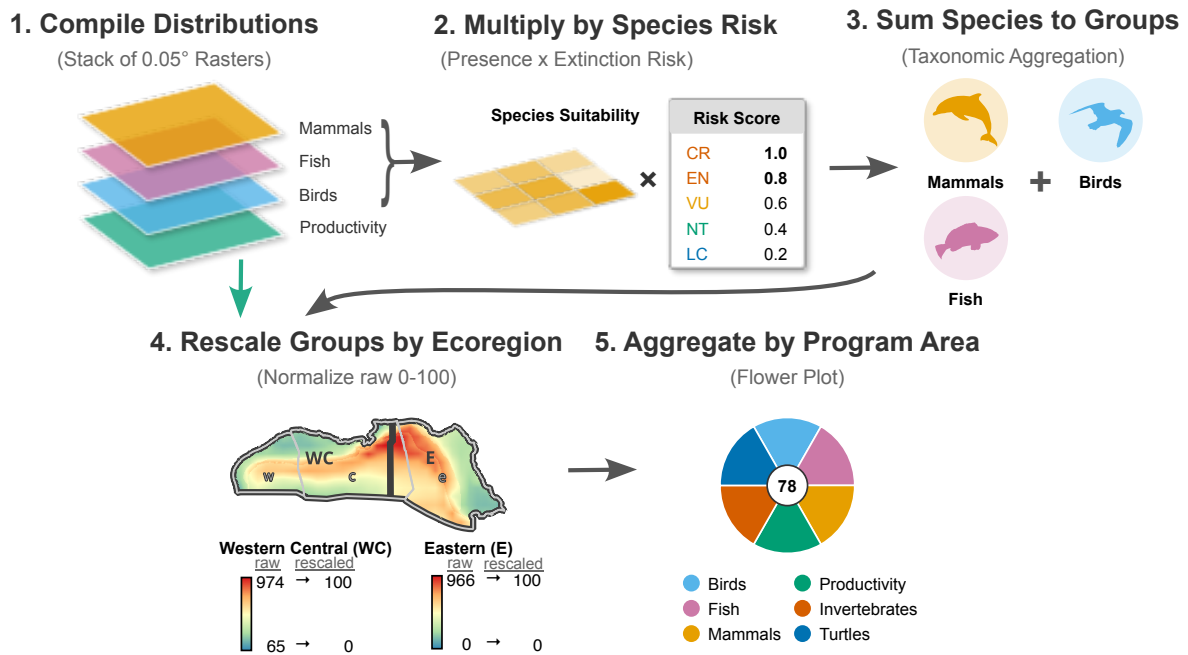


Figure 1.2: Overview of the Marine Sensitivity Toolkit methodology showing the integration of species distribution models, extinction risk scoring, ecoregional rescaling, and spatial aggregation into interactive visualizations.

Part I

Science

2 Science

Interactive Walkthrough

New to the MST scoring pipeline? The [Scoring Walkthrough](#) provides a step-by-step interactive visualization of how input rasters are merged, rescaled by ecoregion, and averaged into final sensitivity scores.

2.1 Conceptual Framework

The term vulnerability (V) is a function of exposure (E), sensitivity (S) and adaptive capacity (A) (Equation 2.1).

$$V = f(E, S, A) \quad (2.1)$$

The more exposed and sensitive an area is—and the less able it is to recover—the more vulnerable it is to impacts from offshore human activities. The MST currently focuses on the **sensitivity** component, quantifying the intrinsic biological sensitivity of marine ecosystems based on the species present and their conservation status. Exposure and adaptive capacity components are planned for future phases (see Chapter 9).

2.2 Cell Vulnerability

For a given cell c , the vulnerability score for a species group g is the sum across all S_g species of the product of species presence probability p_{sc} and extinction risk weight w_s (Equation 2.2):

$$v_c = \sum_{s=1}^{S_g} p_{sc} \cdot w_s \quad (2.2)$$

where:

- v_c = vulnerability score of cell c for species group g

- p_{sc} = probability of species s presence in cell c (0–1 scale, from merged species distribution models)
- w_s = extinction risk weight for species s (0–1 scale, derived from ESA/MMPA/MBTA/IUCN status; see Chapter 5)
- S_g = total number of species in taxonomic group g (e.g., bird, fish, mammal, coral, invertebrate, reptile/turtle)

In other words, if a cell has many species that are both likely to be present and at high risk of extinction, it gets a higher sensitivity score. This helps identify places where rare or threatened species are concentrated.

2.3 Sensitivity Components

The MST assesses sensitivity through three components:

1. **Species** — by taxonomic category (bird, coral, fish, invertebrate, mammal, reptile/turtle): extinction risk weighted species distribution models across 16,153 valid species (see Chapter 3, Chapter 5)
2. **Habitats** — benthic habitats such as coral reefs, seamounts, and hydrothermal vents (planned for future phases)
3. **Primary Productivity** — satellite-derived net primary productivity from the Vertically Generalized Production Model (VGPM), measured as metric tons C km⁻² yr⁻¹ (see Chapter 3)

2.4 Ecoregional Rescaling

Raw cell scores vary naturally across regions due to differences in species richness and oceanographic conditions. To enable meaningful comparison, scores are rescaled to a [0–100] range within each BOEM Ecoregion using min-max normalization (Equation 2.3):

$$v'_c = \frac{v_c - v_{min}}{v_{max} - v_{min}} \times 100 \quad (2.3)$$

where v_{min} and v_{max} are the minimum and maximum cell scores for a given component within the Ecoregion.

Scores are relative within Ecoregions, not absolute. A score of 100 represents the most sensitive 0.05° cell for a given component (e.g., fish, marine mammals, or primary productivity) within a given BOEM Ecoregion—not globally. The final score is an equally weighted average of these ecoregionally rescaled components. This means that a score of 30 in Alaska reflects a different ecological context than a score of 30 in the Pacific.

Why rescale by Ecoregion? Without this step, areas with naturally higher numbers of species (like upwelling zones in the Pacific) would almost always get the highest scores, simply because they have more species. While more species may indicate greater ecological importance, it could also be argued that the fewer species in an ecosystem, the more important each is to its function and resilience. Rescaling by Ecoregion prevents scores from being overwhelmed by species-rich areas. However, comparing *across* Ecoregions would require developing a mathematical relationship between these competing arguments for sensitivity—one that is still a source of earnest academic debate. For this reason, **scores are directly comparable among Program Areas within the same Ecoregion, but should not be compared across Ecoregions.**

2.4.1 Geographic Scope

The current analysis covers 20 BOEM Program Areas from the 11th National Draft Proposed Program (2025) spanning Alaska, the Pacific, and the Gulf of America (Figure 2.1). The Atlantic OCS region is not included in this program cycle, so ecoregions without Program Areas (e.g., Northeast Continental Shelf, Southeast Continental Shelf, Washington/Oregon) are excluded from the analysis. Species whose distributions do not overlap with any Program Area will not appear in the interactive mapping applications.

The Gulf of America Program Areas—GOA Program Area A (GAA) and GOA Program Area B (GAB)—were derived from the larger Gulf of America Planning Areas and span parts of two Ecoregions: Western and Central Gulf of America (WCGOA) and Eastern Gulf of America (EGOA). Because cell scores are rescaled within each Ecoregion before aggregation, cells in GAA that fall within the WCGOA Ecoregion are rescaled relative to WCGOA min/max values, while cells within EGOA use EGOA min/max values. The aggregated Program Area score is then the area-weighted average across all constituent cells, regardless of which Ecoregion they belong to. This approach ensures that the rescaling reflects local ecological context even when a Program Area spans Ecoregion boundaries. See Chapter 7 for implementation details.

See subsequent chapters for detailed descriptions of: data sources (Chapter 3), taxonomic integration (Chapter 4), extinction risk scoring (Chapter 5), model merging (Chapter 6), and scoring methodology (Chapter 7).

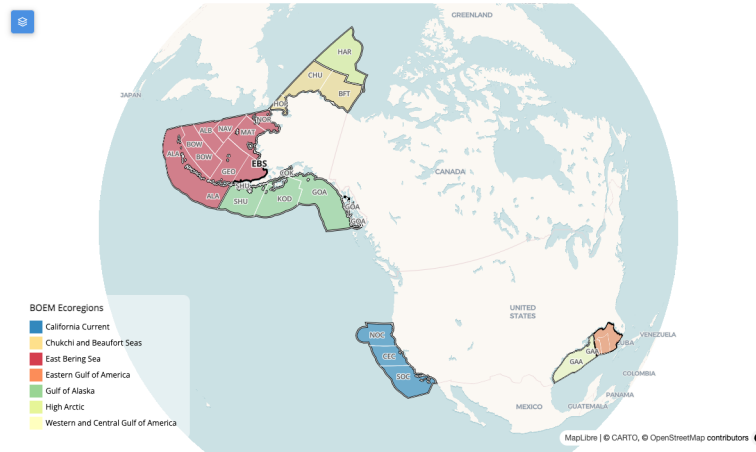


Figure 2.1: BOEM Ecoregions (colored) and Program Area outlines. **Ecoregion keys:** CAC = California Current; CBS = Chukchi and Beaufort Seas; EBS = East Bering Sea; EGOA = Eastern Gulf of America; GOA = Gulf of Alaska; HAR = High Arctic; WCGOA = Western and Central Gulf of America. **Program Area keys:** ALA = Aleutian Arc; ALB = Aleutian Basin; BFT = Beaufort Sea; BOW = Bowers Basin; CEC = Central California; CHU = Chukchi Sea; COK = Cook Inlet; GAA = GOA Program Area A; GAB = GOA Program Area B; GEO = St. George Basin; GOA = Gulf of Alaska; HAR = High Arctic; HOP = Hope Basin; KOD = Kodiak; MAT = St. Matthew-Hall; NAV = Navarin Basin; NOC = Northern California; NOR = Norton Basin; SHU = Shumagin; SOC = Southern California.

3 Data Sources

The MST integrates species distribution data from 8 source datasets plus 1 derived merged dataset, all stored in a DuckDB database at 0.05° resolution (~4 km cells). Each dataset provides spatial predictions of species occurrence, ranging from continuous suitability models (AquaMaps) to habitats associated with a core area, critical habitat or distinct population segment (NMFS, FWS, SWOT+DPS) to those based on simple range maps (BirdLife, IUCN, FWS).

3.1 Dataset Overview

Table 3.1 summarizes all 9 datasets in the MST database.

Table 3.1: Summary of the 9 datasets in the MST database. Cell values represent species presence probability or suitability as a percentage (1–100%). Datasets marked **Is Mask = Yes** can also serve as spatial masks during model merging.

Dataset Key	Display Name	Cell Value Encoding	Is Mask	Sort
am_0.05	AquaMaps SDM	Continuous suitability 0–100%	No	1
ca_nmfs	NMFS Core Area	Core: 100%	Yes	2
ch_nmfs	NMFS Critical Habitat	EN: 100%, TN: 50%	Yes	3
ch_fws	FWS Critical Habitat	EN: 100%, TN: 50%	Yes	4
rng_fws	FWS Range	EN: 100%, TN: 50%, LC: 1%	Yes	5
bl	BirdLife Range	CR: 50%, EN: 25%, VU: 5%, NT: 2%, LC: 1%	Yes	6
rng_iucn	IUCN Range	CR: 50%, EN: 25%, VU: 5%, NT: 2%, LC: 1%	Yes	7
rng_turtle_swot_dps	SWOT+DPS Turtle Range	EN: 100%, TN: 50%	Yes	8
ms_merge	Merged Model	MAX across source datasets	No	0

The **sort order** determines priority during model merging: datasets with continuous predictions (AquaMaps) are considered first, followed by progressively coarser range maps. The merged model (**ms_merge**) takes the MAX value across all available source datasets for each species in each cell.

3.2 Source Datasets

3.2.1 AquaMaps SDM (**am_0.05**)

[AquaMaps](#) provides standardized species distribution models (SDMs) for over 17,000 marine species based on environmental envelope models (Kaschner et al. 2019). Each species model defines habitat suitability as a function of depth, temperature, salinity, primary productivity, ice concentration, and distance from land.

- **Species count:** ~17,550 models
- **Native resolution:** 0.5° (c-squares), downscaled to 0.05° using bilinear interpolation
- **Cell values:** continuous suitability from 0–100%
- **Coverage:** global ocean

AquaMaps serves as the foundational dataset for most species, providing the highest spatial resolution and broadest taxonomic coverage.

3.2.2 NMFS Core Area (**ca_nmfs**)

Core areas delineated by the National Marine Fisheries Service [NMFS; National Marine Fisheries Service (2019)] identify regions of concentrated use for species under NMFS jurisdiction.

- **Species count:** limited to species with designated core areas
- **Cell values:** 100% within core area boundaries
- **Role:** contributes to merged model as a mask dataset

3.2.3 NMFS Critical Habitat (**ch_nmfs**)

Critical habitat designated under the Endangered Species Act (ESA) by NMFS for marine species (National Marine Fisheries Service 2025).

- **Species count:** 34 species
- **Cell values:** Endangered species = 100%, Threatened species = 50%
- **Role:** dual function — contributes cell values AND forms part of the spatial mask for model merging

3.2.4 FWS Critical Habitat (`ch_fws`)

Critical habitat designated under the ESA by the U.S. Fish and Wildlife Service [FWS; U.S. Fish and Wildlife Service (2025b)] for marine and coastal species.

- **Species count:** 29 species
- **Cell values:** Endangered species = 100%, Threatened species = 50%
- **Role:** dual function — contributes cell values AND forms part of the spatial mask

3.2.5 FWS Range (`rng_fws`)

Current range maps maintained by FWS for ESA-listed marine and coastal species (U.S. Fish and Wildlife Service 2025a).

- **Species count:** 106 species
- **Cell values:** Endangered = 100%, Threatened = 50%, Least Concern = 1%
- **Role:** mask dataset, providing spatial extent for species not covered by AquaMaps

3.2.6 BirdLife Range (`bl`)

Expert-reviewed range maps from [BirdLife International's Birds of the World](#) [BOTW; BirdLife International (2024)] dataset, representing the most authoritative global seabird distribution data.

- **Species count:** 573 seabird species
- **Cell values:** scaled by IUCN Red List category — CR: 50%, EN: 25%, VU: 5%, NT: 2%, LC: 1%
- **Role:** mask dataset, providing spatial constraints for seabird species
- **Note:** BirdLife range maps are expert-delineated polygons; cell values reflect conservation status rather than habitat suitability

3.2.7 IUCN Range (`rng_iucn`)

Range maps from the [IUCN Red List of Threatened Species](#) spatial data (IUCN 2025), covering a broad array of marine taxa.

- **Cell values:** scaled by IUCN Red List category — CR: 50%, EN: 25%, VU: 5%, NT: 2%, LC: 1%
- **Role:** mask dataset, providing spatial extent for species with IUCN range data

3.2.8 SWOT+DPS Turtle Range (`rng_turtle_swot_dps`)

Sea turtle distributions from the original AquaMaps and IUCN range maps were overly broad, extending well beyond where species actually occur. Furthermore, several sea turtle species have differential ESA protection across their range according to NOAA Fisheries Distinct Population Segments (DPS).

This dataset replaces the IUCN range map as the **global spatial mask** for all 6 sea turtle species. It combines two data sources:

- **SWOT Global Distributions** from the [State of the World's Sea Turtles](#) (Wallace et al. 2023), which define more realistic species range polygons.
- **NMFS Distinct Population Segments** (DPS) from the NOAA Fisheries ESA Species Range Geodatabase (NOAA Fisheries 2025), which identify sub-populations listed as Endangered under the ESA.

For species with DPS data (Loggerhead CC, Green CM, Olive Ridley LO): Endangered DPS areas are coded at 100%, and the remainder of the SWOT range is coded as Threatened at 50%. For purely Endangered species (Leatherback DC, Hawksbill EI, Kemp's Ridley LK): the entire SWOT range is coded at 100%.

- **Species count:** 6 sea turtle species
- **Cell values:** Endangered = 100%, Threatened = 50%
- **Role:** mask dataset — replaces `rng_iucn` as the global mask for turtles, producing more restrictive distributions dominated by ESA risk values See workflow: [ingest_turtles-swot-dps](#).

3.3 Merged Model (`ms_merge`)

The merged model is the derived output of the model merging pipeline (see Chapter 6). For each of the 16,153 valid species, cell values are computed as:

1. **MAX** across all source dataset values for that species in each cell
2. **Masked** to the spatial extent defined by IUCN, NMFS CH, and FWS CH ranges (when available)
3. **Floored** at minimum values for MMPA-protected species (20%) and MBTA-protected species (10%)

3.4 Standard Grid

All datasets are aligned to a standard $0.05^\circ \times 0.05^\circ$ latitude-longitude grid covering US waters within BOEM Program Areas. At mid-latitudes, each cell represents approximately 4×4 km (16 km^2). The grid is stored as a reference raster with unique `cell_id` values, enabling efficient joins between spatial data and tabular attributes in DuckDB.

3.5 Primary Productivity

Primary productivity is specified in the explicit mandate for BOEM’s management, per the Outer Continental Shelf Lands Act (OCSLA), Section 18(a)(2) of the OCSLA Amendments of 1978 specifying 8 factors the USDOJ must consider in the timing and location of OCS oil and gas activities, including “the relative environmental sensitivity and marine productivity of different areas of the OCS.”

We use satellite-derived net primary productivity (NPP) from the Vertically Generalized Production Model [VGPM; Behrenfeld and Falkowski (1997)] product from [Oregon State’s Ocean Productivity Lab](#), based on VIIRS satellite data for 2014–2023. Values are converted to metric tons C $\text{km}^{-2} \text{ yr}^{-1}$ and averaged across the time period.

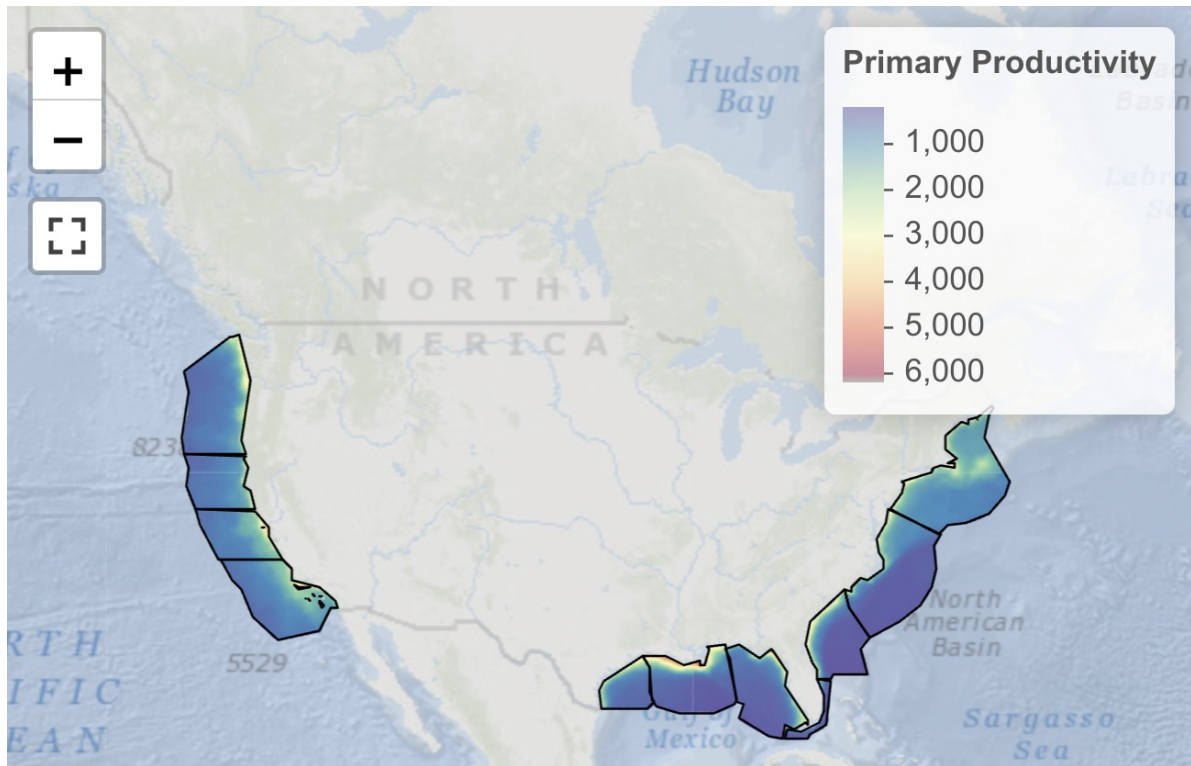


Figure 3.1: Primary productivity throughout the continental United States averaged across months of 2021, as measured by Oregon State's Vertically Generalized Production Model (VGPM). Expansion is anticipated across the entire US EEZ and for more recent years.

4 Taxonomy

Integrating species distribution data from 7 source datasets requires resolving taxonomic identities across multiple naming systems. The MST uses a multi-authority matching pipeline to ensure each species is uniquely identified, enabling accurate merging of models from different data providers.

4.1 Taxonomic Authorities

The following authorities are loaded into a reference database (`spp.duckdb`) for taxonomic reconciliation:

Table 4.1: Taxonomic authorities used in the MST for species name resolution.

Authority	Description	Key Identifier
WoRMS	World Register of Marine Species — authoritative list for marine taxa	<code>worms_id</code> (AphiaID)
GBIF	Global Biodiversity Information Facility Backbone Taxonomy (~6.3M taxa)	<code>gbif_id</code>
ITIS	Integrated Taxonomic Information System — US federal standard	<code>itis_tsn</code>
IUCN Red List	International Union for Conservation of Nature — conservation assessments	<code>iucn_id</code>
BirdLife BOTW	Birds of the World — authoritative for seabird taxonomy	<code>botw_id</code>

4.2 ID Resolution Cascade

Species identifiers are resolved through a cascading lookup process:

1. **ITIS TSN match:** if the source dataset provides an ITIS Taxonomic Serial Number, use the ITIS-to-WoRMS crosswalk for direct matching
2. **WoRMS crosswalk:** look up accepted WoRMS AphiaID via the GBIF backbone, which integrates WoRMS as the marine taxonomy source
3. **Scientific name match:** for records without matching identifiers, attempt exact scientific name matching against WoRMS accepted names
4. **API lookup:** for remaining unresolved names, query the WoRMS REST API (`wm_records_name()`) for fuzzy matching

At each stage, deprecated names are resolved to their accepted synonyms, and the `taxonomicStatus` field is used to determine the preferred name.

4.3 Species Categories

Valid species are classified into 7 categories based on their taxonomic position:

Table 4.2: Species categories used in the MST for scoring and visualization.

Category	Description	Examples
bird	Seabirds and shorebirds	albatross, petrel, tern, pelican
coral	Reef-building and deep-sea corals	stony coral, soft coral, black coral
fish	Bony and cartilaginous fishes	grouper, shark, ray, tuna
invertebrate	Non-coral marine invertebrates	crab, lobster, sea urchin, squid
mammal	Marine mammals	whale, dolphin, seal, manatee
reptile	Sea turtles	loggerhead, green, leatherback
other	Uncategorized marine organisms	worms, tunicates, and bryozoans

4.4 Data Quality

4.4.1 Duplicate Resolution

Multiple source datasets may provide models for the same species. The taxonomic matching step identifies duplicates by resolving all source names to canonical WoRMS or BirdLife identifiers. When a species appears in multiple datasets, the model merging pipeline (see Chapter 6) takes the MAX value across all sources rather than treating them as separate species.

4.4.2 Synonym Handling

Taxonomic names change over time as species are reclassified. The pipeline handles this by:

- resolving all names through `acceptedNameUsageID` in each authority
- tracking the original source name alongside the accepted name
- flagging deprecated WoRMS IDs and updating them to current accepted IDs

4.4.3 Valid Species Filter (`is_ok`)

The source datasets together contain roughly **17,561 taxa** — dominated by the ~17,000 marine species modeled globally by AquaMaps, plus the regulatory and range-map datasets. Not all are appropriate for the sensitivity analysis. A taxon is flagged **valid** (`is_ok = TRUE`) only when it meets *every* one of the following criteria:

- **Taxonomically accepted & marine:** a resolved WoRMS/BirdLife identity with an accepted (or alternative-representation) status, classified as marine (WoRMS `isMarine`, or BirdLife seabird), and not extinct (WoRMS `isExtinct` or IUCN Red List `EX`); non-turtle reptiles (e.g., sea snakes) are excluded.
- **Has a mapped distribution:** the merged model resolves to at least one grid cell with a value — a few stale, empty model records are excluded so the valid count matches the species actually mapped.
- **Expert range intersects the US EEZ:** where an IUCN expert range map exists, it must overlap US waters. Species whose global range falls entirely outside the US EEZ are dropped even if AquaMaps predicts suitable habitat there, since those are edge-of-range model artifacts not supported by expert assessment.

Program-Area overlap is *not* a validity criterion. A species is valid if it has a modeled distribution anywhere in the study area; whether it falls within one of the BOEM Program Areas is determined later by a spatial (zone) query, not by the `is_ok` flag. This was introduced in **v7** — earlier versions silently required Program-Area overlap, which conflated “valid” with “occurs in a Program Area” and hid biodiversity outside the current program cycle (see below).

This filter currently yields **16,153 valid species** across the full study area (the US EEZ), of which **9,230** have modeled distributions within the **20 BOEM Program Areas** of the current program cycle. These figures are pulled live from the </stats.json> endpoint for the current database version (v7).

i How the valid-species count has changed across versions

The definition of “valid species” has tightened over time, so the headline count is **not** directly comparable across versions:

- **v5 (~9,819)** — valid after taxonomic filtering, before excluding IUCN ranges outside the US EEZ.
- **v6 (9,424)** — after excluding ~371 species whose IUCN ranges fall entirely outside the US EEZ. This count *also* silently required overlap with a BOEM Program Area, conflating “valid” with “occurs in a Program Area.”
- **v7 (16,153)** — the Program-Area requirement was removed, so the count reflects the full study-area biodiversity (the ~17,000-species universe minus the taxonomic and range exclusions). The **9,230** within the Program Areas is now reported separately as a spatial subset rather than baked into the flag.

4.4.4 From source taxa to valid species

The table below traces the cumulative effect of each **is_ok** gate — applied in the order used by the merge pipeline — showing the number of species **removed** at each step and the number **remaining**, from the full pool of source-dataset taxa down to the valid species, then the spatial subset that falls within the BOEM Program Areas. The largest reductions are species with no merged distribution (those whose IUCN expert range falls entirely outside the US EEZ produce zero cells) and non-marine taxa. Counts are pulled live for the current database version (v7).

Table 4.3: Cumulative **is_ok** filter funnel: species removed and remaining at each validity gate, then the within-Program-Area spatial subset. Pulled live from the API for the current database version.

Filter step	Removed	Species remaining
Source taxa (all datasets)	—	17,561
Resolved taxon ID	—	17,561
Has a merged model (distribution)	−1,173	16,388
Not extinct	−21	16,367
Marine	−154	16,213
Accepted taxonomy (excl. non-turtle reptiles)	−55	16,158

Table 4.3: Cumulative `is_ok` filter funnel: species removed and remaining at each validity gate, then the within-Program-Area spatial subset. Pulled live from the API for the current database version.

Filter step	Removed	Species remaining
Mapped to 1 cell — valid (<code>is_ok</code>)	—5	16,153
Within BOEM Program Areas (spatial subset)	—6,923	9,230

4.5 Key Function

The taxonomic matching is implemented in `msens::match_taxa()`, which orchestrates the ID resolution cascade and returns a unified taxon table with cross-referenced identifiers from all authorities.

5 Extinction Risk

Extinction risk scoring translates a species' conservation status into a numeric weight that reflects its vulnerability to population-level impacts. The MST integrates protections from three US federal regulatory frameworks and the international IUCN Red List to assign each species a score from 1 to 100.

Interactive Walkthrough

The [Scoring Walkthrough](#) includes a visual comparison of ESA vs. IUCN scoring frameworks and shows how extinction risk weights are applied to species suitability values.

5.1 Regulatory Framework

5.1.1 Endangered Species Act (ESA)

The [Endangered Species Act](#) (1973) provides protections for species listed as Endangered (EN) or Threatened (TN) by either NMFS (for marine species) or FWS (for terrestrial/coastal species).

- **Endangered (EN)**: species in danger of extinction throughout all or a significant portion of its range
- **Threatened (TN)**: species likely to become endangered within the foreseeable future

ESA listings are the primary driver of extinction risk scores for US-listed species, with Endangered receiving the maximum score (100) and Threatened receiving a score of 50.

5.1.2 Marine Mammal Protection Act (MMPA)

The [Marine Mammal Protection Act](#) (1972) protects **all marine mammals** in US waters regardless of their ESA status. This includes whales, dolphins, porpoises, seals, sea lions, walruses, sea otters, polar bears, and manatees.

Because MMPA provides blanket protection, all marine mammals receive a **minimum spatial floor score of 20** in cells where they are present — even species classified as Least Concern

under IUCN. This ensures that the legal protection afforded to marine mammals is reflected in the sensitivity analysis.

The MMPA floor is intentionally higher than the MBTA floor because the MMPA imposes a stronger statutory standard: it affirmatively requires that all marine mammal populations be maintained at Optimum Sustainable Population levels and mandates active population recovery, whereas the MBTA is primarily a prohibitive take statute. Marine mammals also tend to have lower reproductive rates and longer recovery times than most seabirds, making them intrinsically more sensitive to disturbance.

5.1.3 Migratory Bird Treaty Act (MBTA)

The [Migratory Bird Treaty Act](#) (1918) protects migratory birds, which includes most seabird species. Similar to MMPA, all migratory birds receive a **minimum spatial floor score of 10** in cells where they are present.

5.1.4 IUCN Red List

The [IUCN Red List of Threatened Species](#) provides internationally recognized conservation status assessments used for species not listed under US federal laws:

Table 5.1: IUCN Red List categories used in extinction risk scoring.

Category	Abbreviation	Description
Critically Endangered	CR	extremely high risk of extinction in the wild
Endangered	EN	very high risk of extinction in the wild
Vulnerable	VU	high risk of extinction in the wild
Near Threatened	NT	close to qualifying for a threatened category
Least Concern	LC	widespread and abundant
Data Deficient	DD	inadequate information to assess

5.2 Scoring Methodology

Extinction risk scores are computed by `msens::compute_er_score()`, which implements a two-track scoring system depending on whether the species is listed under US federal law.

5.2.1 US-listed Species

For species with NMFS or FWS ESA listings, the score is the **maximum** across three components:

$$\text{er_score} = \max(\text{ESA base}, \text{MMPA floor}, \text{MBTA floor}) \quad (5.1)$$

Table 5.2: US-listed species scoring components. The final score is the MAX across applicable components.

Component	Score
ESA Endangered (EN)	100
ESA Threatened (TN)	50
ESA not listed (LC)	1
MMPA protected	20
MBTA protected	10

Examples:

- ESA Endangered whale (MMPA): $\max(100, 20) = \mathbf{100}$
- ESA Threatened seabird (MMPA + MBTA): $\max(50, 20, 10) = \mathbf{50}$
- Non-ESA marine mammal (MMPA only): $\max(1, 20) = \mathbf{20}$
- Non-ESA migratory bird (MBTA only): $\max(1, 10) = \mathbf{10}$

5.2.2 Species not federally listed (IUCN)

Species not federally listed or mentioned by NMFS ([Species Directory | NOAA Fisheries](#)) or FWS ([Species | U.S. Fish & Wildlife Service](#)) fall back to an extinction risk score based on their IUCN Red List category:

Table 5.3: IUCN-based extinction risk scores for US species not listed or mentioned by NMFS or FWS

IUCN Category	Score
Critically Endangered (CR)	50
Endangered (EN)	25
Vulnerable (VU)	5
Near Threatened (NT)	2
Least Concern / Data Deficient / other	1

Species with no matching extinction risk code (i.e., unmatched taxa) receive a default score of 1.

5.2.3 Scale Note

Scores range from 1 to 100, representing a percentage scale (1–100%). Integer storage is used for computational efficiency; conceptually the scale represents a 0–1 probability-like weight where 100 = 1.0 (maximum extinction risk) and 1 = 0.01 (minimum baseline).

5.3 Implementation

The scoring function is implemented in the `msens` R package:

```
msens::compute_er_score(  
  extrisk_code,      # e.g., "NMFS:EN", "FWS:TN", "IUCN:CR"  
  is_mmpa = FALSE,   # MMPA protection flag  
  is_mmta = FALSE     # MBTA protection flag  
)
```

The `extrisk_code` field follows the format `AUTHORITY:STATUS` where:

- **Authority:** NMFS, FWS, or IUCN
- **Status:** ESA codes (EN, TN, LC) or IUCN codes (CR, EN, VU, NT, LC, DD)

6 Model Merging

Each species may have distribution data in multiple source datasets (see Chapter 3). The model merging pipeline combines these into a single merged model per species, applying spatial masking and minimum floor values to reflect regulatory protections.

Interactive Walkthrough

The [Scoring Walkthrough](#) visualizes the merge process interactively — see how MAX merge, spatial masking, and regulatory floors transform input rasters step by step.

6.1 Pipeline Overview

The merge pipeline processes each valid species through the following steps:

6.1.1 Step 1: Gather Source Models

For each species (identified by `taxa_id`), the pipeline queries all available models across the 8 source datasets. A species may have:

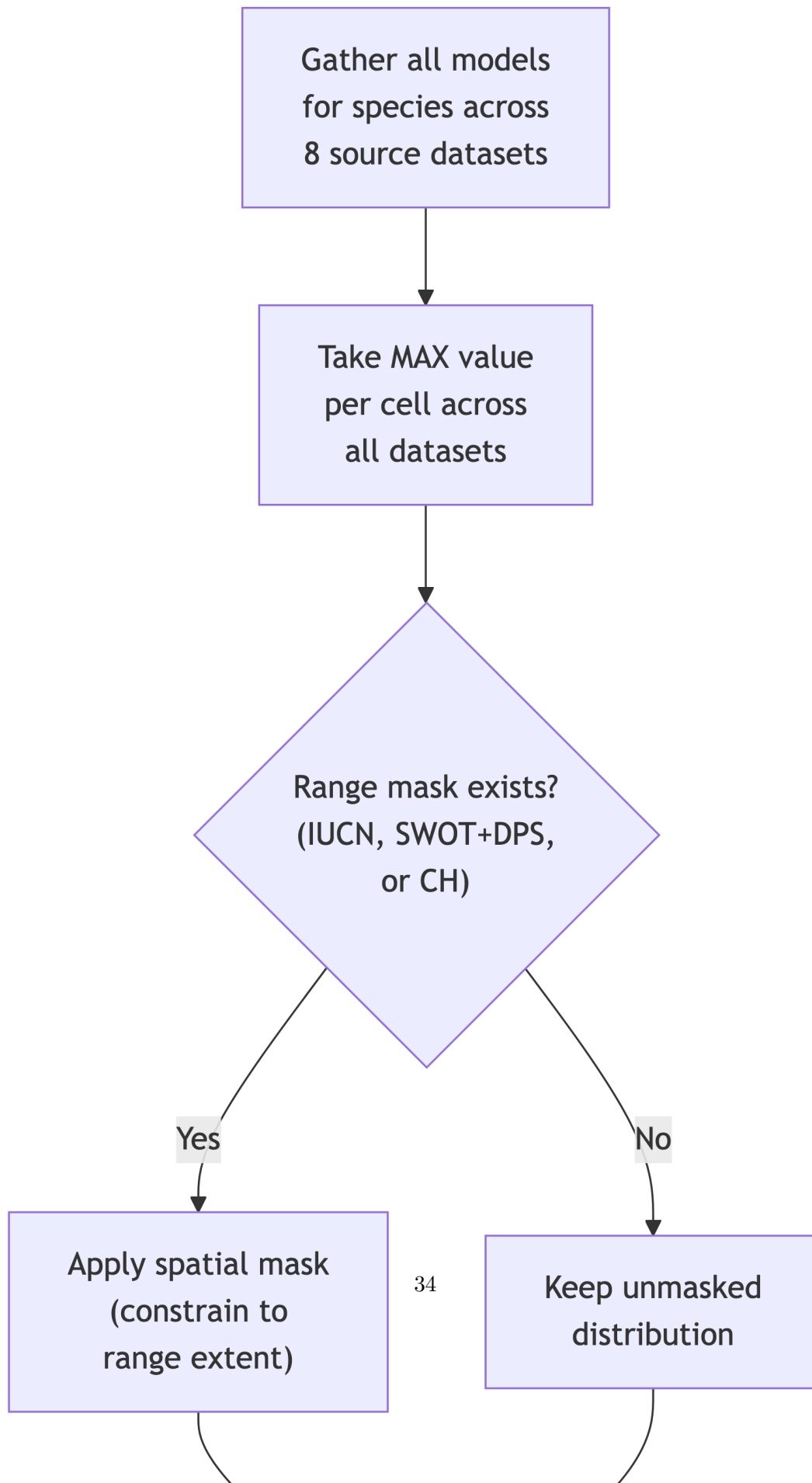
- an AquaMaps SDM (continuous suitability)
- one or more regulatory range/habitat designations (NMFS, FWS)
- a BirdLife or IUCN range map
- a SWOT+DPS range map (sea turtles only)

6.1.2 Step 2: MAX Across Datasets

For each grid cell, the merged value is the **maximum** suitability across all source datasets:

$$v_{merged,c} = \max(v_{am,c}, v_{ca,c}, v_{ch,c}, v_{rng,c}, v_{bl,c}, v_{iucn,c}) \quad (6.1)$$

This ensures that the most informative (highest confidence) prediction is used. For example, if AquaMaps predicts 30% suitability but NMFS has designated the cell as critical habitat for an endangered species (100%), the merged value is 100%.



6.1.3 Step 2b: Turtle Multiplicative Merge

For **sea turtles** (6 species), the merge uses a multiplicative approach instead of simple MAX. The SWOT+DPS dataset provides a spatially varying extinction risk (ER) surface where cell values encode the ESA listing status of the Distinct Population Segment (DPS) that overlaps each cell (Endangered = 100, Threatened = 50). This ER surface is multiplied with the suitability surface (from AquaMaps or other SDMs) during merging:

$$v_{merged,c} = \max\left(1, \text{round}\left(\frac{v_{er,c} \times v_{suit,c}}{100}\right)\right) \quad (6.2)$$

where:

- $v_{er,c}$ = extinction risk value from SWOT+DPS raster (EN: 100, TN: 50)
- $v_{suit,c}$ = suitability value from AquaMaps SDM (1–100)

Critical habitat cells (NMFS or FWS designated) override via MAX — if the critical habitat value exceeds the multiplicative product, the higher value is used.

Because extinction risk is already embedded in the merged values, these species receive `er_score` = 100 at scoring time as a pass-through multiplier (flagged via `is_er_spatial`). This contrasts with the standard approach where all species receive a single scalar extinction risk weight applied during scoring (see Chapter 7).

i Two modeling approaches

The MST uses two distinct merge strategies:

- **Standard (most species)**: MAX across input rasters, then multiply by a **scalar** extinction risk score at scoring time
- **Turtles only**: Multiply suitability raster by an **extinction risk raster** during merging, embedding spatially varying ER directly in the merged model

See the [Scoring Walkthrough](#) for an animated side-by-side comparison.

6.1.4 Step 3: Spatial Masking

When a range map or designated habitat exists for a species, the merged model is **constrained to the spatial extent** of these mask datasets. The mask is formed as the union of all available `is_mask` = TRUE datasets for that species.

For most taxa, the IUCN range map serves as the global mask. For **sea turtles**, the SWOT+DPS dataset (`rng_turtle_swot_dps`; see Section 3.2.8) replaces the IUCN range as the global mask, providing more realistic distributions based on expert-reviewed SWOT

polygons with ESA-based DPS status encoding. NMFS Critical Habitat and FWS Critical Habitat also contribute to the mask when designated.

This prevents the AquaMaps SDM (which often has broad environmental envelope predictions) from extending species presence far beyond their known range. Cells outside the mask are set to zero (species absent). This ensures that suitable habitat is not included outside known ranges, so that model predictions align with expert knowledge of where species actually occur.

! v6 fix: species with ranges outside the US EEZ

Species whose IUCN expert range polygon produces zero cells when rasterized to the US EEZ grid now receive a 0-cell `rng_iucn` entry in the database (rather than being silently skipped during ingestion). The merge clips AquaMaps predictions against this empty mask, producing zero merged cells, and the species is flagged `is_ok=FALSE`. This excluded 371 species (343 fish, 20 invertebrates, 6 corals, 2 mammals) that had spurious AquaMaps edge-of-range predictions in US waters despite IUCN expert assessment placing them entirely outside the US EEZ. The quantitative impact on Program Area scores is negligible because the removed species had near-zero suitability values.

6.1.5 Step 4: MMPA Spatial Floor

For species protected under the Marine Mammal Protection Act (all marine mammals), a spatial minimum floor is applied:

$$v_c = \max(v_{merged,c}, 20) \quad (6.3)$$

This ensures that every cell where a marine mammal is present has a minimum value of 20%, reflecting the legal protection that MMPA affords regardless of ESA status. This includes both NMFS-managed species (whales, dolphins, seals, sea lions) and FWS-managed species (walrus, polar bear, sea otter, manatee, dugong).

! v5 fix: FWS-managed marine mammals

As of v5, the MMPA floor is correctly applied to **all** marine mammals regardless of managing agency. Earlier versions incorrectly excluded FWS-managed species (walrus, polar bear, sea otter, manatee, dugong) from this floor because the `is_mmpa` flag was derived only from the NMFS Species Directory.

6.1.6 Step 5: MBTA Spatial Floor

Similarly, for species protected under the Migratory Bird Treaty Act (most seabirds), a spatial minimum floor of 10% is applied:

$$v_c = \max(v_{merged,c}, 10) \quad (6.4)$$

6.1.7 Step 6: Persist Results

The final merged model is stored in the DuckDB database as:

- **model table:** one row per species with metadata (dataset key `ms_merge`, taxa reference, model sequence)
- **model_cell table:** one row per species $\times$ cell combination with the merged suitability value

6.2 Valid Species Filter

After merging, species are flagged as valid (`is_ok = TRUE`) based on the criteria described in Chapter 4. The filtering rules differ slightly between birds and other taxa:

Birds (from BirdLife BOTW):

- has a `botw_id` (BirdLife identifier)
- IUCN Red List code is not “EX” (Extinct)
- if also in WoRMS: must be marine and not extinct

Other marine taxa (from WoRMS):

- has a `taxa_id` and a merged model (`mdl_seq`)
- IUCN Red List code is not “EX”
- WoRMS `isMarine = TRUE` and `isExtinct != TRUE`
- species category is not “reptile” (except sea turtles, which are reclassified as category “reptile”)

As of v7, an overlap with a BOEM Program Area is **no longer** required for validity (it previously narrowed the set to 9,424); Program-Area scope is now a spatial/zone query so the full study-area biodiversity is retained.

This filtering yields **16,153 valid species** across the study area, of which **9,230** contribute to scores within the 20 BOEM Program Areas of the current program cycle.

7 Scoring

The scoring methodology translates merged species distribution models and extinction risk weights into cell-level sensitivity metrics, which are then rescaled and aggregated to management zones for decision support.

Interactive Walkthrough

See the [Scoring Walkthrough](#) for an animated, step-by-step visualization of this pipeline — from input rasters through MAX merge, ecoregional rescaling, and component averaging.

7.1 Cell-Level Scoring

For each grid cell, sensitivity is computed as the extinction risk weighted sum of species presence within each species category (Equation 7.1):

$$\text{score}_{c,g} = \sum_{s=1}^{S_g} \frac{\text{er_score}_s \times v_{s,c}}{100} \quad (7.1)$$

where:

- $\text{score}_{c,g}$ = sensitivity score for cell c and species category g
- er_score_s = extinction risk score for species s (1–100; see Chapter 5)
- $v_{s,c}$ = merged model value for species s in cell c (0–100%; see Chapter 6)
- S_g = number of valid species in category g

Division by 100 converts the product from a 0–10,000 range back to a 0–100 scale.

In plain terms, for each cell in the ocean the MST adds up the sensitivity contributions of all species found there. If a cell has many species that are both likely to be present and at high risk of extinction, it gets a higher sensitivity score. This helps identify places where rare or threatened species are concentrated.

7.1.1 Two Scoring Approaches

The formula above applies to **most species** (fish, birds, mammals, corals, invertebrates, other), where **er_score** is a **scalar** value per species (1–100) and **v** is the MAX-merged suitability value.

For **sea turtles**, extinction risk is already embedded in the merged model values via the multiplicative merge (see Chapter 6, Step 2b). In this case the scoring formula simplifies to:

$$\text{score}_{c,g} = \sum_{s=1}^{S_g} v_{merged,s,c} \quad (7.2)$$

because **er_score** = 100 is used as a pass-through multiplier (dividing by 100 cancels it out). The spatially varying extinction risk from the SWOT+DPS raster is already encoded in v_{merged} , so cells within an Endangered DPS contribute higher values than cells within a Threatened DPS, even for the same species and suitability level.

7.2 Metric Keys

Cell-level scores are stored as named metrics in the DuckDB database:

Table 7.1: Cell-level metric keys stored in the MST database.

Metric Key	Description
extrisk_bird	extinction risk score for birds
extrisk_coral	extinction risk score for corals
extrisk_fish	extinction risk score for fishes
extrisk_invertebrate	extinction risk score for invertebrates
extrisk_mammal	extinction risk score for marine mammals
extrisk_reptile	extinction risk score for reptiles/turtles
extrisk_other	extinction risk score for other marine organisms
primprod	primary productivity (VGPM, metric tons C km ⁻² yr ⁻¹)

7.3 Ecoregional Rescaling

Raw cell scores vary naturally across ecoregions due to differences in species richness, oceanographic conditions, and biogeographic patterns. To enable meaningful comparison, each metric is rescaled to a [0–100] range within each BOEM Ecoregion:

$$\text{score}'_c = \frac{\text{score}_c - \text{score}_{\min}}{\text{score}_{\max} - \text{score}_{\min}} \times 100 \quad (7.3)$$

where $\text{score}_{\min}$ and $\text{score}_{\max}$ are the minimum and maximum cell values for that metric within the Ecoregion.

Scores are relative within Ecoregions, not absolute. A score of 100 represents the most sensitive 0.05° cell for a given metric within a given BOEM Ecoregion—not globally. The final score is an equally weighted average of these ecoregionally rescaled metrics. This means that a score of 30 in Alaska reflects a different ecological context than a score of 30 in the Pacific.

Why rescale by Ecoregion? Without this step, areas with naturally higher numbers of species (like upwelling zones in the Pacific) would almost always receive the highest scores, simply because they have more species. While more species may indicate greater ecological importance, it could also be argued that the fewer species in an ecosystem, the more important each is to its function and resilience. Rescaling by Ecoregion prevents scores from being overwhelmed by species-rich areas. However, comparing *across* Ecoregions would require developing a mathematical relationship between these competing arguments for sensitivity—one that is still a source of earnest academic debate. For this reason, **scores are directly comparable among Program Areas within the same Ecoregion, but should not be compared across Ecoregions.**

7.3.1 Geographic Scope

The current analysis covers 20 BOEM Program Areas from the 11th National Draft Proposed Program (2025) spanning Alaska, the Pacific, and the Gulf of America (Figure 7.1). The Atlantic OCS region is not included in this program cycle, so ecoregions without Program Areas (e.g., Northeast Continental Shelf, Southeast Continental Shelf, Washington/Oregon) are excluded from the analysis. Species whose distributions do not overlap with any Program Area will not appear in the interactive mapping applications.

7.3.2 Cross-Ecoregion Program Areas

The Gulf of America Program Areas—GOA Program Area A (GAA) and GOA Program Area B (GAB)—span parts of two Ecoregions: Western and Central Gulf of America (WCGOA) and Eastern Gulf of America (EGOA). Because cell scores are rescaled within each Ecoregion before aggregation, cells in GAA that fall within the WCGOA Ecoregion are rescaled relative to WCGOA min/max values, while cells within EGOA use EGOA min/max values. The aggregated Program Area score is then the area-weighted average across all constituent cells, regardless of which Ecoregion they belong to. This approach ensures that the rescaling reflects local ecological context even when a Program Area spans Ecoregion boundaries. See `calc_scores.qmd` for implementation details.

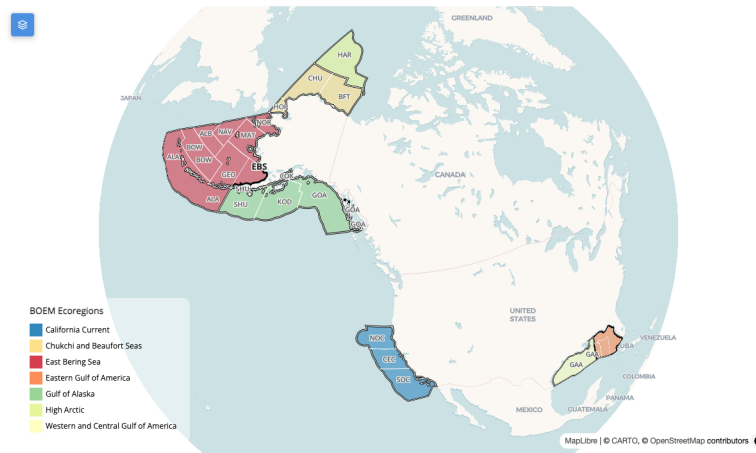


Figure 7.1: BOEM Ecoregions (colored) and Program Area outlines. **Ecoregion keys:** CAC = California Current; CBS = Chukchi and Beaufort Seas; EBS = East Bering Sea; EGOA = Eastern Gulf of America; GOA = Gulf of Alaska; HAR = High Arctic; WCGOA = Western and Central Gulf of America. **Program Area keys:** ALA = Aleutian Arc; ALB = Aleutian Basin; BFT = Beaufort Sea; BOW = Bowers Basin; CEC = Central California; CHU = Chukchi Sea; COK = Cook Inlet; GAA = GOA Program Area A; GAB = GOA Program Area B; GEO = St. George Basin; GOA = Gulf of Alaska; HAR = High Arctic; HOP = Hope Basin; KOD = Kodiak; MAT = St. Matthew-Hall; NAV = Navarin Basin; NOC = Northern California; NOR = Norton Basin; SHU = Shumagin; SOC = Southern California.

7.4 Primary Productivity Scoring

Net primary productivity (NPP) from the VGPM satellite model is included as a separate metric (**primprod**) that does not undergo extinction risk weighting. Instead, the raw NPP values (metric tons C km⁻² yr⁻¹) are directly rescaled within each ecoregion using the same [0–100] normalization.

7.5 Zone Aggregation

Rescaled cell scores are aggregated to three spatial levels using area-weighted averaging:

1. **Program Areas:** the 20 BOEM planning regions used for offshore energy management
2. **Subregions:** intermediate spatial groupings (e.g., Atlantic, Gulf of America, Pacific, Alaska)
3. **Ecoregions:** BOEM-defined ecological regions used as the rescaling baseline

For each zone, the aggregated score is:

$$\text{score}_{\text{zone}} = \frac{\sum_{c \in \text{zone}} \text{score}'_c \times A_c}{\sum_{c \in \text{zone}} A_c} \quad (7.4)$$

where A_c is the area of cell c (accounting for latitude-dependent cell size in the 0.05° grid).

7.6 Visualization

7.6.1 Flower Plot

The flower plot provides an intuitive summary of sensitivity scores by species category, inspired by the Ocean Health Index (Halpern et al. 2012) visualization approach:

- **Petal length** represents the rescaled sensitivity score (0–100) for each species category — longer petals indicate higher sensitivity
- **Petal width** represents the weight of each category contributing to the overall score — wider petals have more influence on the center score
- **Center value** shows the weighted mean across all categories

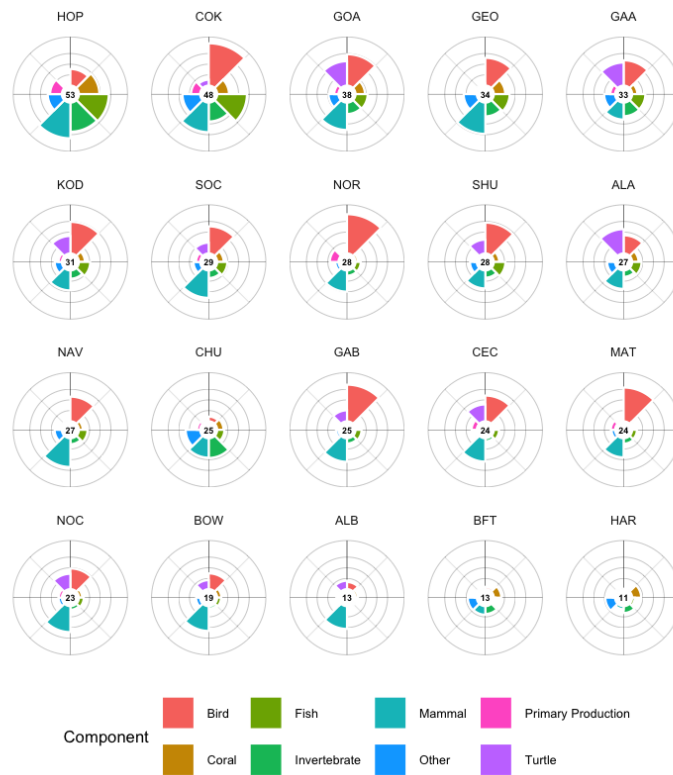


Figure 7.2: Flower plots for all 20 BOEM Program Areas showing ecoregionally rescaled sensitivity scores by species category. Petal length reflects the category score; petal width reflects its weight. Center values are the weighted mean across categories.

Figure 7.2 shows flower plots for all 20 BOEM Program Areas. Each flower summarizes the sensitivity profile of one program area with petals for Bird, Coral, Fish, Invertebrate, Mammal, Other, Turtle, and Primary Production. Alaska program areas (e.g., HOP, COK, GOA) show dominant Fish and Mammal petals, while Gulf areas (GAB, CEC, MAT) show stronger Invertebrate and Coral contributions.

The flower plot lets decision-makers quickly spot which ecological elements drive sensitivity in an area, aiding better planning and impact assessment.

The flower plot is implemented as an interactive **ggiraph** visualization using polar coordinates in R (`ggplot2 + coord_polar()`), with tooltips showing detailed score breakdowns on hover. The `plot_flower()` function in `calc_scores.qmd` generates static versions for reports, while the `mapgl` web application implements an interactive JavaScript version for responsive rendering (see [?@fig-mapgl-flower](#)).

7.6.2 Treemap

The treemap provides a detailed, hierarchical view of species contributions to sensitivity scores (Figure 7.3):

- Audience: scientists and analysts
- Shows the contribution of each species to the overall score within a pixel or management zone
- Elements are hierarchical: first by extinction risk category, then by taxonomic classification
- Interactive zooming allows drilling down to individual species contributions

Zooming into a specific category reveals individual species contributions:

7.6.3 Interactive Maps

The primary visualization tools are two Shiny web applications:

- **Scores**: general sensitivity mapping app showing composite scores by program area with flower plots and treemaps (see [Appendix: mapgl](#))
- **Species**: species distribution viewer showing individual and merged models with regulatory status information (see [Appendix: mapsp](#))

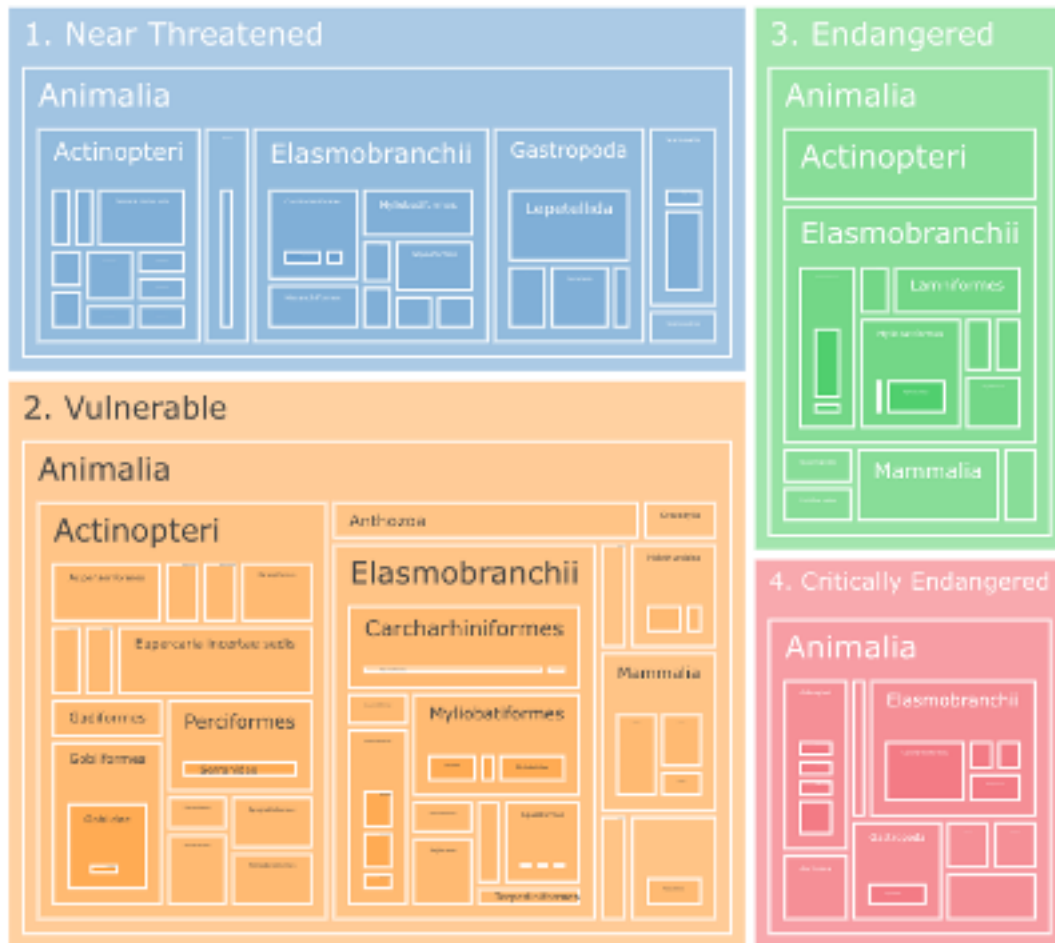


Figure 7.3: Treemap visualization showing species contributions to extinction risk scores, hierarchically organized by risk category and taxonomy.



Figure 7.4: Zoomed treemap showing individual species within a specific risk category and taxonomic group.

8 Scoring Walkthrough

Interactive step-by-step visualization of the sensitivity scoring pipeline

This interactive walkthrough demonstrates how the Marine Sensitivity Toolkit calculates sensitivity scores, from input rasters through model merging, ecoregional rescaling, and final aggregation. Use the arrow keys or navigation buttons to step through the pipeline. Click any step in the top bar to jump directly to it.

For the underlying methodology, see [Scoring](#) and [Model Merging](#). For source code, see the [merge_models](#) and [calc_scores](#) workflows.

9 Stressors

The stressor component of the vulnerability framework (see Chapter 2) addresses the human activities that may impact marine ecosystems. This section is a placeholder for future phases of the MST, which will incorporate stressor-receptor interaction matrices to complete the full vulnerability assessment.

9.1 Offshore Wind Energy

Evaluation of stressors from the offshore wind industry needs to be evaluated based on human activities given the phase of development, whether pre-construction, construction, operation or decommissioning (Figure 9.1).

9.2 Oil & Gas

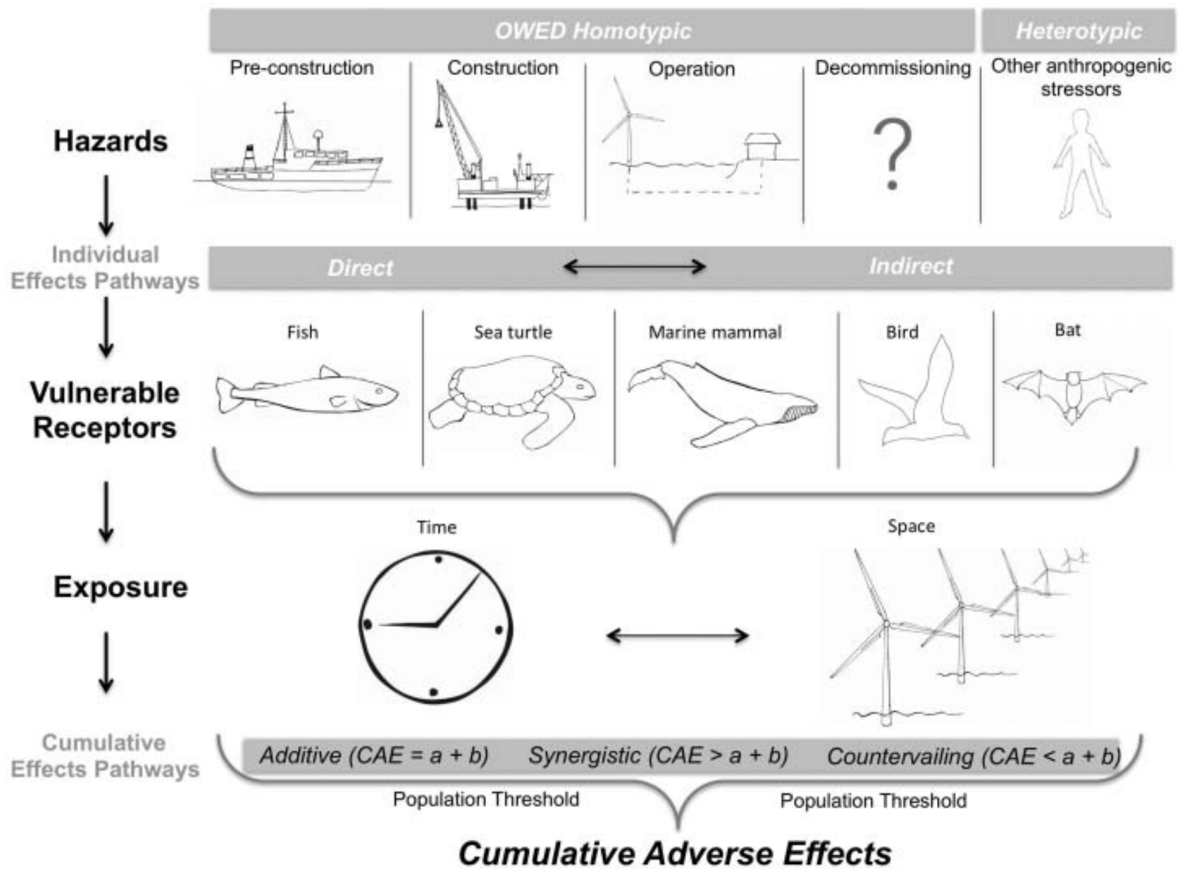


Figure 9.1: Cumulative adverse effects of offshore wind energy development on wildlife (Goodale and Milman 2016).

Part II

Software

10 Software

10.1 Software Overview

The **Marine Sensitivity Toolkit** (MST) is a stack of software components for reproducibly, interactively and hierarchically generating environmental vulnerability maps and scores. Tabular data is collated across studies evaluating sensitivity of species to oil & gas and offshore wind energy development. The best available species and benthic habitat distributions are being mosaicked across the US EEZ. Scores are summarized according to subgroups within Species, Benthic Habitats and Primary Productivity, averaged into an overall score, and visualized as a flower plot, applicable to various levels of BOEM relevancy: regional, ecoregions, protraction diagrams, blocks and aliquots.

The software components for achieving this interactively are:

1. **server** — a Docker Compose configuration orchestrating the reverse proxy, Shiny server, APIs, custom TiTiler tile factory + Varnish cache, and file server;
2. **database** — a columnar, embedded analytic database (**DuckDB**) as the single source of truth for species distributions, per-cell scores and derived per-zone metrics;
3. **workflows** — Quarto notebooks to ingest source datasets (AquaMaps, IUCN, NMFS/FWS, BirdLife, WoRMS, BOEM zones), merge taxonomic authorities, and compute scores that are written back to DuckDB;
4. **APIs** — a custom R **plumber** API for data requests and a custom **TiTiler** factory that renders on-the-fly raster PNG tiles from arbitrary DuckDB queries, cached by Varnish;
5. **libraries** — re-usable documented functions as the **msens** R package, including the tile-URL helpers (`cell_tile_url()`, `cell_stats()`, `add_cell_tiles()`) that the Shiny apps use to talk to the factory;
6. **applications** — interactive Shiny apps (`scores`, `species`, ...) that combine the raster tile factory with static **PMTiles** vector layers and live-pushed choropleth paint rules;
7. **documentation** — this Quarto book with figures, tables, glossary and references in interactive HTML or static DOCX / PDF output.

This toolbox is intended to primarily serve BOEM needs internally, but by being open-source and fully reproducible the hope is to enlist buy-in and even contributions from external partners, whether from other government agencies, academia, NGOs or industry.

We ascribe to the philosophy of sharing all code for the sake of reproducibility, transparency and efficiency (Maitner et al. 2024; Lowndes et al. 2017); i.e. the FAIR principles of Findability, Accessibility, Interoperability, and Reusability (Wilkinson et al. 2016).

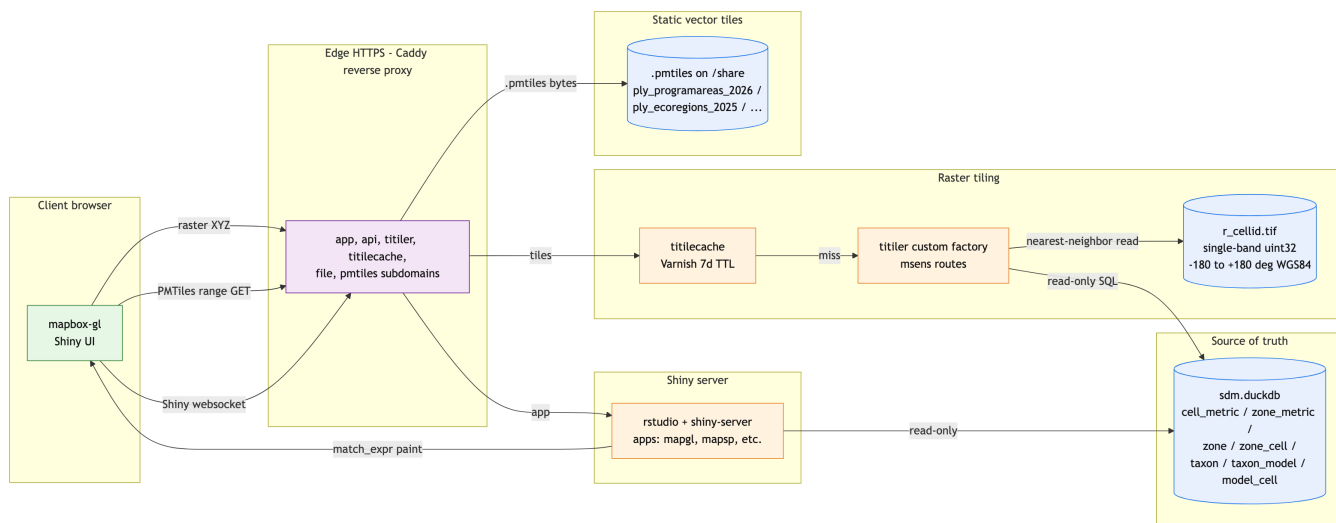


Figure 10.1: System architecture: DuckDB is the authoritative store for cell-level species/score data. A custom TiTiler factory renders those cell values on-the-fly into PNG tiles (cached by Varnish); Program Area / Ecoregion polygons are served as static PMTiles through Caddy; the Shiny apps combine both layer types and push per-Program-Area scores live from DuckDB into the polygon paint spec.

10.1.1 Interactive Applications

We have developed a series of interactive applications to explore the data and results of the MST project. These applications allow users to visualize the data, explore the results, and interact with the data in a more intuitive way. The applications are built using the [shiny](#) package in R (Chang et al. 2024), which allows us to easily create a user interface with complex reactivity for an interactive web application easily accessed through a web browser. The applications are designed to be user-friendly and intuitive, with interactive maps, charts, and tables that allow users to explore the data in a more dynamic way.

10.1.2 Overcoming Challenges with Large Spatial Data

The MST project incorporates many large spatial datasets that are problematic to render in a typical interactive application. For instance, the most common interactive mapping R package [leaflet](#) has a 4 MB limitation for displaying rasters (see “Large Raster Warning” in [Raster Images • leaflet](#)). Vectors (points, lines and polygons) get smoothed when containing many

vertices, but contiguity between polygons is lost and rendering degrades depending on the user’s connection speed.

To work around these limitations we use a “cloud native” stack (see also the [Cloud-Optimized Geospatial Formats Guide](#)) that transfers only the pixels / features visible in the current viewport:

- **Raster cell values** are rendered on-the-fly from DuckDB by a custom TiTiler factory, cached by Varnish, and delivered as small 256×256 PNG tiles.
- **Vector polygons** (Program Areas, Ecoregions, Planning Areas, ...) are served as static [PMTiles](#) archives by the Caddy file server — a single archive per layer, range-read by the browser.
- **Choropleth values** (per-Program-Area scores) are pushed live from DuckDB into the PMTiles paint spec via a `mapbox-gl match` expression, so geometry stays static but color updates as the user changes metric / subregion.

Let’s take a closer look at each.

10.1.2.1 Raster: DuckDB cell values rendered via a custom TiTiler factory

Each metric (e.g., `extrisk_mammal`) resolves, in DuckDB’s `cell_metric` table, to a list of (`cell_id`, `value`) rows — one row per 0.05°-resolution grid cell in the US EEZ. To show these in the browser we would historically have to materialize the full raster (4 GiB for the whole EEZ) and ship it over the wire.

Instead we keep the values in DuckDB and extract **only** the tiles the user’s browser is asking for. The Shiny app builds a small `SELECT`, base64-encodes it and embeds it in the tile URL template. On tile request, our custom [TiTiler](#) factory:

1. validates the SQL (SELECT-only, via `sqlglot`) and executes it once per unique query, LRU-caching the resulting `cell_id` → `value` map in-process;
2. reads the tile’s window of integer cell ids from a pre-baked single-band cell-id COG (`r_cellid.tif`) via **nearest-neighbor** resampling, so integer ids are never interpolated;
3. looks the value up for every pixel, applies rescale + colormap, and returns a PNG.

[Varnish](#) sits in front of the factory and caches each tile URL for 7 days, so once a tile has been rendered for a given SQL + colormap + rescale combination, every subsequent viewer is served from cache in sub-millisecond time (Figure 10.2). A `mtime` query parameter tied to the source DuckDB’s modification time automatically busts the cache when the DB is rebuilt.

The same factory also supports a single-color “mask” render — the “Cells outside Program Areas” overlay in `mapgl` is built this way: a SQL that returns all cells with metric data *not* in any Program Area zone, rendered with `color=#222222` (no colormap, no rescale) as a semi-transparent dark overlay.

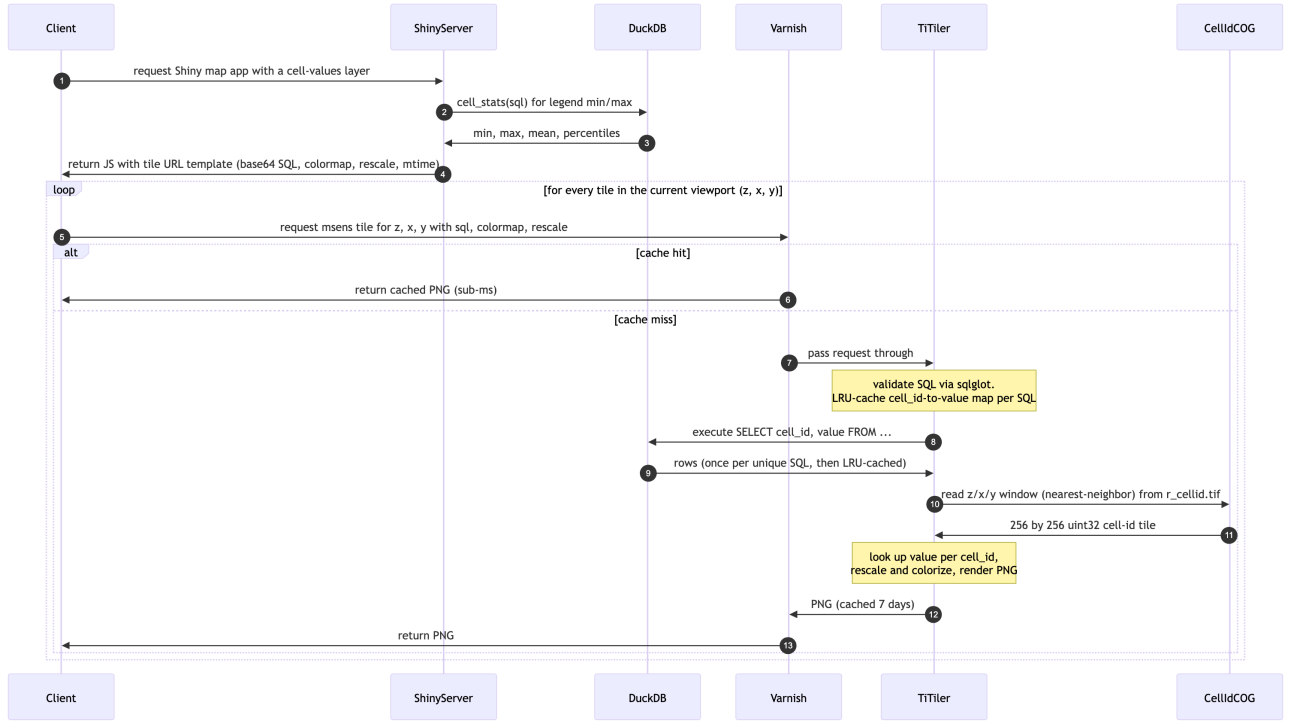


Figure 10.2: Sequence diagram: a cell-values raster layer is rendered on-the-fly by the custom msens TiTiler factory from the DuckDB source of truth. Varnish caches identical tile URLs for 7 days.

10.1.2.2 Vector: static PMTiles served by Caddy

For polygon layers (Program Areas, Ecoregions, Planning Areas, protractions, blocks, aliquots) we build [PMTiles](#) archives offline via [tippecanoe](#) and publish them as static files on the server. A PMTiles archive is a single file with an internal directory structure that lets any HTTP client issue small range requests for individual tiles; no tile server process is required — Caddy’s static file handler is all we need (Figure 10.3).

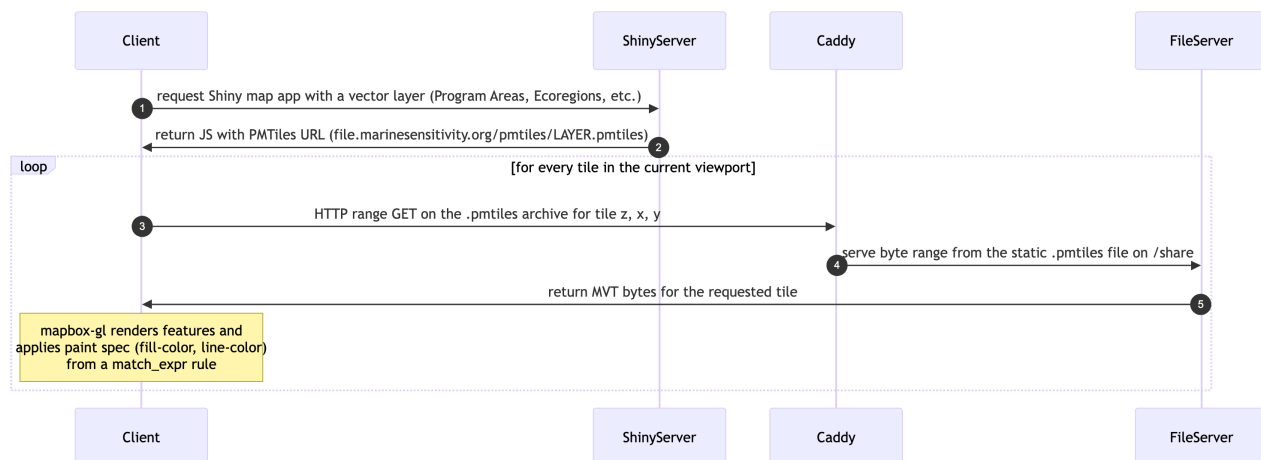


Figure 10.3: Sequence diagram: vector layers are served as static PMTiles archives through Caddy. Each tile is a byte-range read of the archive — no tile server process is in the path.

This is a deliberate simplification from the prior architecture which used PostgreSQL + PostGIS + [pg_tileserv](#). For our workload — polygon layers that change only when the data is rebuilt — a compiled tile server and live database query engine add latency and operational complexity for no benefit over byte-range reads of a pre-baked archive. (The PostgreSQL service remains in `docker-compose.yml` for compatibility with a few legacy apps; see Chapter 11.)

10.1.2.3 Choropleth: live DuckDB values + static PMTiles geometry

The Program Area choropleth is the interesting hybrid case. The *geometry* (Program Area polygons keyed by `programarea_key`) is static — it comes from the same PMTiles file as the polygon outlines. But the *fill color* depends on the currently-selected metric + subregion, which means the per-Program-Area score has to come live from DuckDB.

The Shiny server handles this by querying `zone_metric` (pre-aggregated per-zone values), building a `mapbox-gl match` expression that maps `programarea_key` → `fill_color`, and

pushing that paint rule to the map through `mapgl`'s Shiny proxy. No tiles are re-rendered; only the paint specification changes (Figure 10.4).

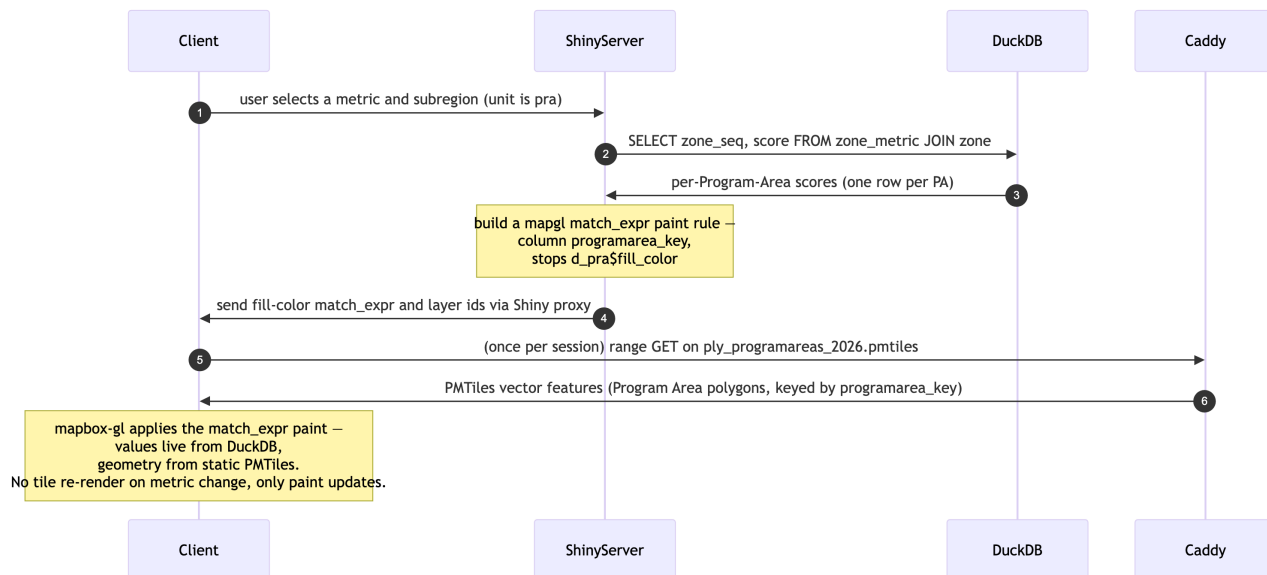


Figure 10.4: Sequence diagram: the Program Area choropleth. Polygon geometry is static (PMTiles from Caddy file server); per-area fill values come live from DuckDB's `zone_metric` table and are pushed into mapbox-gl's paint spec via a match expression. Switching metrics updates only the paint rule, not the tiles.

10.1.2.4 Data pipeline — how the source of truth is built

The three serving paths above all sit downstream of the same data pipeline: ingest workflows populate DuckDB from 7 source datasets (AquaMaps, IUCN, NMFS/FWS, BirdLife, WoRMS/eBird taxa, BOEM zones, ...); `calc_scores.qmd` computes per-cell and per-zone metrics back into DuckDB; and a one-shot `make_cellid_cog.py` bakes the cell-id COG consumed by the TiTiler factory. PMTiles archives are built offline from the same zone geopackages used to populate the `zone` table (Figure 10.5).

10.1.3 Github Repositories

repo	description
.github	organization README
analytics	NA
api	application programming interface (API) using R Plumber package

repo	description
apps	Shiny applications
docs	documentation for BOEM's offshore environmental sensitivity index products
manuscripts	Manuscripts with review of sensitivities by industry and receptors (species, habitats, human uses)
MarineSensitivity.github.io	default website
msens	R library of functions for mapping marine sensitivities, sponsored by BOEM
objectives	repository for issues spanning multiple repositories and doing big picture roadmapping
server	server setup for R Shiny apps, RStudio IDE, Caddy web server, h3 TiTiler service and Varnish cache
stac-sdm	STAC extension for Species Distribution Models (SDMs)
workflows	scripts for testing data analytics and visualization as well as production workflows

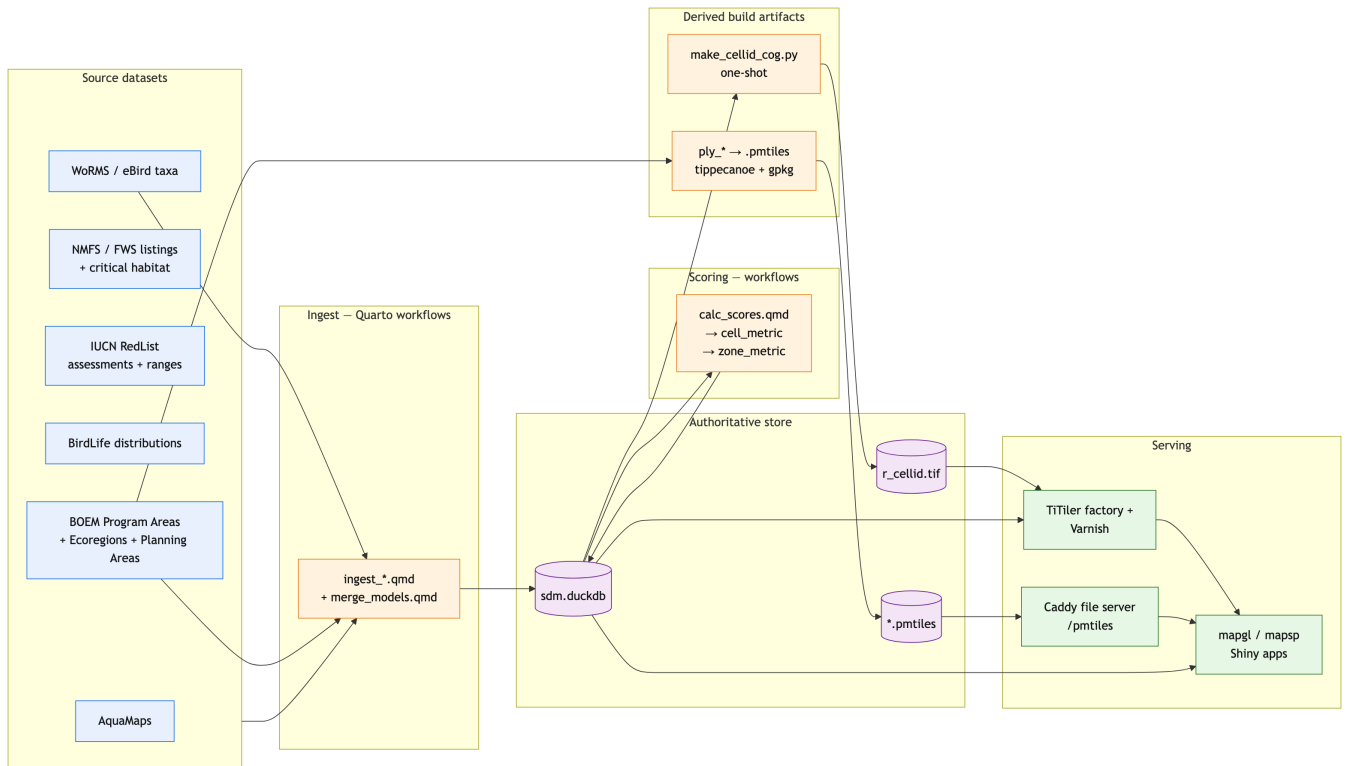
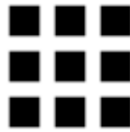


Figure 10.5: Data pipeline: ingest workflows populate DuckDB; scoring workflows compute `cell_metric` and `zone_metric` back into DuckDB; a one-shot COG regen and offline PMTiles build produce the derived artifacts the serving tier reads.

10.1.4 Software Components



APIs
application programming
interfaces (APIs)



Apps
interactive applications using
Shiny



Database
PostgreSQL database
extended spatially with
PostGIS



Docs
technical documentation



Libraries
documented functions as an R
package



Server
server software configuration
using Docker



Workflows
scripts for exploring plus
production workflows

11 Server

The server is for serving up any web services outside those of Github (e.g., [website](#), [docs](#) and R package [msens](#)) using [Docker](#) (see the [docker-compose.yml](#); with reverse proxying from subdomains to ports by [Caddy](#)).

The full architecture is summarized in the [Software overview](#). This chapter documents the concrete services running on the host and how Caddy maps subdomains to container ports.

11.1 Setup

For the latest instructions on launching an Amazon instance and installing the server software, see [Server Setup · MarineSensitivity/server Wiki](#), which is pasted below for convenience...

on AWS as EC2 instance using Docker

11.1.1 launch instance

name: **msens1**:

- Software Image (AMI)
Canonical, **Ubuntu**, 22.04 LTS, amd64 jammy image build on 2023-09-19 ami-0fc5d935ebf8bc3bc
- Virtual server type (instance type)
t2.xlarge (4 vCPU, 16 GB memory)
- Firewall (security group)
New security group
- Storage (volumes)
2 volume(s)
 - **20 GB**
/ server software, disposable
 - **60 GB**
/share for all data, persistent and to be backed up

11.1.1.1 allocate IP address

- [Elastic IP addresses | EC2 | us-east-1](#) for persistent IP address
- Allocated IPv4 address: 100.25.173.0
- Associate Elastic IP address

11.1.2 ssh to server

```
pem='/Users/bbest/My Drive/private/msens_key_pair.pem'  
ssh -i $pem ubuntu@msens1.marinesensitivity.org
```

11.1.2.1 set hostname

- [Change the hostname of your Amazon Linux instance - Amazon Elastic Compute Cloud](#)

```
sudo vi /etc/cloud/cloud.cfg  
# preserve_hostname: true  
sudo hostnamectl set-hostname msens1.marinesensitivity.org  
sudo reboot
```

11.1.2.2 mount volume

The extra volume (60 GB for /share) was added during EC2 launch instance wizard, but needs to be mounted before available for use.

- [Make an Amazon EBS volume available for use on Linux - Amazon Elastic Compute Cloud](#)

```
df -H
```

Filesystem	Size	Used	Avail	Use%	Mounted on
/dev/root	21G	2.3G	19G	11%	/
tmpfs	8.4G	0	8.4G	0%	/dev/shm
tmpfs	3.4G	898k	3.4G	1%	/run
tmpfs	5.3M	0	5.3M	0%	/run/lock
/dev/xvda15	110M	6.4M	104M	6%	/boot/efi
tmpfs	1.7G	4.1k	1.7G	1%	/run/user/1000

```
lsblk
```

NAME	MAJ:MIN	RM	SIZE	RO	TYPE	MOUNTPOINTS
loop0	7:0	0	24.6M	1	loop	/snap/amazon-ssm-agent/7528
loop1	7:1	0	55.7M	1	loop	/snap/core18/2790
loop2	7:2	0	63.5M	1	loop	/snap/core20/2015
loop3	7:3	0	111.9M	1	loop	/snap/lxd/24322
loop4	7:4	0	40.8M	1	loop	/snap/snapd/20092
xvda	202:0	0	20G	0	disk	
xvda1	202:1	0	19.9G	0	part	/
xvda14	202:14	0	4M	0	part	
xvda15	202:15	0	106M	0	part	/boot/efi
xvdb	202:16	0	60G	0	disk	

```
sudo file -s /dev/xvdb
# /dev/xvdb: data
```

So no file system on /dev/xvdb yet.

```
sudo mkfs -t xfs /dev/xvdb
sudo mkdir /share
sudo mount /dev/xvdb /share
```

```
sudo cp /etc/fstab /etc/fstab.orig
sudo blkid
# /dev/xvdb: UUID="bc766dfb-1c42-49cf-9320-2242a2d48a2e" BLOCK_SIZE="512" TYPE="xfs"
sudo vim /etc/fstab
# UUID=bc766dfb-1c42-49cf-9320-2242a2d48a2e /share xfs defaults,nofail 0 2

df -h
sudo umount /share ; df -h
sudo mount -a ; df -h
```

11.1.3 install docker

Following:

- [Step-by-Step Guide to Install Docker on Ubuntu in AWS | by Srija Anaparthi | Medium](#)

```
sudo apt-get update
#OLD: sudo apt-get install docker.io -y
```

NEW: [[Migrate to docker compose]]

```
sudo systemctl start docker
sudo docker run hello-world
sudo systemctl enable docker
docker --version
# Docker version 24.0.6, build ed223bc
sudo usermod -a -G docker $(whoami)
```

11.1.3.1 run docker compose

- [/Users/bbest/My Drive/private/msens_server_env-password.txt](#)

```
sudo chown -R ubuntu:ubuntu /share
mkdir -p /share/github/MarineSensitivity
cd /share/github/MarineSensitivity
# clone server repo
git clone https://github.com/MarineSensitivity/server.git
cd server

# add password, used as $PASSWORD in docker-compose.yml
echo 'PASSWORD=*****' > .env

# launch docker instances
sudo docker-compose up -d
```

11.1.4 Backup /share with snapshots

Per [Automate snapshot lifecycles - Amazon Elastic Compute Cloud](#), created two policies:

- **bkup_msens-share__daily** every 24 hrs at 09:00 UTC, max of 7
- **bkup_msens-share__weekly** every Monday 09:00 UTC, max of 8

11.2 Docker compose

The Docker compose file is used to define and run multi-container Docker applications. Here is the [docker-compose.yml](#) file for the server pasted for convenience ...

```
version: "3.9"

services:
  caddy:
    container_name: caddy
    build: ./caddy
    ports:
      - 80:80
      - 443:443
    restart: unless-stopped
    volumes:
      - ./caddy/Caddyfile:/etc/caddy/Caddyfile
      - /share:/share
      - /share/caddy/data:/data
      - /share/caddy/config:/config

  rstudio:
    container_name: rstudio
    build: ./rstudio
    environment:
      ROOT: 'true'
      # DEFAULT_USER, not USER: rocker's /rocker_scripts/init_userconf.sh opens with
      #   DEFAULT_USER=${DEFAULT_USER:-rstudio}
      #   USER=${DEFAULT_USER}
      # so a `USER:` set here is Clobbered on line 6 before it is ever read. This
      # said `USER: admin`, so no admin account was ever created - only `rstudio` -
      # and logging into rstudio.marinesensitivity.org as admin failed with
      # "Incorrect or invalid username/password" while the password was perfectly
      # correct. Takes effect on the next container recreate; until then the login
      # is `rstudio`.
      DEFAULT_USER: admin
      PASSWORD: ${PASSWORD}
      ADD: shiny
      # usage-log Sheet endpoint (Apps Script /exec) for the Shiny apps'
      # msens::ga_js() beacon. Set MSENS_LOG_URL in the server .env alongside
      # PASSWORD; if unset the Sheet leg is a silent no-op and only GA4 receives
      # events. The apps hold no credential - the endpoint is the whole secret,
```

```

    # so it belongs in .env (untracked), never in the app source.
    MSENS_LOG_URL: ${MSENS_LOG_URL:-}
ports:
  - 8787:8787 # rstudio
  - 3838:3838 # shiny
restart: unless-stopped
volumes:
  - /share:/share
  - /share/shiny_apps:/srv/shiny-server

# plumber: DuckDB-backed custom API (postgres stack retired 2026-06). Serves
# the OBIS H3 summary endpoint /h3 (obisindicators::obis_h3t_sql) from the same
# store as the h3t tile service. Caddy already routes
# api.marinesensitivity.org -> plumber:8888. No postgis dependency: the
# postgres-backed endpoints degrade (see api/plumber.R lazy `con`).
plumber:
  container_name: plumber
  build: ./plumber
  environment:
    MSENS_OBIS_DUCKDB: /share/data/obis/obis_h3.duckdb
  ports:
    - 8888:8888 # api
  volumes:
    - /share:/share
  restart: unless-stopped

# --- postgres stack retired (2026-06) to free resources for OBIS h3 build ---
# plumber:
#   container_name: plumber
#   build: ./plumber
#   ports:
#     - 8888:8888 # api
#   restart: unless-stopped
#   volumes:
#     - /share:/share
#   depends_on:
#     - postgis

# postgis:
#   container_name: postgis
#   image: postgis/postgis:latest
#   environment:

```

```

# POSTGRES_DB: msens
# POSTGRES_USER: admin
# POSTGRES_PASSWORD: ${PASSWORD}
# ANON_PASSWORD: ${ANON_PASSWORD}
# PGDATA: /share/postgis/data
# volumes:
#   - ./postgis/init.sh:/docker-entrypoint-initdb.d/init.sh
#   - /share:/share
#   - /share/postgis:/var/lib/postgresql
# restart: unless-stopped
# healthcheck:
#   test: 'exit 0'
# ports:
#   - 5432:5432

# pgadmin:
#   container_name: pgadmin
#   image: dpape/pgadmin4:8.14
#   restart: always
#   environment:
#     PGADMIN_DEFAULT_EMAIL: ben@ecoquants.com
#     PGADMIN_DEFAULT_PASSWORD: ${PASSWORD}
#     PGADMIN_LISTEN_PORT: 8088
#   ports:
#     - 8088:8088
#   volumes:
#     - /share/pgadmin:/var/lib/pgadmin
#   depends_on:
#     - postgis

# pgbkups:
#   container_name: pgbkups
#   image: prodrigestivill/postgres-backup-local
#   restart: always
#   user: postgres:postgres
#   volumes:
#     - /share/postgis_backups:/backups
#   links:
#     - postgis
#   depends_on:
#     - postgis
#   environment:

```

```

# - POSTGRES_HOST=postgis
# - POSTGRES_DB=msens
# - POSTGRES_USER=admin
# - POSTGRES_PASSWORD=${PASSWORD}
# - POSTGRES_EXTRA_OPTS=-Z6 --blobs
# - SCHEDULE=@daily
# - HEALTHCHECK_PORT=8088
# - BACKUP_KEEP_DAYS=2
# - BACKUP_KEEP_WEEKS=1
# - BACKUP_KEEP_MONTHS=2

# tile:
#   container_name: tile
#   environment:
#     DATABASE_URL: 'postgresql://admin:${PASSWORD}@postgis:5432/msens'
#   image: pramsey/pg_tileserv:latest
#   depends_on:
#     - postgis
#   ports:
#     - 7800:7800

# tilecache:
#   container_name: tilecache
#   image: varnish:latest
#   volumes:
#     - /share:/share
#   ports:
#     - 6081:6081
#   environment:
#     VARNISH_BACKEND_HOST: tile
#     VARNISH_BACKEND_PORT: 7800
#     VARNISH_HTTP_PORT: 6081
#   restart: always
#   depends_on:
#     - "tile"
# --- end postgres stack ---

# rest:
#   container_name: rest
#   environment:
#     PGRST_DB_URI: 'postgresql://anon:${ANON_PASSWORD}@postgis:5432/msens'
#     PGRST_OPENAPI_SERVER_PROXY_URI: http://127.0.0.1:3000

```

```

#   PGRST_DB_ANON_ROLE: anon
#   image: postgrest/postgrest
#   depends_on:
#     - postgis
#   ports:
#     - "3000:3000"

# swagger:
#   container_name: swagger
#   image: swaggerapi/swagger-ui
#   depends_on:
#     - rest
#   ports:
#     - "8080:8080"
#   expose:
#     - "8080"
#   environment:
#     API_URL: https://rest.MarineSensitivity.org/

titiler:
  container_name: titiler
  build: ./titiler
  environment:
    PORT: 8000
    # WORKERS_PER_CORE: 1
    MSENS_CELLID_COG: /share/data/derived/r_cellid.tif
    MSENS_DUCKDB: /share/data/big/latest/sdm.duckdb
    MSENS_MAX_ROWS: "1000000"
    MSENS_LRU_SIZE: "128"
  volumes:
    - /share:/share
  ports:
    - "8000:8000"
  restart: always

# v8 parallel serving: a tiny view-only DuckDB whose model_cell is a VIEW over the
# partitioned Parquet on S3 (marine-atlas/v8), rendered on the global [-180,180] cell COG.
# v7 titiler above stays fully live (A/B). serve.duckdb is KB - no multi-GB rsync.
titiler-v8:
  container_name: titiler-v8
  build: ./titiler
  environment:

```



```

    PORT: 8000
    MSENS_CELLID_COG: /share/data/derived/r_cellid_global.tif
    MSENS_DUCKDB: /share/data/big/v8/serve.duckdb
    MSENS_MAX_ROWS: "1000000"
    MSENS_LRU_SIZE: "128"
  volumes:
    - /share:/share
  ports:
    - "8001:8000"
  restart: always

# stac-api: a SEARCHABLE STAC API with one Item per model, complementing (not
# replacing) the static catalog at file.marinesensitivity.org/stac.
#
# The static catalog is dataset-level: one Item per collection whose assets are
# S3 PREFIXES (.../native/am_native/, which 403s) and whose bbox is the whole
# dataset envelope - so it can answer neither "the asset for this model" nor
# "which models cover this area". This serves 25,643 per-model Items with real
# asset hrefs and real footprints.
#
# Backend reads stac-geoparquet via DuckDB - the same Parquet-on-DuckDB shape as
# the rest of this stack. Items and collection.json come from the tree that
# workflows/publish_stac_api.qmd builds and rsyncs; PARQUET_URLS_JSON (collection
# id -> parquet path) is passed by that notebook from parquet_urls.json, so
# adding a dataset needs no edit here.
#
# Pinned to an exact upstream commit built straight from GitHub - the project is
# marked EXPERIMENTAL by both its repo and the stac-fastapi docs (pgstac and
# Elasticsearch are the production-ready backends), so a moving default branch
# is not something to serve from. Treat this as a trial surface, not load-bearing.
stac-api:
  container_name: stac-api
  build: ./stac-api # upstream pinned + ST_Intersects(WKB) patch; see its Dockerf
  environment:
    APP_HOST: 0.0.0.0
    APP_PORT: "8084"
    ENVIRONMENT: production
    WEB_CONCURRENCY: "4"
    BACKEND: duckdb
    STAC_FILE_PATH: /app/stac_collections
    PARQUET_URLS_JSON: ${PARQUET_URLS_JSON}
  volumes:

```

```

    # read-only: the API must never mutate the published catalog
    - /share/data/derived/stac-api/v8:/app/stac_collections:ro
ports:
  - "8085:8084"
restart: unless-stopped
command: bash -c "python -m stac_fastapi.duckdb.app"

titilecache:
  container_name: titilecache
  image: varnish:latest # 7.4.2 # last updated: 2023-12-26
  volumes:
    - /share:/share
    - "/varnish/titiler.vcl:/etc/varnish/default.vcl:ro"
  ports:
    - 6082:6082
  environment:
    VARNISH_BACKEND_HOST: titiler
    VARNISH_BACKEND_PORT: 8000
    VARNISH_HTTP_PORT: 6082
  command: "-p default_keep=604800" # 7d, matches vcl TTL
  restart: always
  depends_on:
    - "titiler"

# h3t: OBIS biodiversity-by-hex tile factory (DuckDB SELECT -> h3j JSON tiles)
h3t:
  container_name: h3t
  build: ./h3t
  environment:
    H3T_DBS: "obis:/share/data/obis/obis_h3.duckdb"
    H3T_DEFAULT_DB: obis
    H3T_MAX_ROWS: "50000"
    # common filtered maps now hit precomputed idx_h3 / idx_h3_taxon layers;
    # these caps are the stopgap for the remaining live occ_h3 paths (year
    # filters, multi-value / fine-rank taxa). raise the timeout above the
    # app-side client timeout in apps/h3-db/app.R.
    H3T_STMT_TIMEOUT_MS: "8000"
    H3T_THREADS: "4" # cap serving CPU (don't peg all cores)
    H3T_MEMORY_LIMIT: "6GB" # cap serving RAM (occ_h3 rollups can spill)
  volumes:
    - /share:/share
  ports:

```

```

- "8889:8889"
restart: always

h3tcache:
  container_name: h3tcache
  image: varnish:latest # 7.4.2 # last updated: 2023-12-26
  volumes:
    - /share:/share
    - "/varnish/h3t.vcl:/etc/varnish/default.vcl:ro"
  ports:
    - 6083:6083
  environment:
    VARNISH_BACKEND_HOST: h3t
    VARNISH_BACKEND_PORT: 8889
    VARNISH_HTTP_PORT: 6083
  command: "-p default_keep=604800" # 7d, matches vcl TTL
  restart: always
  depends_on:
    - "h3t"

```

11.3 DNS

The domain name server (DNS) records are managed by [SquareSpace](#). The subdomains point to the server on Amazon at 100.25.173.0, whereas the main website is hosted by Github servers, per [Managing a custom domain for your GitHub Pages site - GitHub Docs](#).

Host	Type	Data
@	A	185.199.111.153
@	A	185.199.110.153
@	A	185.199.109.153
@	A	185.199.108.153
api	A	100.25.173.0
app	A	100.25.173.0
file	A	100.25.173.0
msens1	A	100.25.173.0
pgadmin	A	100.25.173.0
pmtiles	A	100.25.173.0
rstudio	A	100.25.173.0
shiny	A	100.25.173.0
tile	A	100.25.173.0

Host	Type	Data
tilecache	A	100.25.173.0
titiler	A	100.25.173.0
titilecache	A	100.25.173.0
www	CNAME	marinesensitivity.org

11.4 Caddyfile

The Caddyfile parameterizes the reverse proxying between the external subdomains and the Docker's internal ports. Here is the [Caddyfile](#) pasted for convenience ...

```
# docker exec caddy caddy reload --config /etc/caddy/Caddyfile
{
    order pmtiles_proxy before file_server
}

(cors) {
    @origin header Origin *
    header @origin {
        Access-Control-Allow-Origin "*"
        Access-Control-Request-Method GET
    }
}

api.marinesensitivity.org {
    reverse_proxy plumber:8888
}

file.marinesensitivity.org {
    import cors

    # raw pmtiles for client-side PMTiles protocol (add_pmtiles_source)
    handle_path /pmtiles/* {
        root * /share/data/derived/pmtiles
        file_server browse
    }

    # legacy versioned pmtiles routes (backwards compat for old apps)
    handle_path /pmtiles/v4/* {
        root * /share/data/derived/v4/pmtiles
    }
}
```

```

        file_server browse
    }
    handle_path /pmtiles/v3/* {
        root * /share/data/derived/v3/pmtiles
        file_server browse
    }

    # stac catalog (sdm extension) - static JSON tree
    handle_path /stac/* {
        root * /share/data/derived/stac
        file_server browse
    }

    # versioned derived data (geoparquet, cogs, gpkg) referenced by stac assets
    handle_path /derived/* {
        root * /share/data/derived
        file_server browse
    }

    # force reports to download (especially HTML which otherwise renders inline)
    handle_path /reports/* {
        root * /share/public/reports
        header Content-Disposition "attachment"
        file_server
    }

    handle {
        root * /share/public
        file_server browse {
            precompressed zstd br gzip
        }
    }
}

# Browsable front door for the PUBLIC S3 bucket.
#
# Why this exists: S3 serves OBJECTS, not directories. A URL ending in "/" 404s
# unless an object literally has that key, and anonymous ListBucket is DENIED on
# this bucket (403 AccessDenied), so a browser cannot enumerate anything. The
# workflows notebook publish_storage_index.qmd generates a real index.html per
# directory; this vhost rewrites folder URLs to them:
#

```

```

# storage.marinesensitivity.org/marine-atlas/v8/ -> ../v8/index.html
# storage.marinesensitivity.org/marine-atlas/latest.txt -> the object
#
# Named "storage" rather than "s3" so the URL survives a move to another
# provider.
storage.marinesensitivity.org {
    encode zstd gzip

    log {
        output file /share/logs/caddy/storage.log {
            roll_size 64MiB
            roll_keep 14
            roll_keep_for 2160h
        }
        format json
    }

    # robots.txt inline, BEFORE the bucket handler, so it cannot be lost by an
    # index regeneration. The catalog pages are welcome in search results; the
    # objects behind them are not. Every byte is PROXIED through this server
    # (reverse_proxy, not a redirect), and the tree is 18 GB across ~90,000
    # objects, so a crawl would be billed as VM egress.
    handle /robots.txt {
        header Content-Type "text/plain; charset=utf-8"
        respond `# Catalog pages: crawl away. Data objects: please do not.
User-agent: *
Crawl-delay: 10
Disallow: /*.tif$
Disallow: /*.parquet$
Disallow: /*.duckdb$
Disallow: /*.pmtiles$
Disallow: /*.gz$
Disallow: /*.zip$
` 200
    }

    # ONLY the published atlas. The same bucket holds backups/ -- including a
    # multi-GB database dump -- which is technically public but must not be
    # advertised or reachable here, even by exact path. This allow-list is the
    # server-side half of the restriction the index generator applies.
    @atlas path_regexp atlas ~/marine-atlas(/|$)
    handle @atlas {

```

```

    # folder URL -> the index.html object standing in for a listing
    @dir path_regexp dir ^(.*)/$
    rewrite @dir {re.dir.1}/index.html
    reverse_proxy https://s3.us-east-1.amazonaws.com {
        header_up Host s3.us-east-1.amazonaws.com
        rewrite /oceanmetrics.io-public{uri}
    }
}

# the bare host lands on the atlas index
redir / /marine-atlas/ 302

# anything else: refuse, and say where to go
handle {
    header Content-Type "text/plain; charset=utf-8"
    respond `Not found.

Browsable: /marine-atlas/
Start here: https://storage.marinesensitivity.org/marine-atlas/
` 404
}

pgadmin.marinesensitivity.org {
    reverse_proxy pgadmin:8088
}

# rest.marinesensitivity.org {
#     reverse_proxy rest:3000
# }

# searchable STAC API (one Item per model) - complements the STATIC catalog served
# from file.marinesensitivity.org/stac, which stays exactly as it is.
stac-api.marinesensitivity.org {
    reverse_proxy stac-api:8084
}

rstudio.marinesensitivity.org {
    reverse_proxy rstudio:8787
}

shiny.marinesensitivity.org {

```

```

    reverse_proxy rstudio:3838
}

app.marinesensitivity.org {
    reverse_proxy rstudio:3838
}

shiny.oceanmetrics.io {
    reverse_proxy rstudio:3838
}

# swagger.marinesensitivity.org {
#     reverse_proxy swagger:8080
# }

tile.marinesensitivity.org {
    reverse_proxy tile:7800
}

tilecache.marinesensitivity.org {
    reverse_proxy tilecache:6081
}

titiler.marinesensitivity.org {
    reverse_proxy titiler:8000
}

titiler-v8.marinesensitivity.org {
    reverse_proxy titiler-v8:8000
}

titilecache.marinesensitivity.org {
    reverse_proxy titilecache:6082
}

h3t.marinesensitivity.org {
    reverse_proxy h3t:8889
}

h3tcache.marinesensitivity.org {
    reverse_proxy h3tcache:6083
}

```



```

pmtiles.marinesensitivity.org {
  import cors

  # server-side tile decoding: ZXY + TileJSON (for QGIS, leaflet, etc.)
  handle_path /tiles/* {
    pmtiles_proxy {
      bucket file:///share/data/derived/v4/pmtiles
      # note: v3 pmtiles also available at /share/data/derived/v3/pmtiles
      cache_size 256
      public_url https://pmtiles.marinesensitivity.org/tiles
    }
  }
}

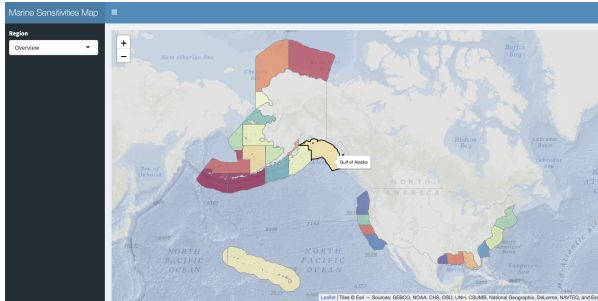
```

11.5 Services

The server is running the following services. Subdomains with `internal` are internal (admin / development); others are public.

11.5.1 App + data plane

- [app](#) / [shiny](#)
interactive Shiny applications (scores, species, ...) — proxied to the `rstudio` container's Shiny Server on internal port 3838



[Shiny Server docs](#)

- [file](#)
static file server for PMTiles archives, report downloads, reference GeoPackages — served by Caddy's `file_server` directive directly from the `/share` bind-mount. This replaces the old `pg_tileserv` vector-tile path for Program Areas, Ecoregions, Planning Areas, protraction, blocks and aliquots; each layer is a single `.pmtiles` file that the browser range-reads for only the tiles it needs.

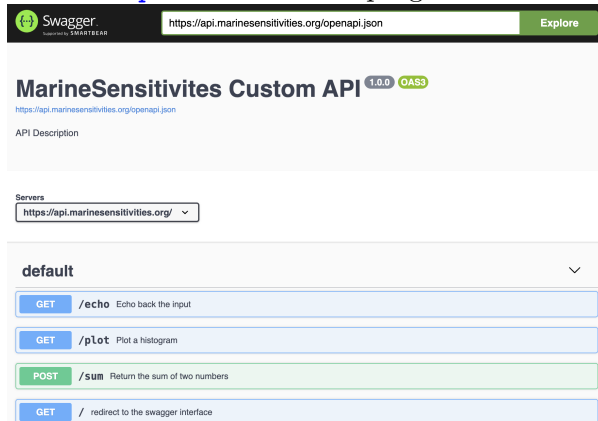
- **titiler** + **tilecache**

*custom msens **TiTiler** factory fronted by a **Varnish** cache.*

The factory exposes `/msens/*` routes that execute a validated SELECT against DuckDB, look the result up against a pre-baked cell-id COG, and return on-the-fly colorized PNG tiles (see Chapter 14). **tilecache** keys on the full URL (including base64 SQL) and caches each tile for 7 days with `default_keep=604800`; the custom VCL (`server/varnish/titiler.vcl`) normalizes query-param order via `std.queriesort()` so semantically-identical URLs share a cache entry, strips cookies / authorization (the factory is stateless), and adds an `X-Cache: HIT | MISS` debug header.

- **api**

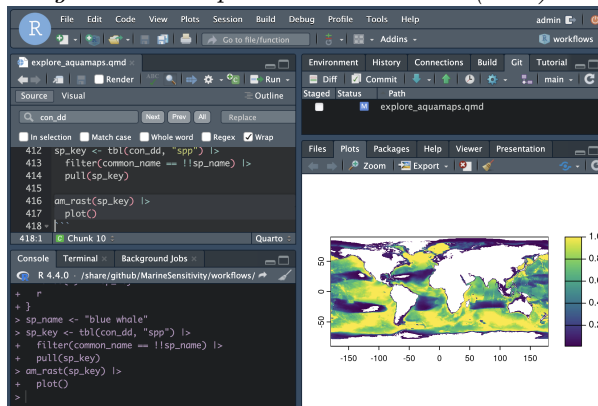
*custom R **plumber** API for programmatic access to DuckDB — see Chapter 14*



11.5.2 Admin

- **rstudio**

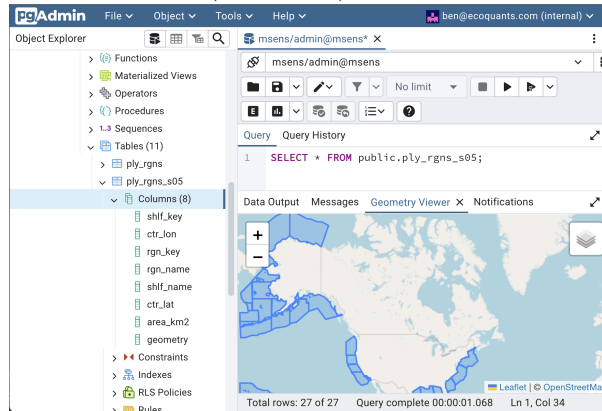
integrated development environment (IDE) to code and debug directly on the server.



Posit RStudio Server

- **pgadmin**

PostgreSQL database administration interface — used for legacy apps that still connect to PostgreSQL (see below).



pgAdmin

11.5.3 Legacy (carried for backward-compat; not in the critical path)

As of 2026-04, the new raster tile path (`titiler` + `tilecache`) and `PMTiles-via-file` replace what was previously handled by PostgreSQL + `pg_tileserv`. The **authoritative store is now DuckDB** (`/share/data/big/latest/sdm.duckdb`). The following services remain in `docker-compose.yml` only for backward compatibility with a small set of pre-2026 apps (`indicators`, `bird_hotspots`) that still read from PostgreSQL. New app code should target DuckDB + the `TiTiler` factory + `PMTiles`.

- **postgis** (`postgis/postgis:latest`, internal 5432)
PostgreSQL + PostGIS — formerly the primary spatial database; now read by legacy apps only.
- **tile** (`pramsey/pg_tileserv:latest`, internal 7800) + **tilecache** (Varnish, 6081)
vector tiles from PostGIS tables — superseded by static `PMTiles` served by Caddy.
- **rest** and **swagger** — `PostgREST` + `Swagger UI`, both **commented out** in `docker-compose.yml`.

...

12 Database

The **source of truth** for the Marine Sensitivity Toolkit is a single [DuckDB](#) file, `sdm.duckdb`, stored on the server at `/share/data/big/latest/sdm.duckdb` (a symlink to the current versioned directory, currently `v6`). DuckDB is a columnar, embedded analytic database — it's fast enough to power live `cell_metric` queries from the TiTiler factory while also being a single-file artifact that can be copied, versioned and served read-only to multiple processes concurrently.

All species distributions (merged from AquaMaps, BirdLife, IUCN and others), per-cell scores (`cell_metric`) and per-zone aggregates (`zone_metric`) live here. The Shiny apps, the custom plumber API and the TiTiler tile factory all read the same file in read-only mode; workflows under `workflows/` are the only writers.

Note

Why DuckDB and not PostgreSQL? Earlier versions of the stack used PostgreSQL + PostGIS as the primary store. As the dataset stabilized around a fixed 0.05° cell grid and a small number of zone layers, we no longer needed live geometry operations or concurrent writes — just fast analytic reads. DuckDB gives us columnar scans over tens of millions of `cell_metric` rows in well under a second, without a separate database server process. PostgreSQL is retained in `docker-compose.yml` only for a handful of legacy apps (see Chapter 11).

12.1 Table and column naming conventions

- Table names are plural and use lower case.
- Unique identifier columns are suffixed with:
 - `*_id` for unique integer keys (stable across rebuilds);
 - `*_key` for unique string keys (human-readable);
 - `*_seq` for auto-incrementing sequence integer keys (internal join keys).
- Column names are singular and use `snake_case`.
- Foreign keys use the singular form of the referenced table, followed by the primary-key suffix (`_id`, `_key`, or `_seq`).
- Primary keys are named `id` when a table has a single natural key.

12.2 Core tables

The stack revolves around 10 tables; the ones touched by the tile factory and the Shiny apps are:

Table	Rows (v6)	Purpose	Read by
<code>cell</code>	660 k	0.05°-resolution grid cells covering the US EEZ footprint.	workflows, apps
<code>cell_metric</code>	10 M	(<code>cell_id</code> , <code>metric_seq</code> , <code>value</code>) — the per-cell score for each metric.	TiTiler factory , apps
<code>metric</code>	20	Metric metadata (<code>metric_key</code> , <code>metric_seq</code> , <code>description</code>).	factory, apps
<code>zone</code>	100	Zone registry (<code>zone_seq</code> , <code>tbl</code> , <code>fld</code> , <code>value</code>) — keyed by the source GeoPackage table, field and value, so Program Areas, Ecoregions, Subregions, Planning Areas all share one schema.	factory (mask SQL), apps
<code>zone_cell</code>	3 M	(<code>zone_seq</code> , <code>cell_id</code>) — membership mapping, used for subregion / Program Area filters.	factory, apps

Table	Rows (v6)	Purpose	Read by
<code>zone_metric</code>	2 k	Pre-aggregated (<code>zone_seq</code> , <code>metric_seq</code> , <code>score</code>) — the per-Program-Area / per-subregion score values that drive the choropleth.	apps
<code>taxon</code>	17 k	Unified taxonomy (one row per species) merged from WoRMS, eBird, IUCN.	apps
<code>taxon_model</code>	20 k	(<code>taxon_id</code> , <code>ds_key</code> , <code>mdl_seq</code>) — which species is represented by which distribution model(s).	apps
<code>model_cell</code>	30 M	(<code>mdl_seq</code> , <code>cell_id</code> , <code>value</code>) — raw species-distribution-model output per cell (0–100 suitability or similar).	apps (species explorer)
<code>dataset</code>	10	Source datasets (AquaMaps, BirdLife, NMFS CH, FWS CH, IUCN, ...).	apps

12.2.1 How the core tables feed the serving tier

- **Raster cell tiles** (Figure 10.2) — the TiTiler factory receives a base64-encoded SELECT that joins `cell_metric` with `metric` (and optionally `zone` / `zone_cell` when the user picks a subregion), returning (`cell_id`, `value`) rows. The factory LRU-caches this result per-SQL and looks each `cell_id` up against the pre-baked cell-id COG to colorize the tile.
- **“Cells outside Program Areas” mask** — a specific SQL returning every `cell_metric` `cell_id` that is **not** in any row of `zone_cell` where `zone.fld =`

- 'programarea_key'. Rendered with the factory's `color=` single-color mask path.
- **Program Area choropleth** (Figure 10.4) — Shiny queries `zone_metric` joined to `zone` (filtered to `fld = 'programarea_key'`), maps `score` → `color`, and pushes a `match_expr` paint rule to `mapbox-gl`. Geometry comes from static PMTiles; only the paint spec changes on metric switch.

The entity-relationship diagram (ERD) and detailed schema are built by the `create_sdm-tables.qmd` workflow:

- [Create SDM Tables](#)

And ingestion from a specific source is illustrated in:

- [Ingest GoMex cetacean & sea turtle SDMs](#)

See Chapter 13 for the full list of ingest / score workflows.

13 Workflows

Workflows are scripts — mostly [Quarto](#) notebooks — that ingest source datasets, merge taxonomic authorities, compute scores, and build the derived artifacts consumed by the serving tier. They are the only writers of `sdm.duckdb`; everything downstream (Shiny apps, the TiTiler factory, the `plumber` API) reads it in read-only mode.

Links:

- marinesensitivity.org/workflows — rendered HTML
- github.com/MarineSensitivity/workflows — source

See Figure 10.5 for a visual summary and Chapter 12 for the target schema.

13.1 Ingest workflows (populate `sdm.duckdb`)

Each source dataset has a dedicated `ingest_*.qmd` that writes rows to `cell_metric`, `model_cell`, `taxon`, `dataset` and related tables:

Workflow	Source
<code>ingest_aquamaps_to_sdm_duckdb.qmd</code>	AquaMaps predictions (16,800 models)
<code>ingest_sdm-nc.qmd</code> , <code>ingest_sdm-gm.qmd</code>	NOAA/GEBCO cetacean + sea-turtle distribution models
<code>ingest_taxon.qmd</code>	Unified taxonomy from WoRMS , eBird
<code>ingest_nmfs-fws-listings.qmd</code> , <code>ingest_iucn*.qmd</code>	Conservation status (ESA listings, IUCN Red List)
<code>ingest_prot.qmd</code> , <code>ingest_blocks.qmd</code> , <code>ingest_mregions.qmd</code>	Spatial zones (protractions, blocks, marine regions)
<code>merge_models.qmd</code>	Authority merging — unified <code>taxon</code> table, <code>taxon_model</code> junction, merged <code>mdl_seq</code> per species

13.2 Scoring workflow (`cell_metric` + `zone_metric`)

- **`calc_scores.qmd`** — the primary scoring pipeline. Loads `sdm.duckdb` read-write, computes extinction-risk metrics by species category (e.g., `extrisk_bird`, `extrisk_fish`, `extrisk_mammal`), writes rows to `cell_metric`, and aggregates into `zone_metric` for both Program Area and subregion zones. Also exports helper caches (`taxon.csv`, `layers_v6.csv`, `flower_default_subregions.csv`) used by the Shiny apps at startup.
- **`update_scores.qmd`** — targeted remap when the score scale changes (e.g., the 2026 migration from the old 70/90 critical-habitat scale to the unified `er_score` EN=100 / TN=50 / LC=1 scale).

13.3 Derived artifacts for the serving tier

Two one-shot build steps produce the binary artifacts that sit next to `sdm.duckdb` on `/share`:

- **`server/titiler/scripts/make_cellid_cog.py`** — one-shot script that reads band 1 of the source multi-band raster (`r_bio-oracle_planarea.tif`, stored in 0–360° longitude convention for contiguous US EEZ coverage across the Pacific), casts `NaN` → 0 + `float32` → `uint32`, wraps longitudes to standard `−180..180°`, and writes a single-band tiled GeoTIFF at `/share/data/derived/r_cellid.tif`. No overviews — nearest-neighbor reads at native resolution are fast enough for `z = 4` and interpolating overviews would corrupt integer cell ids.
- **`PMTiles builds`** — offline `tippecanoe` runs over the source GeoPackages (`ply_programareas_2026`, `ply_ecoregions_2025`, `ply_planareas_2025`, ...) produce `.pmtiles` archives, published to `/share/pmtiles/` and served by Caddy at `file.marinesensitivity.org/pmtiles/<layer>.pmtiles`.

Re-run either when the underlying data changes; the Shiny app's `mtime` cache-busting parameter (tied to `sdm.duckdb`'s `mtime`) invalidates cached tiles automatically when the DB is rebuilt.

14 APIs

Two APIs sit in front of the DuckDB source of truth:

1. tilecache.marinesensitivity.org — a custom [TiTiler](#) factory that renders on-the-fly PNG raster tiles from arbitrary DuckDB SELECTs, cached by Varnish.
2. api.marinesensitivity.org — a custom R [plumber](#) API for programmatic (non-tile) data access.

Static vector data (PMTiles) is served directly by Caddy’s file server and doesn’t need an API in the traditional sense — see Chapter 11.

14.1 msens TiTiler factory

The factory is a FastAPI subclass of `titiler.core.factory.TilerFactory`, registered at the `/msens` prefix on top of the default TiTiler application. It runs in the `titiler` container (image built from `server/titiler/`) and is fronted by a Varnish cache (`tilecache`).

Source: [server/titiler/factory.py](#).

14.1.1 Endpoints

All `/msens/*` routes accept a `sql` query parameter: a **base64url-encoded SELECT** returning exactly two columns, `cell_id` (integer) and `value` (numeric). The SQL is parsed with `sqlglot` (SELECT-only, no write ops, no dynamic I/O like `read_*` / `httpfs` / `load_*`), executed against a read-only DuckDB connection, and the resulting `cell_id` → `value` map is cached in-process with an LRU keyed on the canonical SQL string. A `mtime` parameter can be passed for cache-busting (typically the mtime of `sdm.duckdb`).

Endpoint	Purpose
GET <code>/msens/bounds</code>	Geographic (EPSG:4326) bounds of the cell-id COG.

Endpoint	Purpose
GET /msens/statistics?sql=...	{n, min, max, mean, std, p2, p50, p98} of the SQL result — used by clients to set a stable legend rescale without fetching any tiles. Cached by Varnish.
GET /msens/tilejson.json?sql=...&colormap=...&rescale=...	TileJSON 2.2.0 document with the tile URL template — lets any TileJSON-aware client (QGIS, MapLibre, Mapbox GL) consume the layer.
GET /msens/tiles/{z}/{x}/{y}.png?sql=...&colormap=...&rescale=...	The tile endpoint. Renders a 256×256 RGBA-PNG by: (a) executing the SQL and LRU-caching the <code>cell_id → value</code> map; (b) reading the z/x/y window of the cell-id COG with nearest-neighbor resampling; (c) looking the value up per pixel, rescaling and colorizing; (d) returning PNG with <code>Cache-Control: public, max-age=604800</code> .
GET /msens/tiles/{z}/{x}/{y}.png?sql=...&colormap={hex}#f0f0f0	Single-color mask variant: every valid (non-zero, non-null) cell-value pixel is painted with the given hex color; <code>colormap</code> and <code>rescale</code> are ignored. Used for binary overlays like “cells outside Program Areas”.
GET /msens/debug/cog	Inspect the underlying cell-id COG (bounds, CRS, dtype, nodata, overviews, <code>rio_tiler</code> / <code>rasterio</code> versions).
GET /msens/debug/tile/{z}/{x}/{y}	Inspect what <code>rio_tiler.Reader.tile()</code> returns for a given tile (cell-id range, mask valid-pixel count, top-10 cell-ids by count).

A typical Scores Shiny app composes tile URLs via `msens::cell_tile_url()` and fetches stats via `msens::cell_stats()`; see Chapter 15.

14.1.2 Caching

[Varnish](#) fronts the factory at `titilecache:6082`. The custom VCL (`server/varnish/titiler.vcl`):

- Normalizes query-param order via `std.queriesort()` so semantically-identical URLs share a cache entry.
- Strips `Cookie` / `Authorization` (the factory is stateless).
- Caches 2xx responses on `/msens/(tiles|tilejson|bounds|statistics)` for 7 days (`grace=1h, keep=24h`); errors for 60 s to prevent stampede.

- Adds X-Cache: HIT | MISS for debugging.

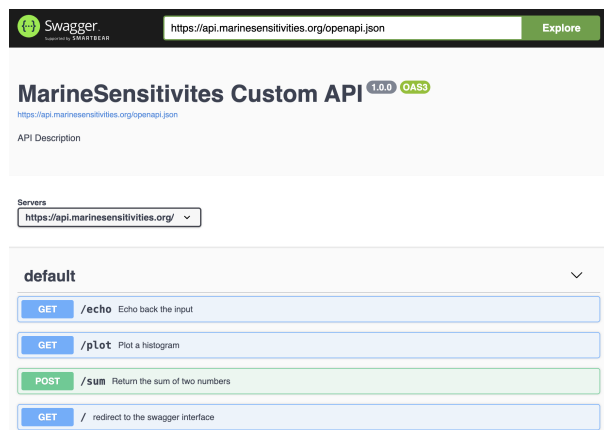
When the source DuckDB regenerates, the Shiny app bumps the `mtime` query parameter to invalidate all cached tiles at once.

14.1.3 Derived COG: `r_cellid.tif`

The factory looks values up against a single-band uint32 cell-id COG at `/share/data/derived/r_cellid.tif`. This file is rebuilt by a one-shot script (`server/titiler/scripts/make_cellid_cog.py`) any time the grid definition changes. See Chapter 13 for details.

14.2 Custom plumber API

The plumber API at api.marinesensitivity.org covers non-tile use cases — CSV exports, species lookups by feature, on-demand PDF/DOCX report generation. Source: [MarineSensitivity/api](https://api.marinesensitivity.org).



14.3 Legacy APIs (deprecated)

The following APIs were part of the pre-2026 PostgreSQL-backed stack and are **no longer in the critical path** for `scores` / `species` apps. Endpoints remain reachable for backward compatibility with legacy apps (`indicators`, `bird_hotspots`) but should not be targeted by new code.

- `swagger.marinesensitivity.org` — PostgREST over the PostgreSQL DB (currently commented out in `docker-compose.yml`).

- `tile.marinesensitivity.org` — `pg_tileserv` vector tiles from PostGIS. Superseded by static PMTiles served by Caddy's `file` subdomain; see Chapter [10](#).

15 Libraries

By creating an R package, we document functions and make them easily available across workflows and apps.

- [msens](#)

R library of functions for mapping marine sensitivities, sponsored by BOEM

Functions `read` data from Chapter 12, `analyze` it, and `visualize` results. The Shiny apps in Chapter 16 consume `msens::*` helpers exclusively — app code never talks to the tile factory or PMTiles URL templates directly.

15.1 Map-tile helpers (as of 0.3.4)

The following helpers bridge the Shiny app to the `msens` TiTiler factory (Chapter 14):

Helper	Purpose
<code>cell_tile_url(sql, colormap, rescale, color, mtime, base)</code>	Compose an XYZ tile URL template for the factory, base64url-encoding <code>sql</code> and omitting <code>colormap/rescale</code> when <code>color=</code> is set (single-color mask).
<code>cell_stats(sql, mtime, base)</code>	GET <code>/msens/statistics</code> — returns <code>list(n, min, max, mean, std, p2, p50, p98)</code> for a stable legend rescale.
<code>add_cell_tiles(m, tile_url, ...)</code>	Wraps <code>mapgl::add_raster_source()</code> + <code>mapgl::add_raster_layer()</code> so the Shiny app adds a cell-values (or mask) layer to a map in one call.
<code>add_cells(m, r, colors, ...)</code> (retained)	Legacy image-source path for static report figures that need terra <code>SpatRaster</code> rendering (e.g., Quarto-to-DOCX output that can't reach the tile server).

The SQL passed to `cell_tile_url()` and `cell_stats()` must return exactly two columns, `cell_id` (integer) and `value` (numeric) — see the examples in Chapter 12.

16 Apps

The MST provides two primary interactive web applications for exploring sensitivity data, plus several legacy applications from earlier development phases.

These tools allow stakeholders, scientists, and decision-makers to understand the underlying components contributing to an area's sensitivity to offshore energy development, supporting better planning and impact assessment.

16.1 Current Applications

16.1.1 Scores — mapper of aggregate sensitivity scores

The [Scores](#) application is the primary tool for exploring composite sensitivity scores across BOEM Program Areas. Key features include:

- **Multi-resolution viewing:** toggle between cell-level (0.05° grid) and Program Area aggregations
- **Study area selection:** focus on specific Program Areas or subregions
- **Layer selection:** view individual species category scores or composite metrics
- **Flower plot:** interactive summary showing scores by species category with weighted center value
- **Species table:** sortable table of species present in selected area with extinction risk details
- **Treemap:** hierarchical view of species contributions to sensitivity
- **CSV export:** download data for selected area or species

Under the hood, the Scores app composes three types of map layers — all driven from the same DuckDB source of truth (Figure [10.1](#)):

- **Cell-values raster** — on-the-fly PNG tiles from the msens TiTiler factory. The Shiny server builds the SELECT via `cell_sql(metric_key, subregion_key)`, fetches the stats for a stable legend via `msens::cell_stats()`, and passes the tile URL template to `msens::add_cell_tiles()`. The browser fetches only the tiles intersecting its current viewport; Varnish caches each tile for 7 days. No 4 GB full-raster transfer on initial load (see Figure [10.2](#)).

- **“Cells outside Program Areas” mask** — same factory, single-color `color=#222222` render path, with a SQL returning `cell_metric` cells **not** in any `zone` row where `fld='programarea_key'`.
- **Program Area choropleth** — when the user switches unit to Program Area, polygon geometry comes from a static PMTiles file served by Caddy; per-area fill colors come live from DuckDB’s `zone_metric` table via a `mapbox-gl` `match` expression pushed through the Shiny proxy (see Figure 10.4).

See [Appendix: scores](#) for details.

16.1.2 Species — mapper of individual species distributions

The [Species](#) application enables exploration of individual species distribution models. Key features include:

- **Species selection:** search by common or scientific name, filter by species category
- **Model display:** view merged models or individual source dataset layers
- **IUCN masking toggle:** compare masked vs. unmasked distributions
- **Species info panel:** ESA listing, IUCN Red List category, MMPA/MBTA protection flags, extinction risk score
- **Deep-linking:** URL parameters support direct links to specific species

See [Appendix: species](#) for details.

16.2 Legacy Applications

Earlier versions of the MST included several prototype applications developed during the research phase. These are documented in the Appendix under [Applications \(Legacy\)](#) for reference.

17 Apps User Guide

Step-by-step walkthrough of the Scores and Species applications

This guide walks through the two interactive applications — **Scores** (app.marinesensitivity.org/scores) and **Species** (app.marinesensitivity.org/species) — one control at a time. Each numbered step below mirrors a step in that app’s built-in **Take a Tour** (click *About* → *Take a Tour* in either app to walk the same path live), with the relevant control highlighted on the screenshot.

For a feature-by-feature reference instead, see the [Scores](#) and [Species](#) appendix pages.



How this guide stays in sync

The steps, captions, and highlighted screenshots are generated directly from the apps’ in-app tours (the `Conductor$new()$step(...)` chains in each `app.R`). Running `libs/apps-guide-regenerate.R` re-parses both tours, re-highlights the same CSS element each tour points at, and re-captures the screenshots, writing `figures/apps-guide/steps.json`. So whenever a tour changes, re-running that one script refreshes this whole page.

17.1 Scores app

17.1.1 1. Study Area

Pick a region to focus on. ‘Full study area’ shows all US federal waters from the Aleutians to the Caribbean and out to the Pacific Island Territories. The other choices zoom into a single subregion.

17.1.2 2. Spatial Units

Toggle between fine-grained raster cells (0.05°) and aggregated BOEM Program Area polygons. Cell mode lets you click any pixel; Program Area mode shows pre-aggregated zone scores.

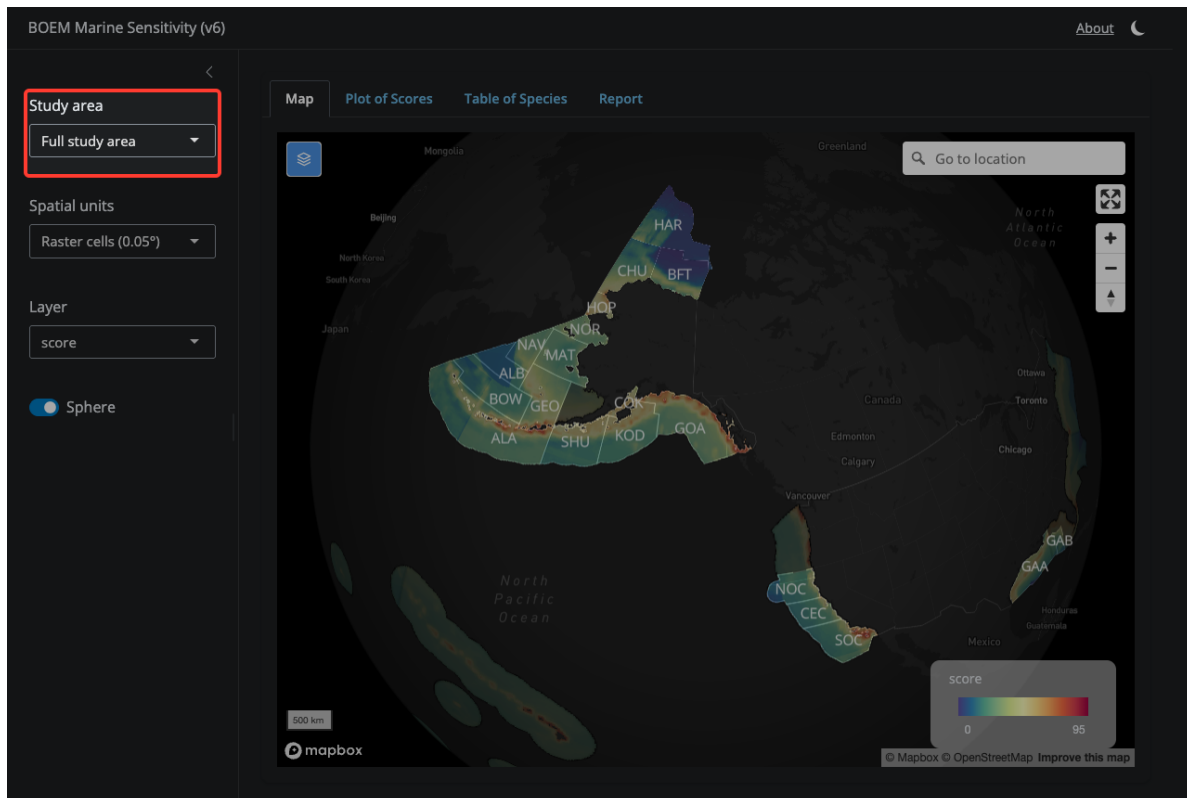


Figure 17.1: Study Area

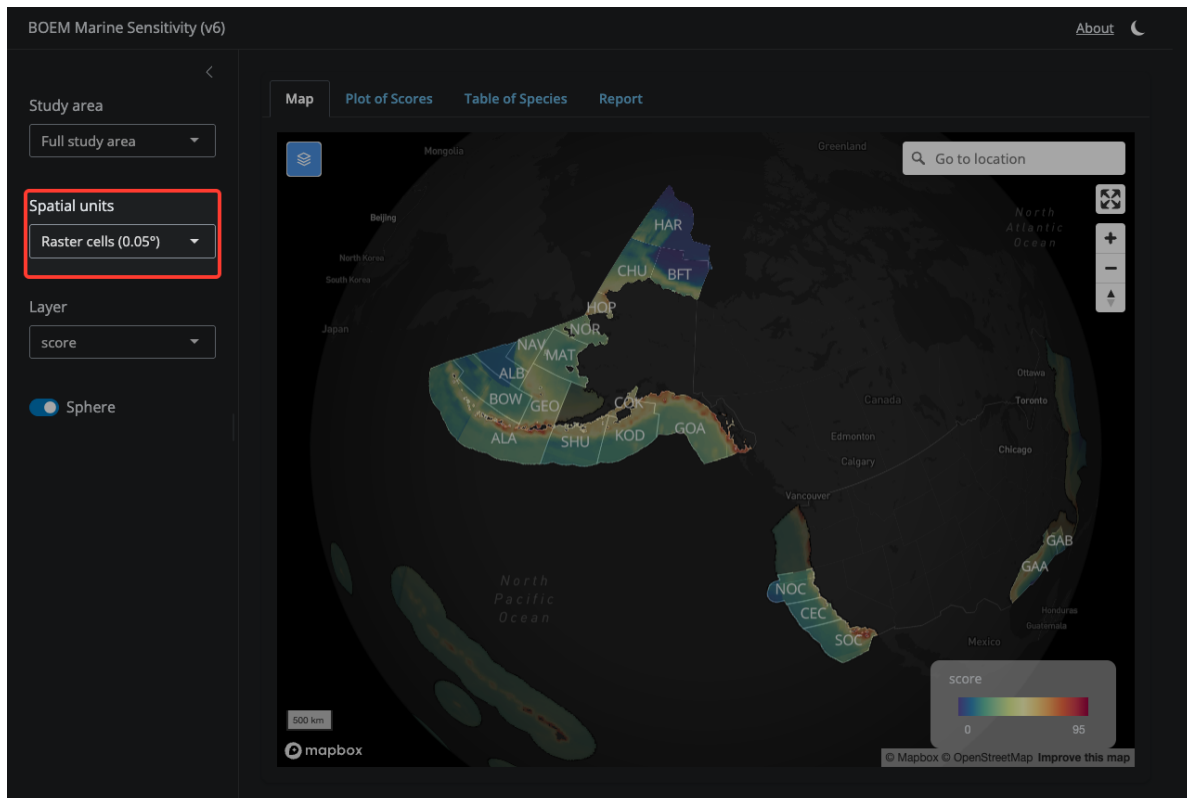


Figure 17.2: Spatial Units

17.1.3 3. Layer Selection

Choose which sensitivity metric to display — composite score, individual species categories (bird, fish, mammal, etc.), or primary productivity. Note that some cells (Atlantic, Gulf of America, Hawaii, Puerto Rico, Pacific Islands) have scores but lie outside any v6 BOEM Program Area; they appear dimmed under a ‘Cells outside Program Areas’ overlay.

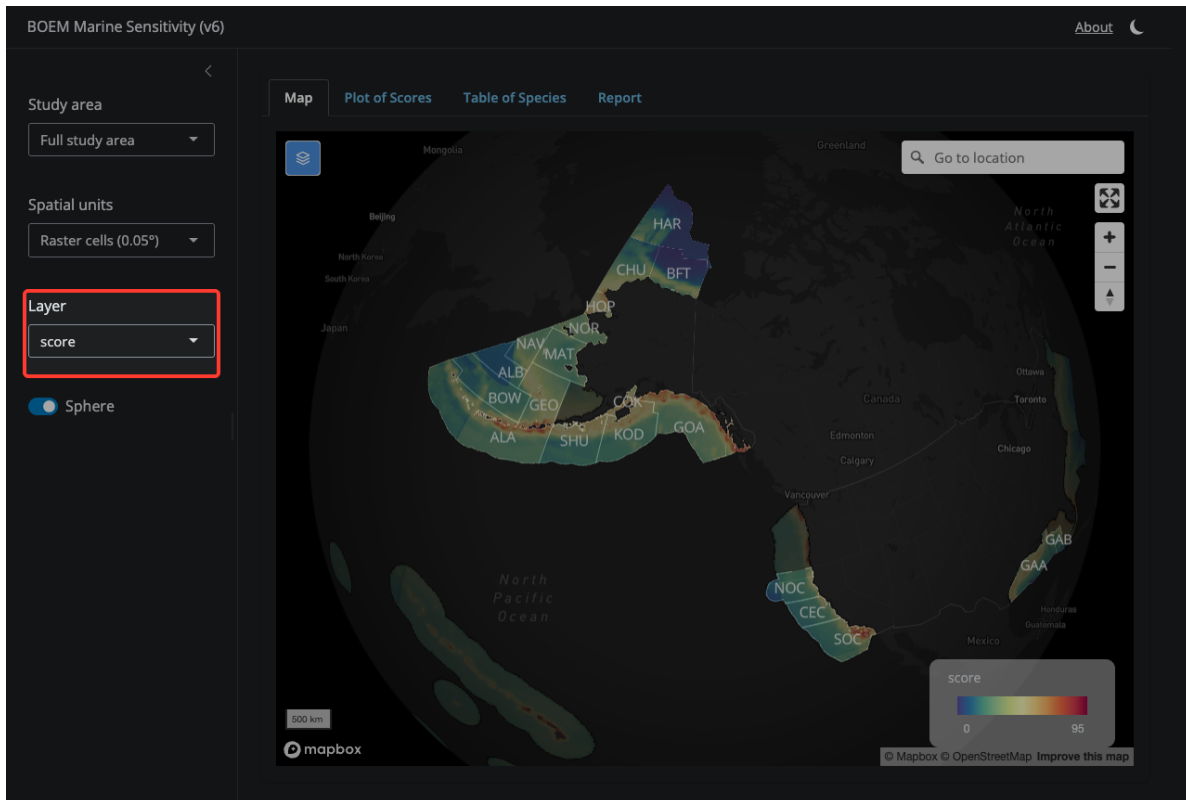


Figure 17.3: Layer Selection

17.1.4 4. Map tab

The Map tab is where you explore scores spatially — click cells or click Program Areas to see their scores.

17.1.5 5. Layers control

Toggle individual map layers on and off — program area outlines, ecoregion outlines, raster cells, and the gray ‘Cells outside Program Areas’ overlay.

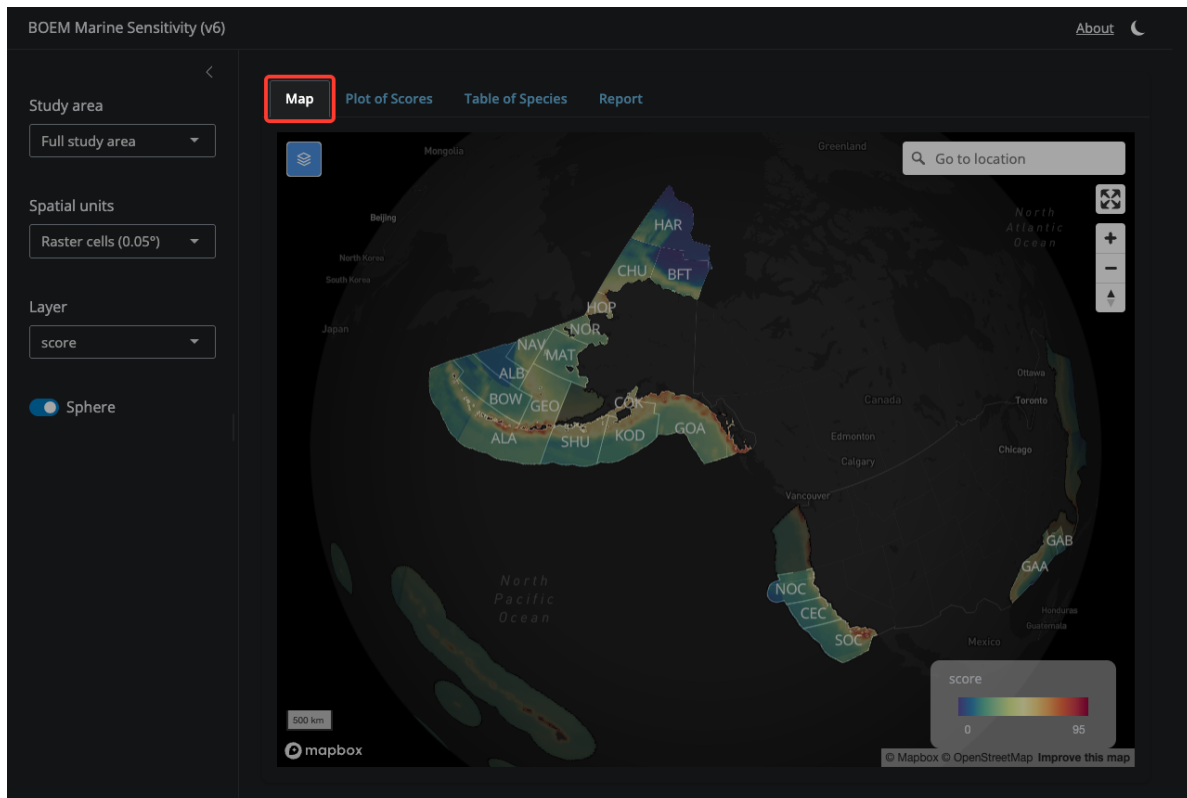


Figure 17.4: Map tab

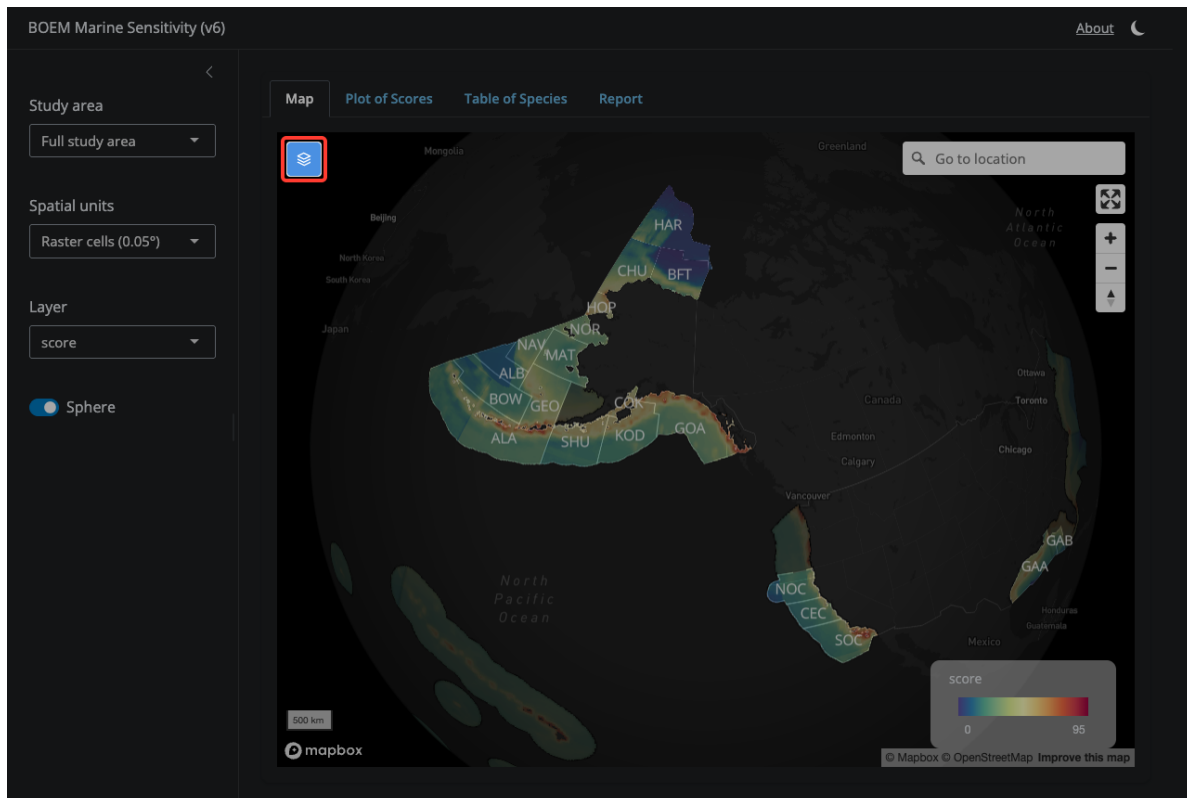


Figure 17.5: Layers control

17.1.6 6. Go to location

Search for a place name and the map will fly there — useful for jumping to a specific Program Area, port, or feature.

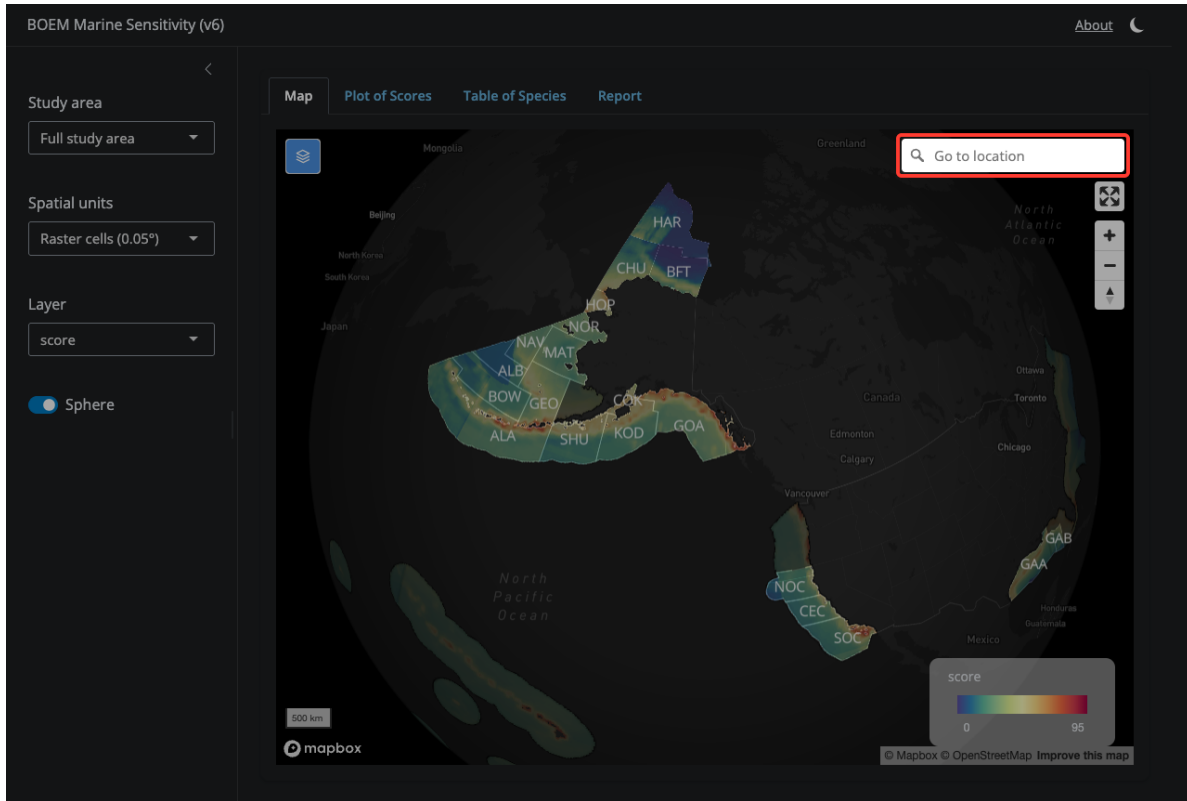


Figure 17.6: Go to location

17.1.7 7. Full screen

Expand the map to fill the entire window.

17.1.8 8. Zoom in / out

Zoom and reset the view. You can also use the mouse wheel, pinch gesture, or double-click.

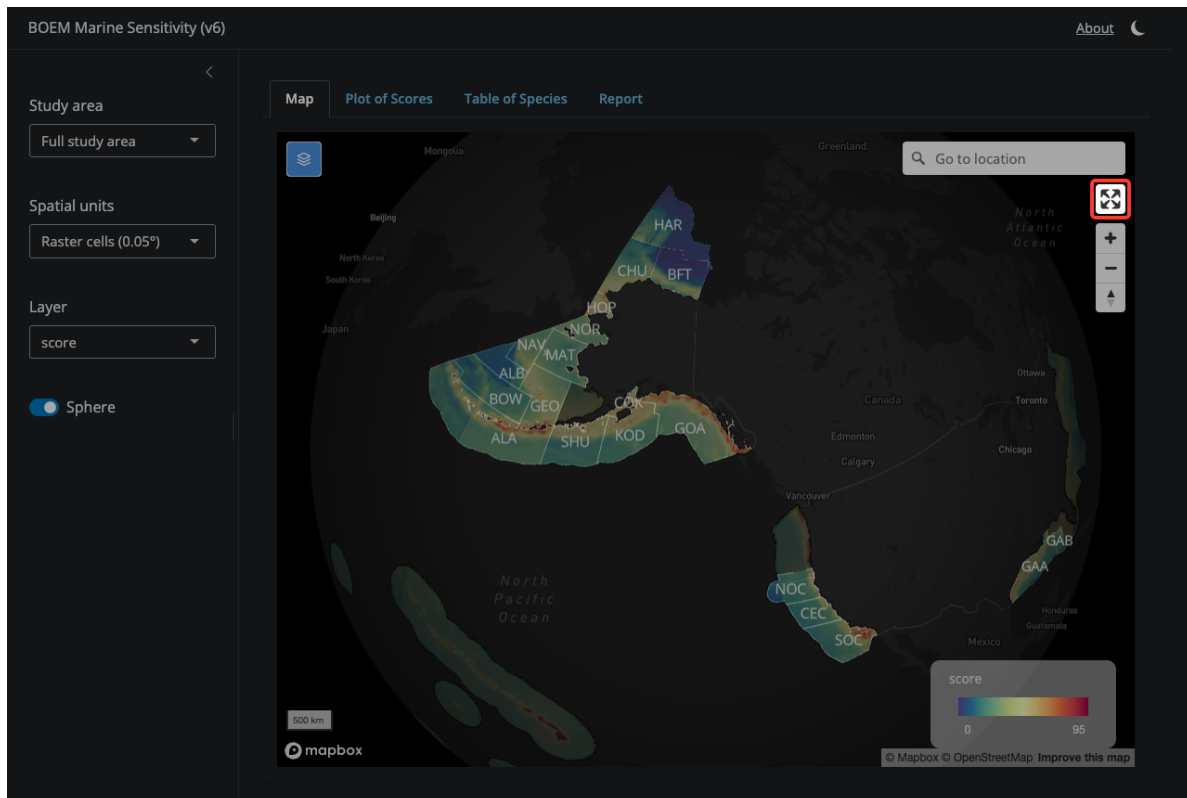


Figure 17.7: Full screen

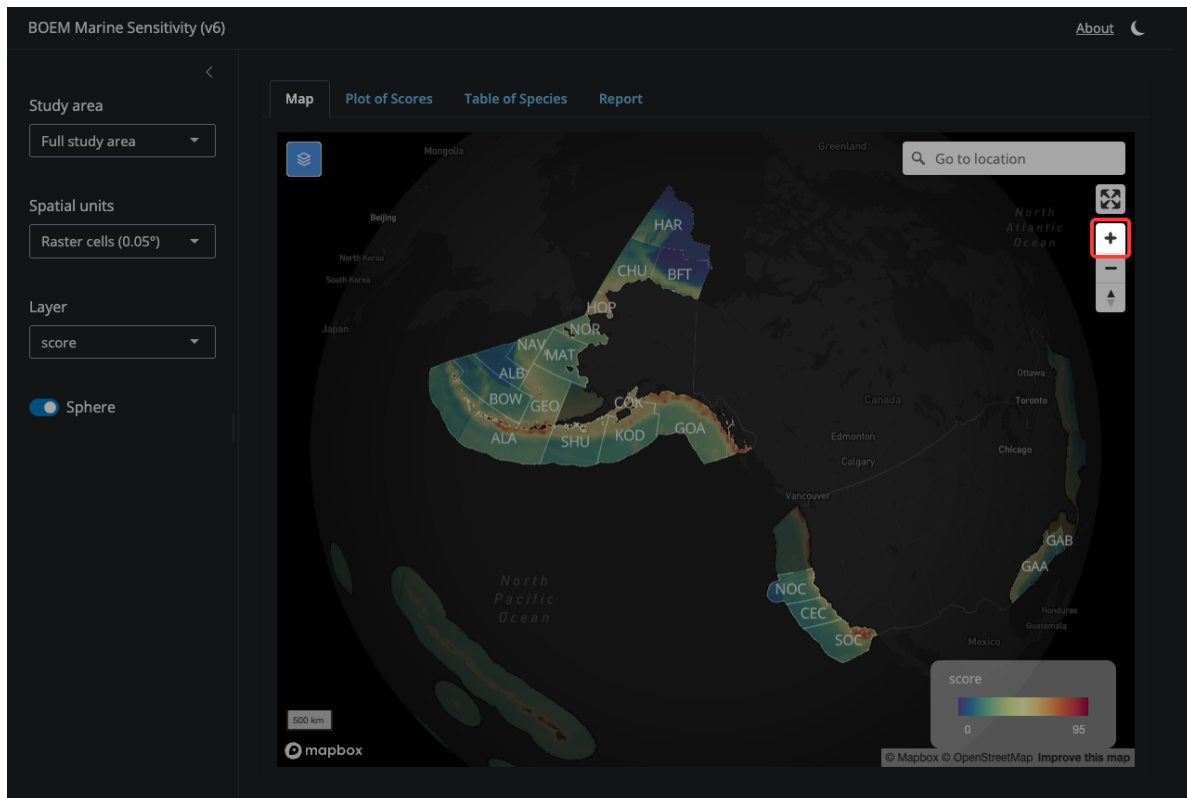


Figure 17.8: Zoom in / out

17.1.9 9. Plot of Scores tab

Switch here to see the flower plot of aggregated sensitivity scores for the current selection, broken out by species category. Petal length = score (0–100); center number = weighted mean. Defaults to ‘All USA’ until you click a cell or click a Program Area.

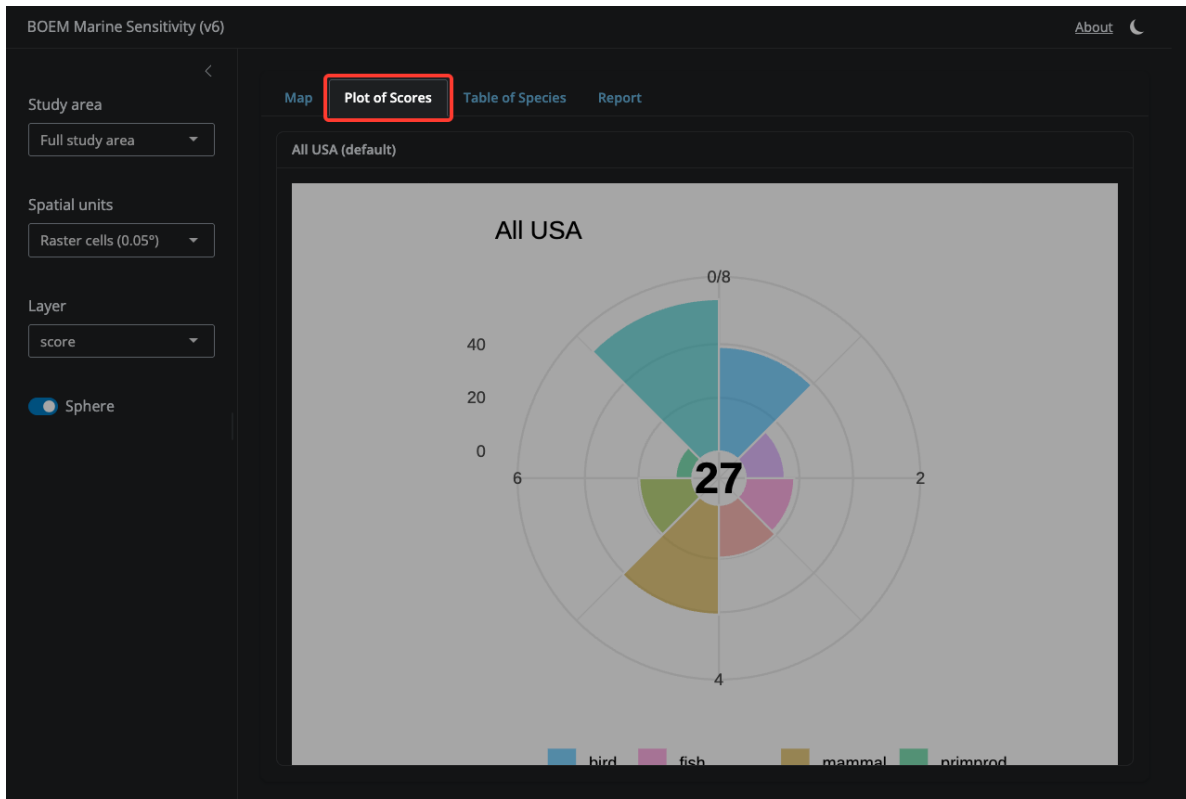


Figure 17.9: Plot of Scores tab

17.1.10 10. Table of Species tab

A sortable, downloadable table of every species in the currently selected area, with extinction-risk codes, areas, and per-category contributions.

17.1.11 11. Report tab

Generate a sensitivity report for one or more custom areas. Build up a list of labeled areas by drawing a polygon on the map (using the polygon tool) and/or by selecting a Program Area,

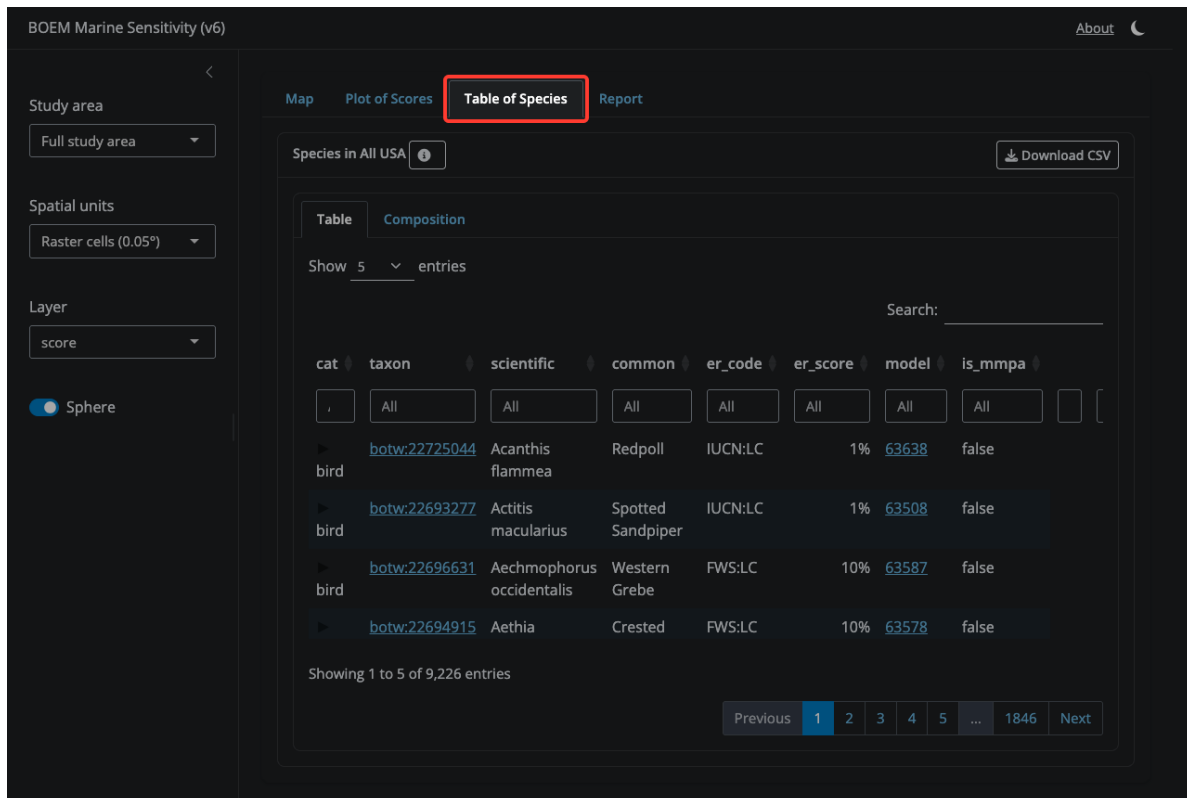


Figure 17.10: Table of Species tab

then clicking ‘Add’ for each. Set a title, data version, and output format (HTML, Word, or PDF), then click ‘Generate report’.

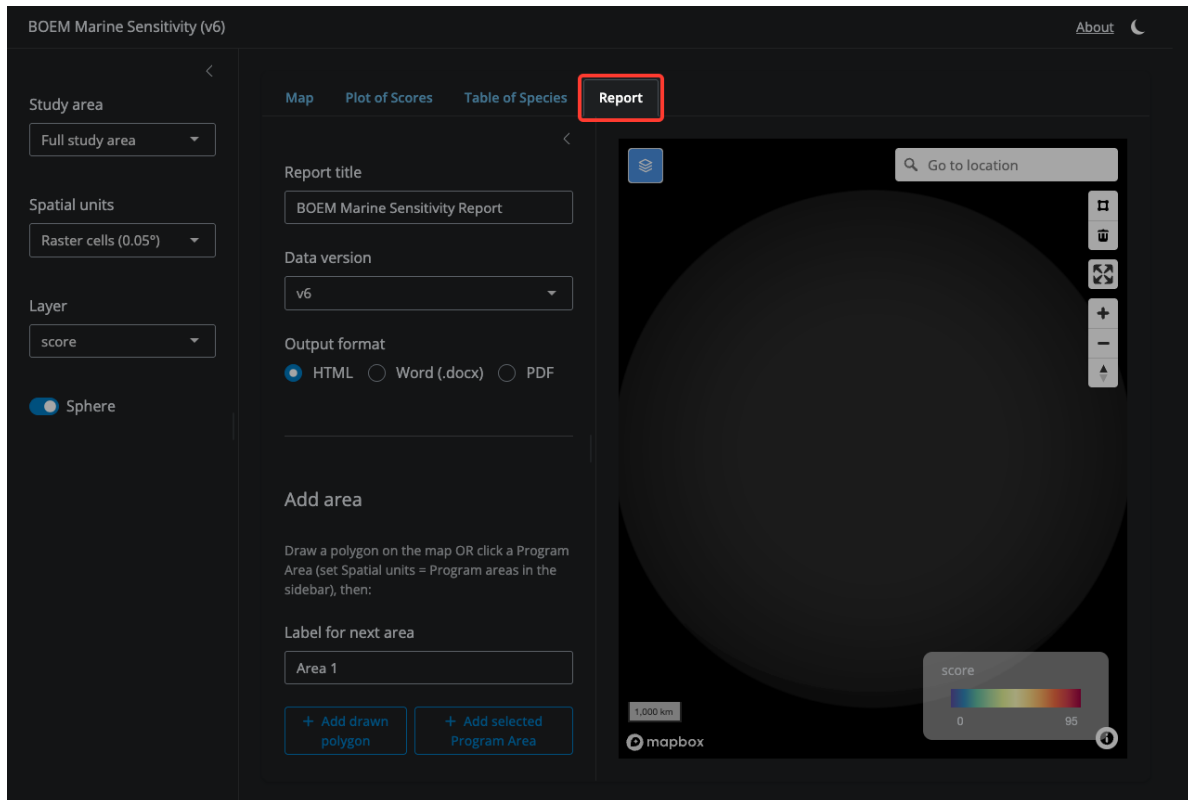


Figure 17.11: Report tab

17.2 Species app

17.2.1 1. Select a Species

Search by common or scientific name. Results are grouped by category (bird, fish, mammal, etc.).

17.2.2 2. Mask Selection

Choose whether to overlay BOEM Program Area or Ecoregion boundaries on the map.

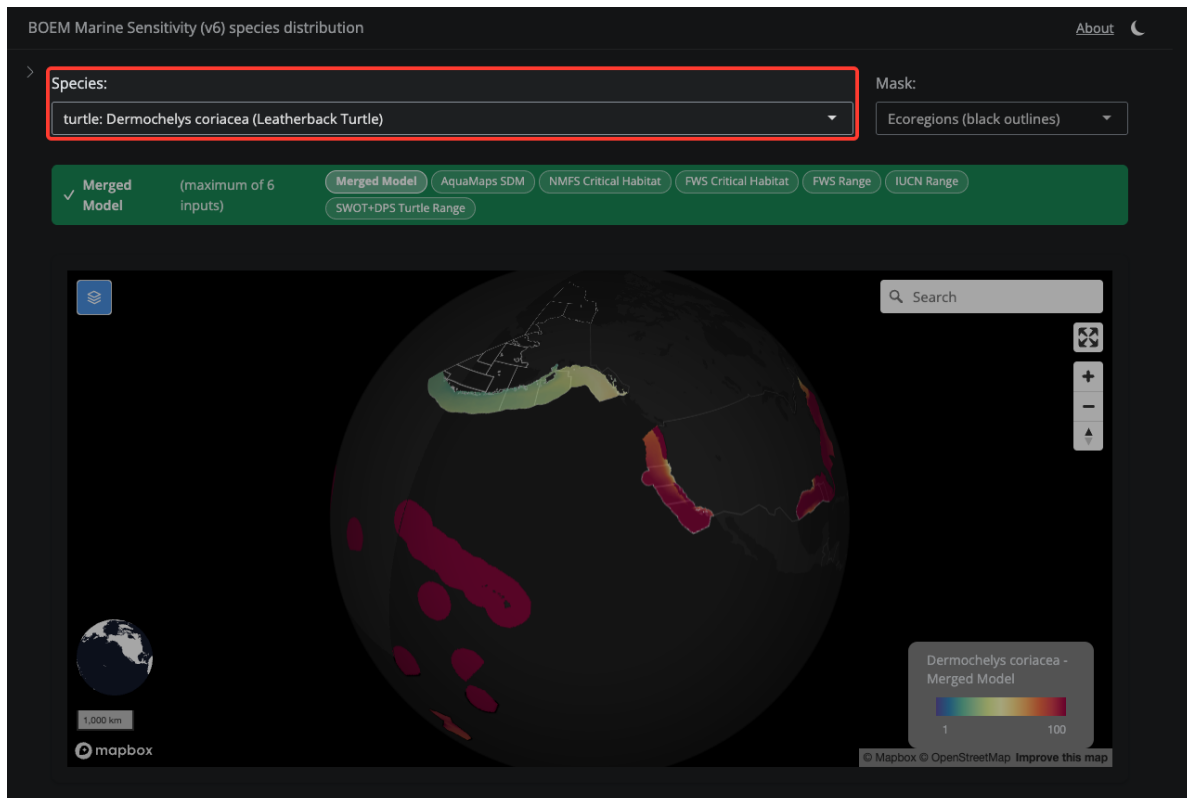


Figure 17.12: Select a Species

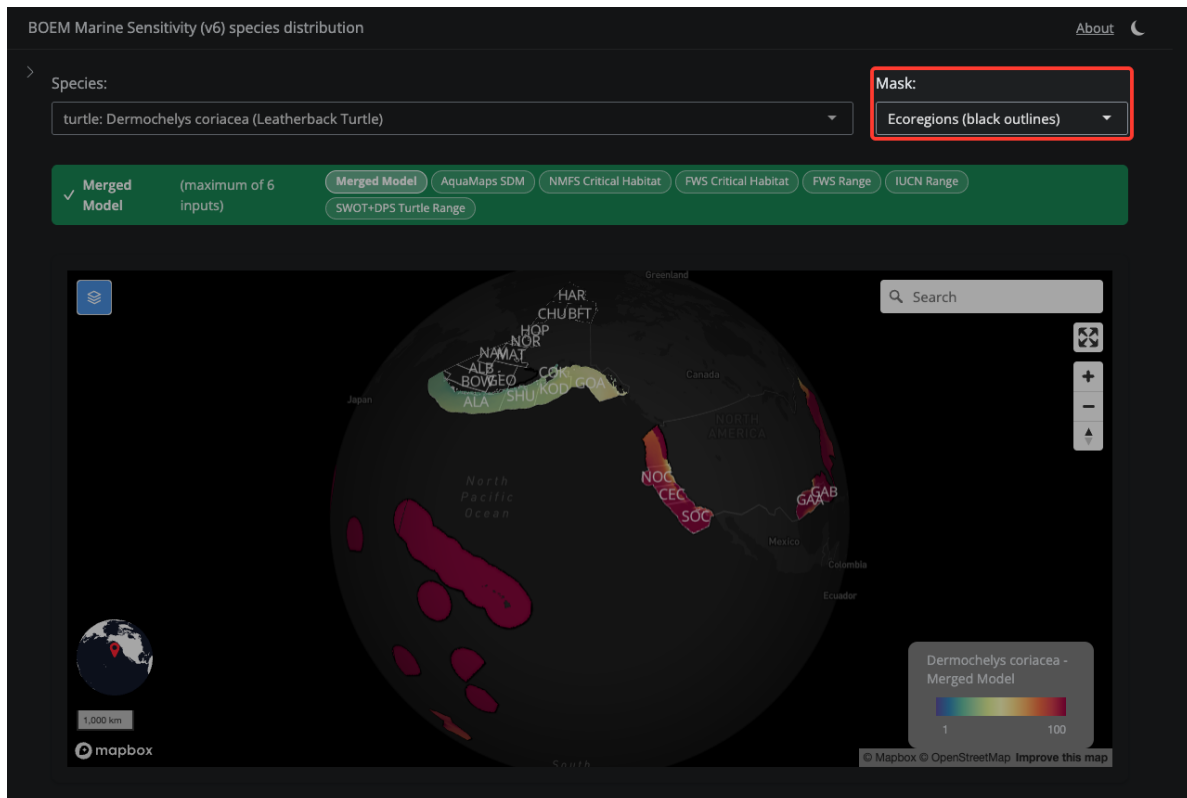


Figure 17.13: Mask Selection

17.2.3 3. Layer Selector

The colored bar shows which model layer is displayed. Green means the final Merged Model; orange means you are viewing a single input. Click the pills to switch layers.

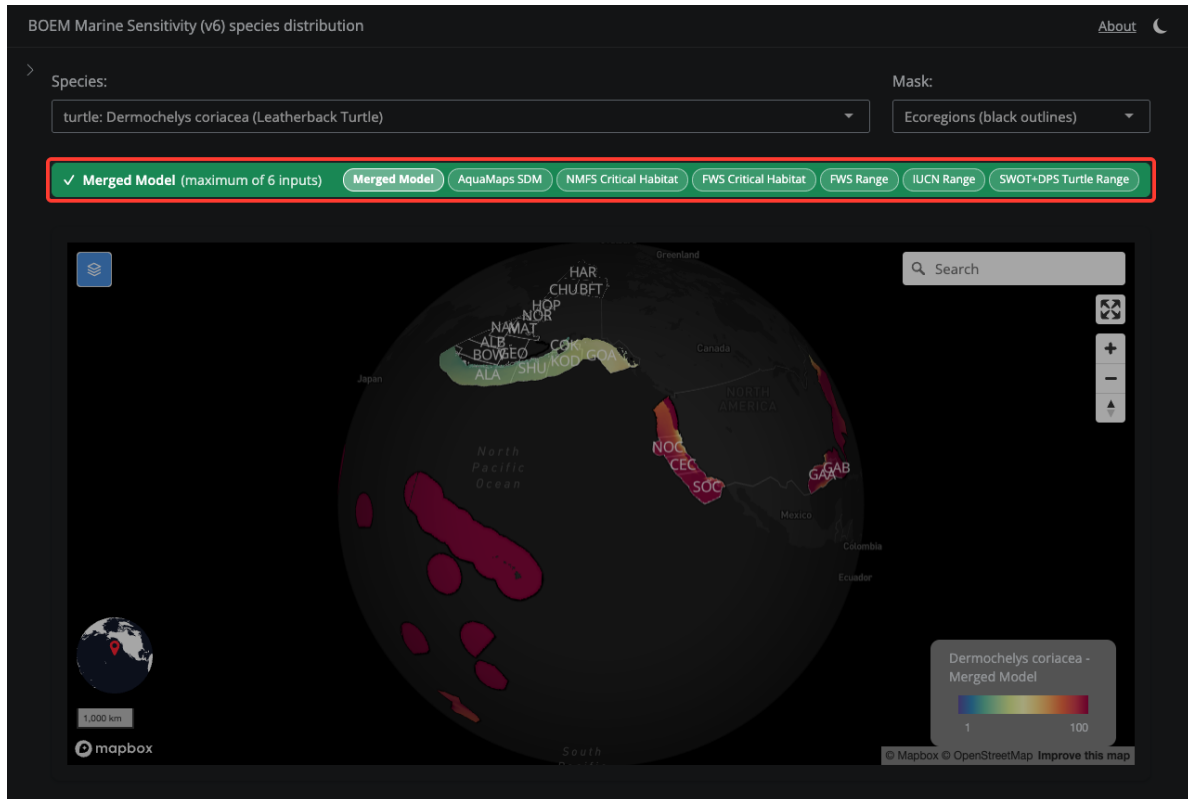


Figure 17.14: Layer Selector

17.2.4 4. Species Map

The map shows the species distribution model. Cell values range from 1 (low sensitivity) to 100 (high sensitivity). Click cells for details.

17.2.5 5. Species Info

Open the sidebar to see ESA listing, IUCN status, MMPA/MBTA flags, and extinction risk score for the selected species.

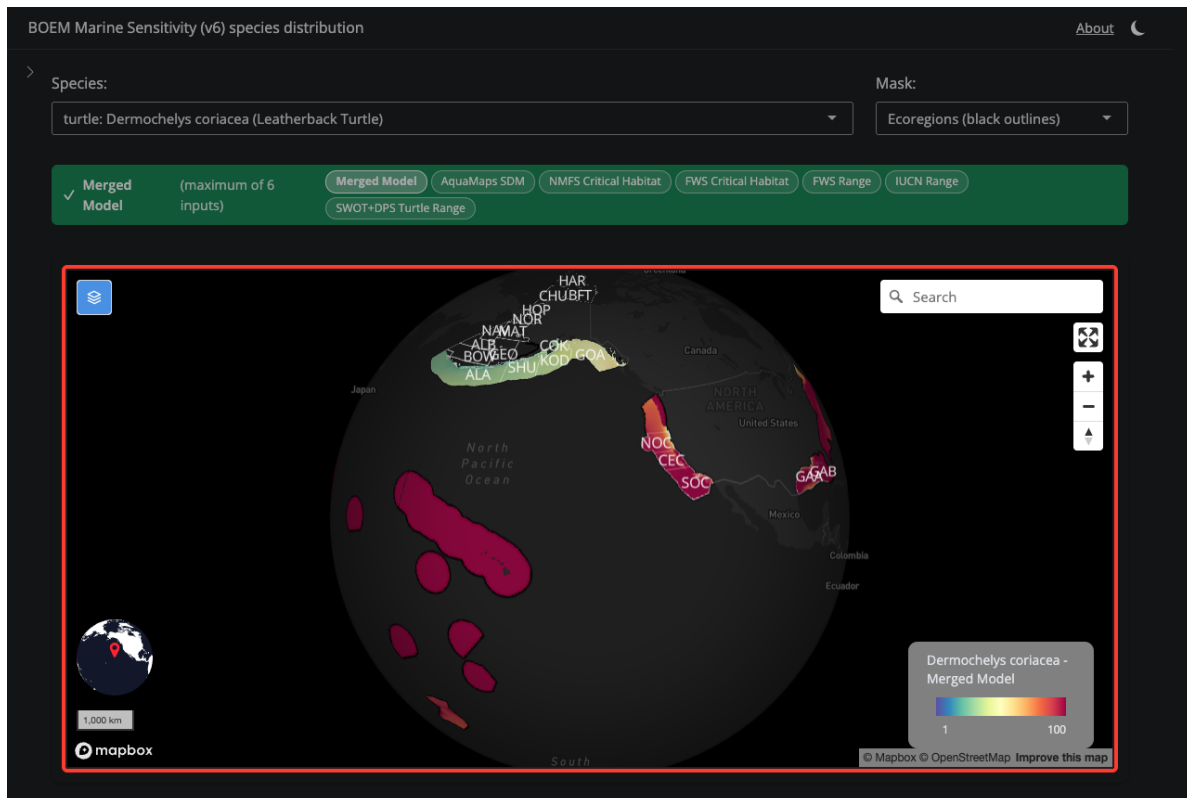


Figure 17.15: Species Map

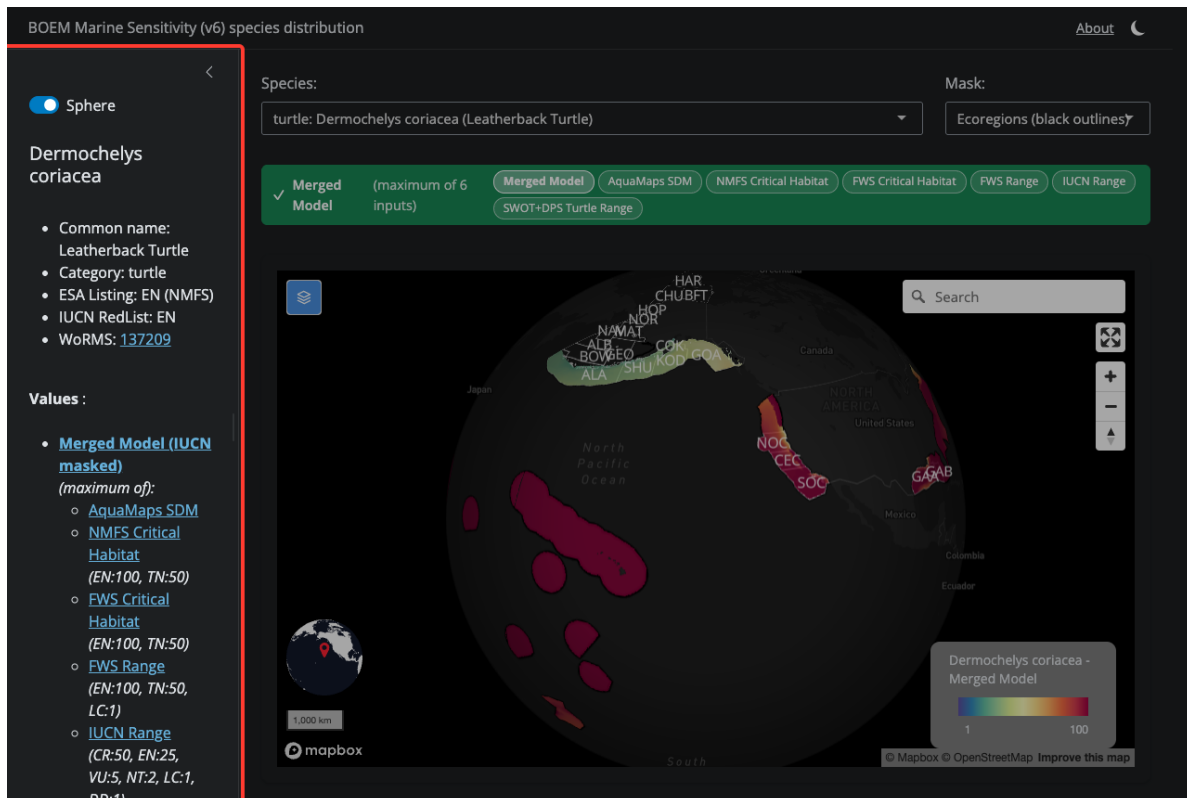


Figure 17.16: Species Info

18 Docs

Technical documentation is principally in this book:

- marinesensitivity.org/docs
the main documentation site
- github.com/MarineSensitivity/docs
source code in the Github repository

But there are also some other self-documenting resources:

- marinesensitivity.org/msens
documented R functions

19 Summary

The Marine Sensitivity Toolkit (MST) provides a comprehensive, reproducible system for assessing the sensitivity of marine ecosystems to offshore energy development across US waters.

19.1 Key Findings

- **16,153 valid marine species** mapped at 0.05° resolution (~4 km cells) across the full US EEZ study area (**9,230** within the **20 BOEM Program Areas**), integrating data from 7 source datasets including AquaMaps SDMs, NMFS and FWS regulatory designations, BirdLife seabird ranges, and IUCN range maps
- **Extinction risk scoring** (1–100 scale) incorporates protections under the Endangered Species Act (ESA), Marine Mammal Protection Act (MMPA), and Migratory Bird Treaty Act (MBTA), ensuring that federally protected species are properly weighted in the sensitivity analysis
- **Model merging** combines multiple distribution datasets per species using a MAX-value approach, constrained by spatial masks from IUCN/Critical Habitat ranges and enhanced with MMPA (floor=20) and MBTA (floor=10) spatial minimums
- **Ecoregional rescaling** normalizes sensitivity scores to a [0–100%] range within each BOEM ecoregion, enabling meaningful cross-region comparison despite natural differences in species richness between tropical and polar waters
- **Interactive tools** (Scores and Species applications) provide decision makers and scientists with accessible interfaces for exploring sensitivity patterns, individual species distributions, and score breakdowns at multiple spatial scales

19.2 Future Directions

Several enhancements are planned for future phases of the MST:

- **OBIS/GBIF validation:** refine species distributions using occurrence data from the Ocean Biodiversity Information System and Global Biodiversity Information Facility with temporal filters

- **H3 hexagonal grid:** explore Uber’s H3 hierarchical hexagonal indexing system for multi-resolution spatial analysis, replacing the current latitude-longitude grid
- **Cloud-Optimized GeoTIFFs (COGs) and STAC catalogs:** export raster data as COGs with SpatioTemporal Asset Catalog metadata for interoperability with geospatial platforms
- **DAG pipeline:** implement a formal directed acyclic graph (DAG) workflow using the `targets` R package for automated, reproducible data processing
- **Stressor-receptor matrices:** expand from sensitivity-only assessment to full vulnerability analysis by integrating exposure data for offshore wind, oil & gas, and other marine activities (see Chapter 9)
- **Enhanced applications:** arbitrary area selection tools, taxonomic tree browsing, automatic high-quality static figure generation, and story-map/scrollytelling explainer interfaces
- **Infrastructure transfer:** support transition of data and applications to BOEM server infrastructure for long-term operational deployment

References

- Behrenfeld, Michael J., and Paul G. Falkowski. 1997. "Photosynthetic Rates Derived from Satellite-Based Chlorophyll Concentration." *Limnology and Oceanography* 42 (1): 1–20. <https://doi.org/10.4319/lo.1997.42.1.0001>.
- BirdLife International. 2024. *BirdLife International and Handbook of the Birds of the World*. <https://datazone.birdlife.org/species/requestdis>.
- Chang, Winston, Joe Cheng, JJ Allaire, et al. 2024. *Shiny: Web Application Framework for r*. <https://CRAN.R-project.org/package=shiny>.
- Goodale, M. Wing, and Anita Milman. 2016. "Cumulative Adverse Effects of Offshore Wind Energy Development on Wildlife." *Journal of Environmental Planning and Management* 0 (0): 1–21. <https://doi.org/10.1080/09640568.2014.973483>.
- Halpern, Benjamin S., Catherine Longo, Darren Hardy, et al. 2012. "An Index to Assess the Health and Benefits of the Global Ocean." *Nature*, ahead of print. <https://doi.org/10.1038/nature11397>.
- IUCN. 2025. *Digital Distribution Maps on the IUCN Red List of Threatened Species*. <https://www.iucnredlist.org/resources/spatial-data-download>.
- Kaschner, K., K. Kesner-Reyes, C. Garilao, et al. 2019. *AquaMaps: Predicted Range Maps for Aquatic Species*. <https://www.aquamaps.org>.
- Lowndes, Julia S. Stewart, Benjamin D. Best, Courtney Scarborough, et al. 2017. "Our Path to Better Science in Less Time Using Open Data Science Tools." *Nature Ecology & Evolution* 1 (6): 0160. <https://doi.org/10.1038/s41559-017-0160>.
- Maitner, Brian, Paul Efren Santos Andrade, Luna Lei, et al. 2024. "Code Sharing in Ecology and Evolution Increases Citation Rates but Remains Uncommon." *Ecology and Evolution* 14 (8): e70030. <https://doi.org/10.1002/ece3.70030>.
- Morandi, A., S. Berkman, J. Rowe, et al. 2018. *Environmental Sensitivity and Associated Risk to Habitats and Species on the Pacific West Coast and Hawaii with Offshore Floating Wind Technologies*. OCS Study BOEM 2018-031. US Depart-

- ment of the Interior, Bureau of Ocean Energy Management, Pacific OCS Region. <https://www.boem.gov/sites/default/files/environmental-stewardship/Environmental-Studies/Pacific-Region/Studies/BOEM-2018-031-Vol1.pdf>.
- National Marine Fisheries Service. 2019. *Rice's Whale Core Distribution Area Map GIS Data*. <https://www.fisheries.noaa.gov/resource/map/rices-whale-core-distribution-area-map-gis-data>.
- National Marine Fisheries Service. 2025. *National ESA Critical Habitat Geodatabase*. <https://www.fisheries.noaa.gov/resource/map/national-esa-critical-habitat-mapper>.
- NOAA Fisheries. 2025. *Endangered Species Act Species Range Geodatabase*. National Oceanic; Atmospheric Administration (NOAA), National Marine Fisheries Service (NMFS), Office of Protected Resources (OPR). <https://noaa.maps.arcgis.com/home/search.html?searchTerm=turtle+dps>.
- Ross, Pauline M., Elliot Scanes, Maria Byrne, et al. 2023. *Surviving the Anthropocene: The Resilience of Marine Animals to Climate Change*. CRC Press.
- U.S. Fish and Wildlife Service. 2025a. *USFWS Complete Current Range Maps*. <https://ecos.fws.gov/ecp/report/current-range-all>.
- U.S. Fish and Wildlife Service. 2025b. *USFWS Threatened and Endangered Species Active Critical Habitat Report*. <https://ecos.fws.gov/ecp/report/table/critical-habitat.html>.
- Wallace, Bryan P., Zachary A. Posnik, Bryan J. Hurley, et al. 2023. "Marine Turtle Regional Management Units 2.0." *Endangered Species Research* 52: 209–23. <https://doi.org/10.3354/esr01243>.
- Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, et al. 2016. "The FAIR Guiding Principles for Scientific Data Management and Stewardship." *Scientific Data* 3 (1): 160018. <https://doi.org/10.1038/sdata.2016.18>.

Glossary

acclimatisation the adjustment of an organism to environmental conditions in the field or environment rather than the laboratory without an adjustment in their genetics. Acclimatisation has been used to describe phenotypically plastic responses in natural conditions. Source: Ross et al. (2023).

adaptation the evolutionary mechanism where natural selection of traits is genetically passed on, typically over many generations, to create an organism suited to the environment. Source: [rossRoss et al. (2023)

adaptive capacity the capacity of the ecosystem or organism to improve and reorganise in response to stress such as climate change through phenotypic plasticity (acclimation, acclimatisation) or adaptation, distributional shifts, and rapid evolution of traits suited to new conditions. Source: Ross et al. (2023).

BOEM Bureau of Ocean Energy Management — US federal agency responsible for managing offshore energy and mineral resources on the Outer Continental Shelf.

COG Cloud-Optimized GeoTIFF — a raster file format optimized for efficient access over HTTP, enabling range requests for partial reads.

DuckDB an embedded analytical database engine used for storing and querying species distribution models, extinction risk scores, and sensitivity metrics.

ecoregion a BOEM-defined ecological region used as the spatial unit for rescaling sensitivity scores to enable meaningful cross-region comparison.

epigenetics the modification of phenotype plasticity of an organism through altered gene expression without an alteration to the DNA sequence. ‘Epi’ means above the DNA and includes DNA methylation, modification of histones, and non-coding RNA. Source: Ross et al. (2023).

ESA Endangered Species Act (1973) — US federal law providing protections for species listed as Endangered or Threatened by NMFS or FWS.

exposure the magnitude of the change in the environment

fecundity the maximum physiological potential reproductive output of an organism to produce offspring (reproductive output). This differs from fertility, which is the number of offspring born. Source: Ross et al. (2023)

MBON Marine Biodiversity Observation Network; see [MarineBON.org](https://marinebon.org)

MBTA Migratory Bird Treaty Act (1918) — US federal law protecting migratory birds, including most seabird species. MBTA-protected species receive a minimum spatial floor score of 10.

MMPA Marine Mammal Protection Act (1972) — US federal law protecting all marine mammals in US waters. MMPA-protected species receive a minimum spatial floor score of

20.

MST Marine Sensitivity Toolkit — the cloud-native system developed for BOEM to assess marine ecosystem sensitivity to offshore energy development.

OCSLA Outer Continental Shelf Lands Act — US federal law governing the management of offshore resources, including the mandate to consider environmental sensitivity and marine productivity.

resilience the capacity of an ecosystem, society, or organism to absorb disturbance and reorganise while undergoing change so as to retain essentially the same function, structure, identity, and feedbacks. Resilience reflects the degree to which a complex adaptive system is determined by its capacity to reorganise and adapt in order to avoid being disturbed again. Source: Ross et al. (2023).

SDM Species Distribution Model — a statistical or environmental envelope model predicting the spatial distribution of a species based on environmental variables.

sensitivity the magnitude of response to the change

stressor the stimulus that causes stress to an organism

vulnerability combination of exposure and sensitivity

Part III

Applications (Current)

A Scores — mapper of aggregate sensitivity scores

A.1 Overview

The **Scores** application (app.marinesensitivity.org/scores) is the primary interactive tool for exploring composite marine sensitivity scores across BOEM Program Areas. It enables users to visualize extinction risk weighted species presence at multiple spatial resolutions, with integrated flower plot and treemap summaries.

Figure C.1 shows the global view of cell-level sensitivity scores across all US waters. The sidebar provides controls for study area, spatial units (raster cells or program areas), and layer selection.

Figure C.2 shows the flower plot panel that appears when a program area is selected. Here, the Southern California program area displays a weighted mean score of 29, with petals representing species categories (bird, coral, fish, invertebrate, mammal, other, turtle) and primary productivity.

A.2 Features

A.2.1 Map Display

- **Cell-level view:** 0.05° grid cells colored by sensitivity score for the selected metric
- **Program area view:** aggregated scores by BOEM Program Area, with boundaries overlaid
- **Base maps:** multiple MapBox base layers (satellite, streets, light, dark)
- **Study area selection:** click a program area to focus the view and update summary panels

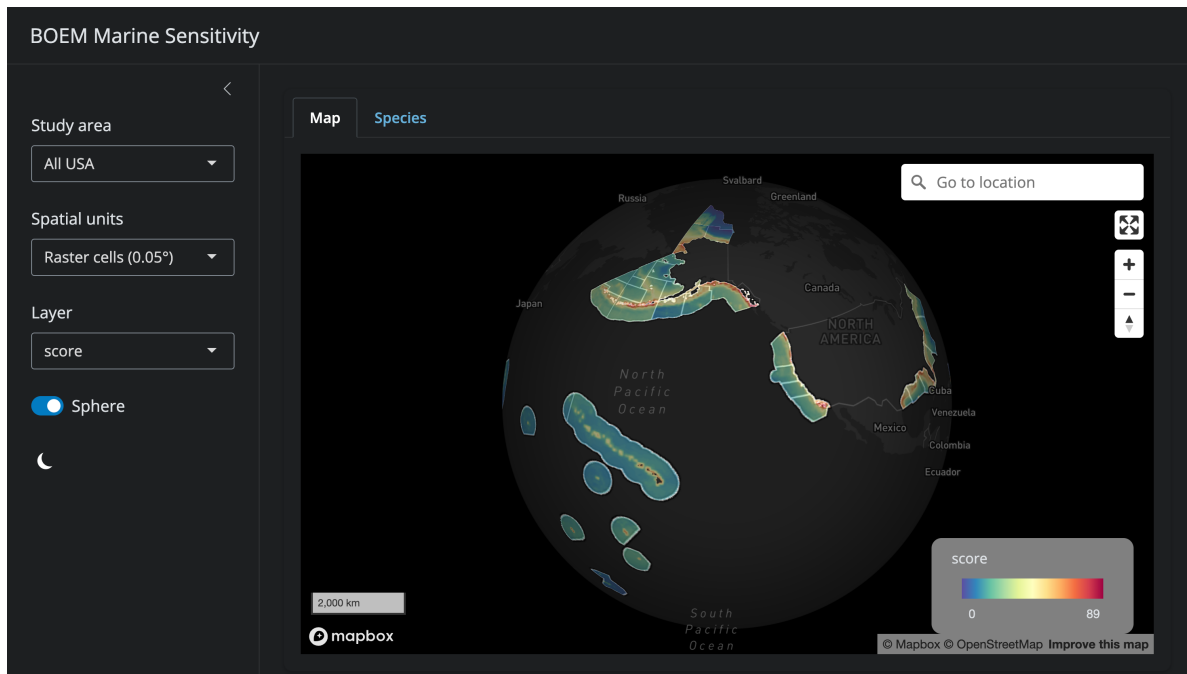


Figure A.1: The Scores application displaying cell-level sensitivity scores across US waters. The sidebar provides controls for study area selection, spatial resolution (raster cells vs. program areas), and layer choice. The globe view shows scores from 0 (blue) to 89 (red) with MapBox basemap.

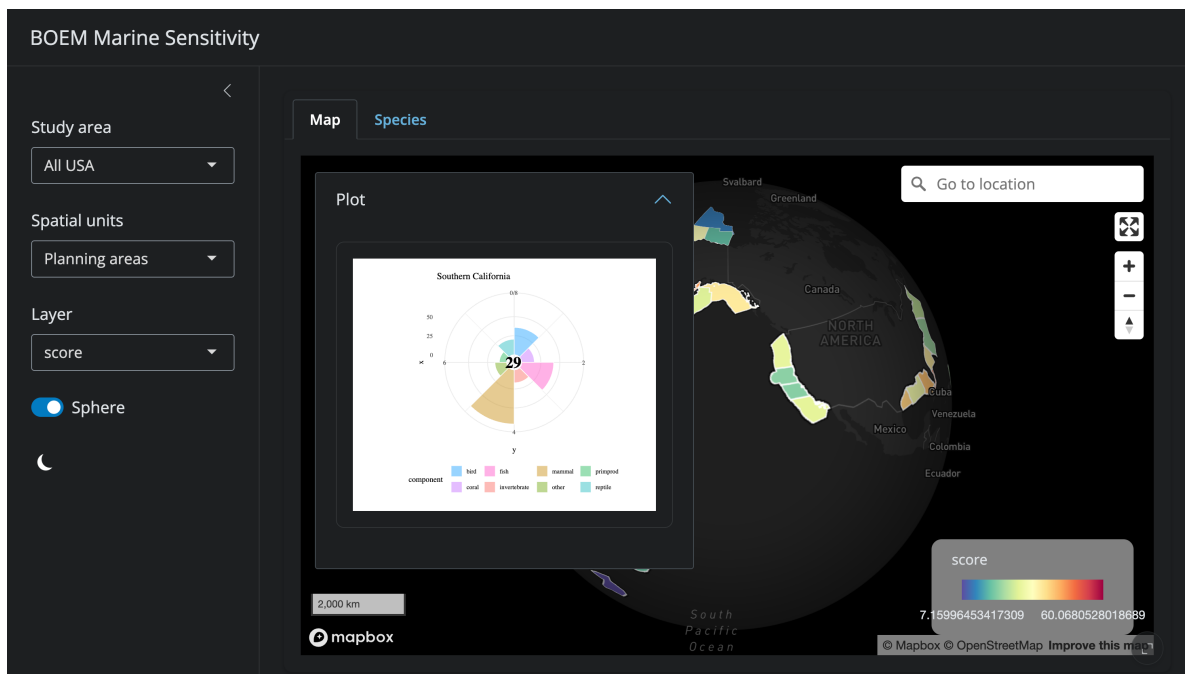


Figure A.2: The Scores application showing the flower plot panel for Southern California. Petal length represents each category's rescaled sensitivity score; petal width represents its weight. The center value (29) is the weighted mean across all categories.

A.2.2 Sensitivity Metrics

The map can display any of the following layers:

- Composite extinction risk (weighted sum across all species categories)
- Individual category scores: bird, coral, fish, invertebrate, mammal, reptile/turtle
- Primary productivity

All scores are ecoregionally rescaled to [0–100] for meaningful cross-region comparison (see Chapter 7).

A.2.3 Flower Plot

Selecting a program area displays a flower plot summarizing the sensitivity profile:

- Each petal represents a species category
- Petal length = rescaled sensitivity score (0–100)
- Petal width = category weight
- Center value = weighted mean across categories

A.2.4 Species Table

A sortable table of all valid species present in the selected area, showing:

- Scientific and common name
- Species category
- ESA listing, IUCN Red List status
- MMPA/MBTA protection flags
- Extinction risk score

A.2.5 Data Export

Users can download CSV files of scores and species lists for the selected program area or custom area of interest.

A.3 Generating Screenshots

A.4 Source Code

The application source code is at: [apps/mapgl/app.R](#)

B Species — mapper of individual species distributions

B.1 Overview

The **Species** application (app.marinesensitivity.org/species) enables exploration of individual species distribution models within the MST database. Users can view merged and source-level models, compare masked vs. unmasked distributions, and access detailed species regulatory status information.

Figure [B.1](#) shows the merged distribution model for the blue whale (*Balaenoptera musculus*). The map displays cell values from 1 (low suitability, blue) to 100 (high suitability, red) across all US program areas where the species occurs.

B.2 Features

B.2.1 Species Selection

- **Search:** find species by common name or scientific name
- **Category filter:** narrow results by species category (bird, coral, fish, invertebrate, mammal, reptile/turtle)
- **Alphabetical browsing:** scroll through species lists within each category

B.2.2 Model Display

- **Merged model:** the final merged distribution showing MAX values across all source datasets, with masking and spatial floors applied
- **Source datasets:** toggle individual source layers (AquaMaps, NMFS CH, FWS CH, FWS Range, BirdLife, IUCN Range) to see each dataset's contribution
- **IUCN masking toggle:** compare the unmasked AquaMaps SDM with the masked merged model to visualize the effect of range constraints

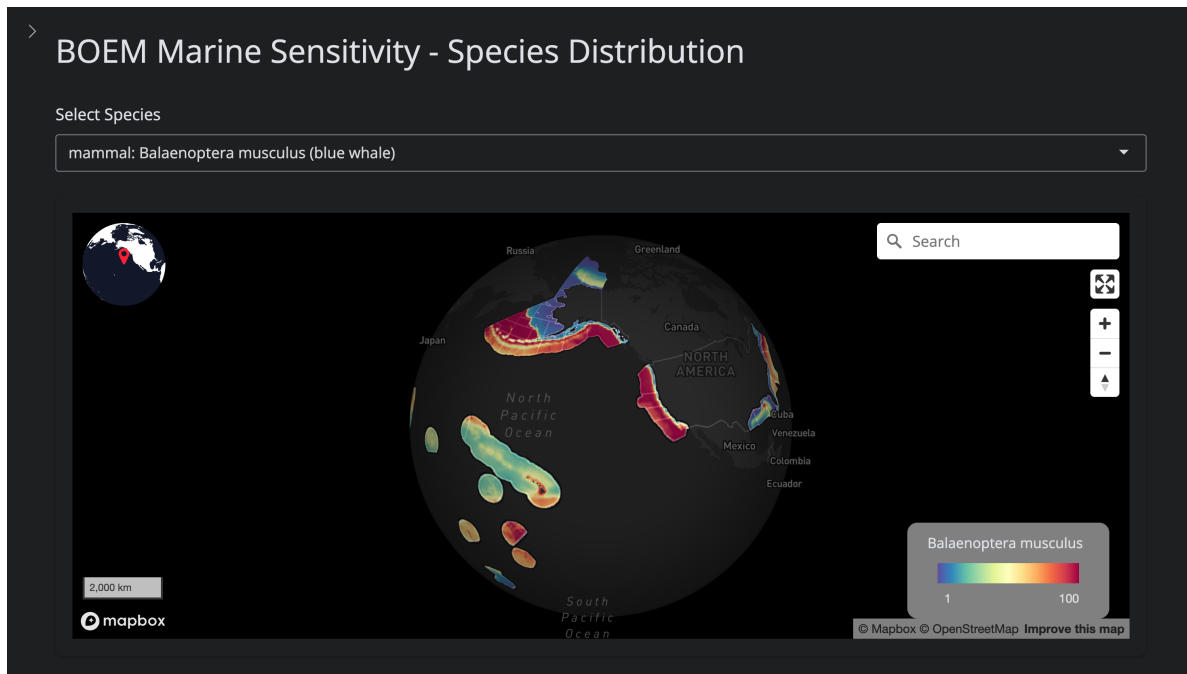


Figure B.1: The Species application showing the merged distribution model for blue whale (*Balaenoptera musculus*). Cell values range from 1 (low suitability) to 100 (high suitability). The species selector dropdown allows searching by common or scientific name with category prefix.

B.2.3 Species Info Panel

Detailed information for the selected species:

- Scientific name and common name
- Species category (bird, fish, mammal, etc.)
- ESA listing status (Endangered, Threatened, or not listed)
- IUCN Red List category
- MMPA protection flag
- MBTA protection flag
- Computed extinction risk score (1–100)
- Number of source datasets contributing to the merged model

B.2.4 Deep-Linking

The application supports URL parameters for direct linking to specific species, enabling bookmarking and sharing of specific views:

`https://shiny.marinesensitivity.org/species/?species=Balaenoptera+musculus`

B.3 Generating Screenshots

B.4 Source Code

The application source code is at: [apps/mapsp/app.R](#)

Part IV

Applications (Legacy)

Areas of Interest

Map high resolution Areas of Interest (using vector tiles) for visualization (and later summarization).

- [website](#)
- [code](#)

Area Explorer

<

Areas of Interest

Regions ▾

last edited: POLYGON
((-141.0639 46.55907,
-136.6324 42.0423,
-132.0404 48.99068,
-135.5994 49.97881,
-141.0639 46.55907))

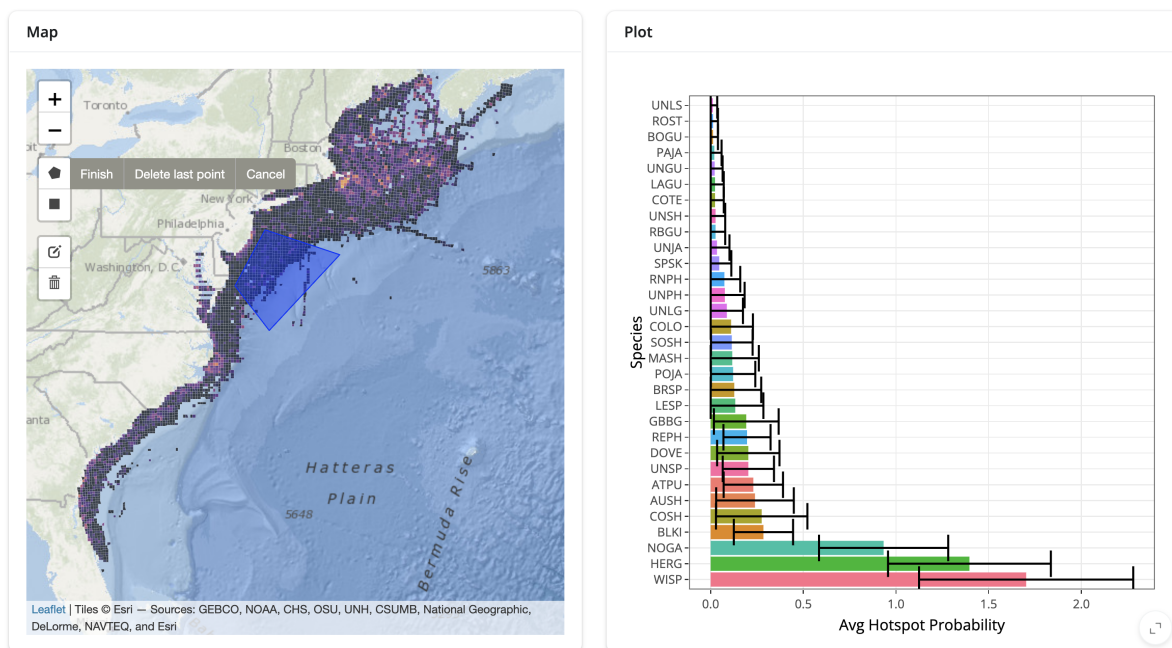
Map

aoi	
shlf_key	AK
shlf_name	Alaska
rgn_key	ALA
rgn_name	Aleutian Arc
ctr_lon	-178.56252307137368
ctr_lat	51.149538291898025
area_km2	860518.1237176519
geometry	null

Bird Hotspots

Bird hotspots application showing hotspot probability for species present given drawn Area of Interest.

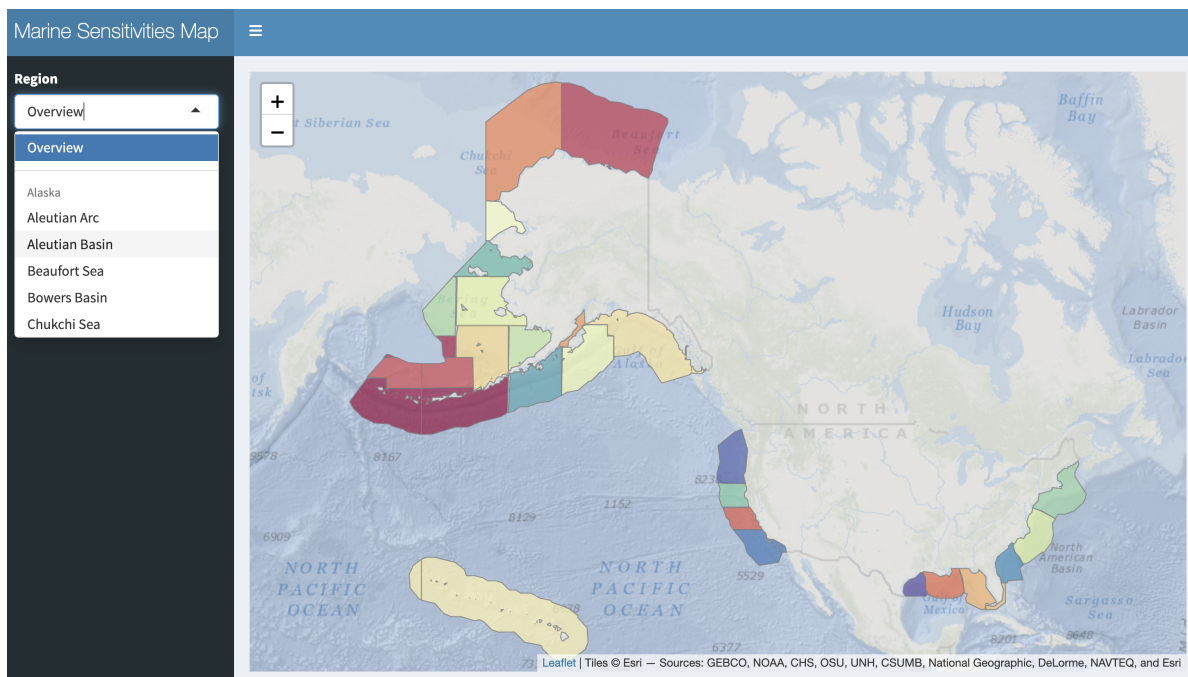
- [website](#)
- [code](#)



Regional Map

Basic interactive map of BOEM regions.

- [website](#)
- [code](#)



C Scores — mapper of aggregate sensitivity scores

C.1 Overview

The **Scores** application (app.marinesensitivity.org/scores) is the primary interactive tool for exploring composite marine sensitivity scores across BOEM Program Areas. It enables users to visualize extinction risk weighted species presence at multiple spatial resolutions, with integrated flower plot and treemap summaries.

Figure C.1 shows the global view of cell-level sensitivity scores across all US waters. The sidebar provides controls for study area, spatial units (raster cells or program areas), and layer selection.

Figure C.2 shows the flower plot panel that appears when a program area is selected. Here, the Southern California program area displays a weighted mean score of 29, with petals representing species categories (bird, coral, fish, invertebrate, mammal, other, turtle) and primary productivity.

C.2 Features

C.2.1 Map Display

- **Cell-level view:** 0.05° grid cells colored by sensitivity score for the selected metric
- **Program area view:** aggregated scores by BOEM Program Area, with boundaries overlaid
- **Base maps:** multiple MapBox base layers (satellite, streets, light, dark)
- **Study area selection:** click a program area to focus the view and update summary panels

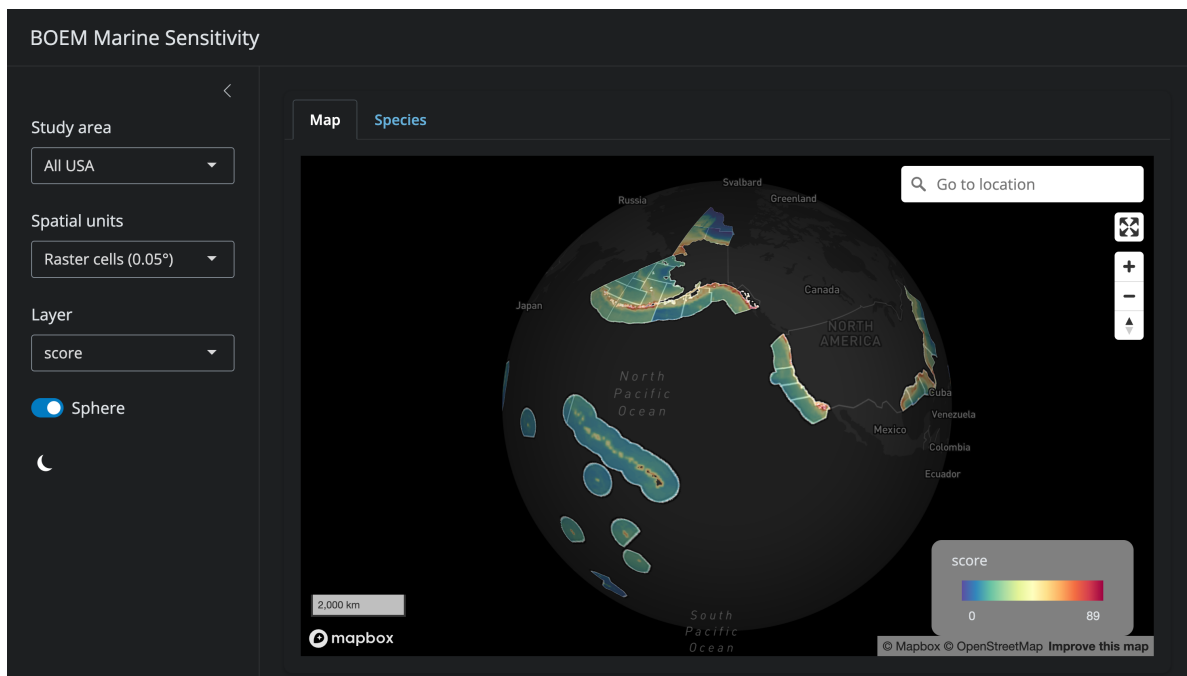


Figure C.1: The Scores application displaying cell-level sensitivity scores across US waters. The sidebar provides controls for study area selection, spatial resolution (raster cells vs. program areas), and layer choice. The globe view shows scores from 0 (blue) to 89 (red) with MapBox basemap.

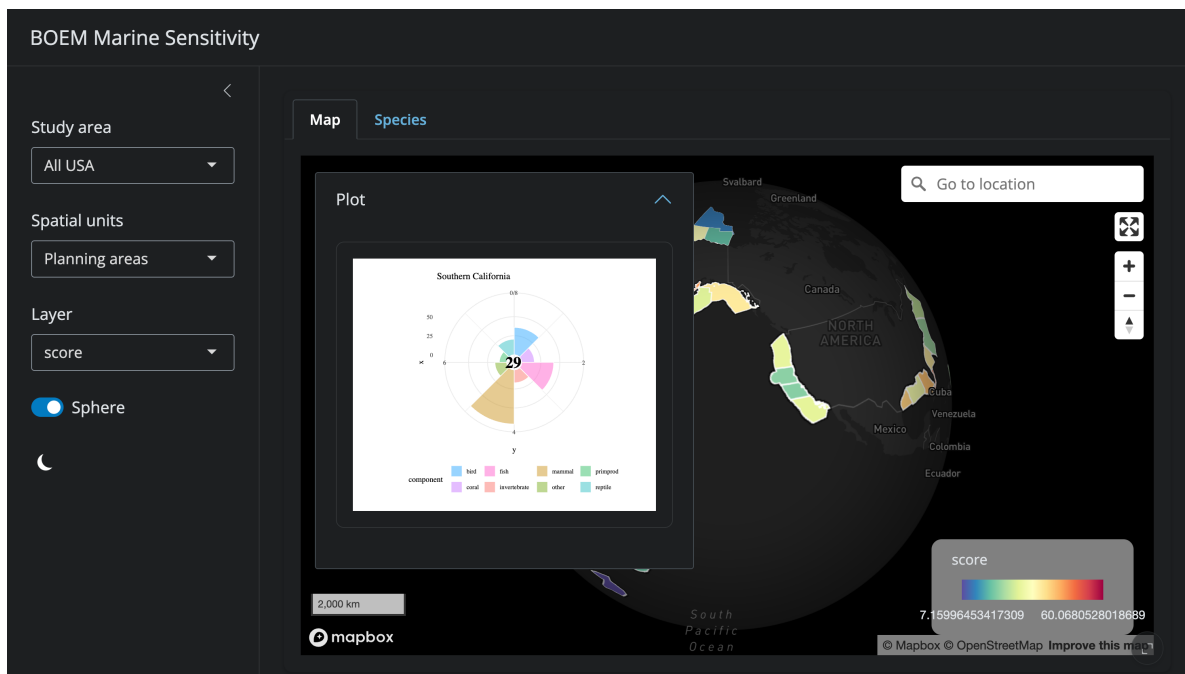


Figure C.2: The Scores application showing the flower plot panel for Southern California. Petal length represents each category's rescaled sensitivity score; petal width represents its weight. The center value (29) is the weighted mean across all categories.

C.2.2 Sensitivity Metrics

The map can display any of the following layers:

- Composite extinction risk (weighted sum across all species categories)
- Individual category scores: bird, coral, fish, invertebrate, mammal, reptile/turtle
- Primary productivity

All scores are ecoregionally rescaled to [0–100] for meaningful cross-region comparison (see Chapter 7).

C.2.3 Flower Plot

Selecting a program area displays a flower plot summarizing the sensitivity profile:

- Each petal represents a species category
- Petal length = rescaled sensitivity score (0–100)
- Petal width = category weight
- Center value = weighted mean across categories

C.2.4 Species Table

A sortable table of all valid species present in the selected area, showing:

- Scientific and common name
- Species category
- ESA listing, IUCN Red List status
- MMPA/MBTA protection flags
- Extinction risk score

C.2.5 Data Export

Users can download CSV files of scores and species lists for the selected program area or custom area of interest.

C.3 Generating Screenshots

C.4 Source Code

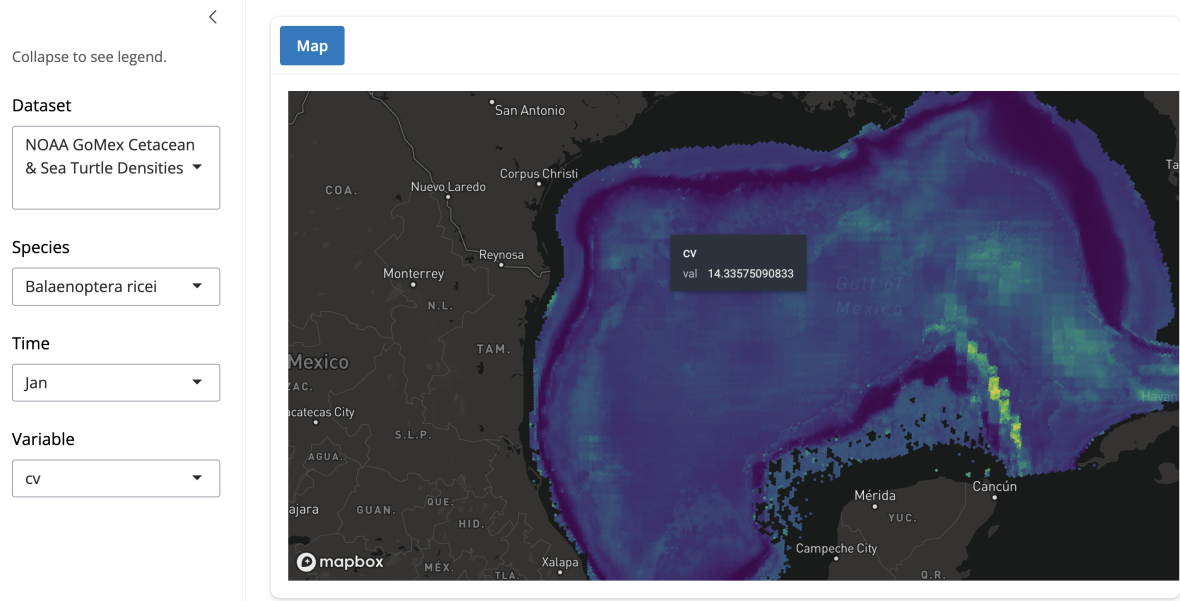
The application source code is at: [apps/mapgl/app.R](#)

Distributions, Vector

Show species distributions with high resolution vectors (as vector tiles).

- [website](#)
- [code](#)

SDM Explorer

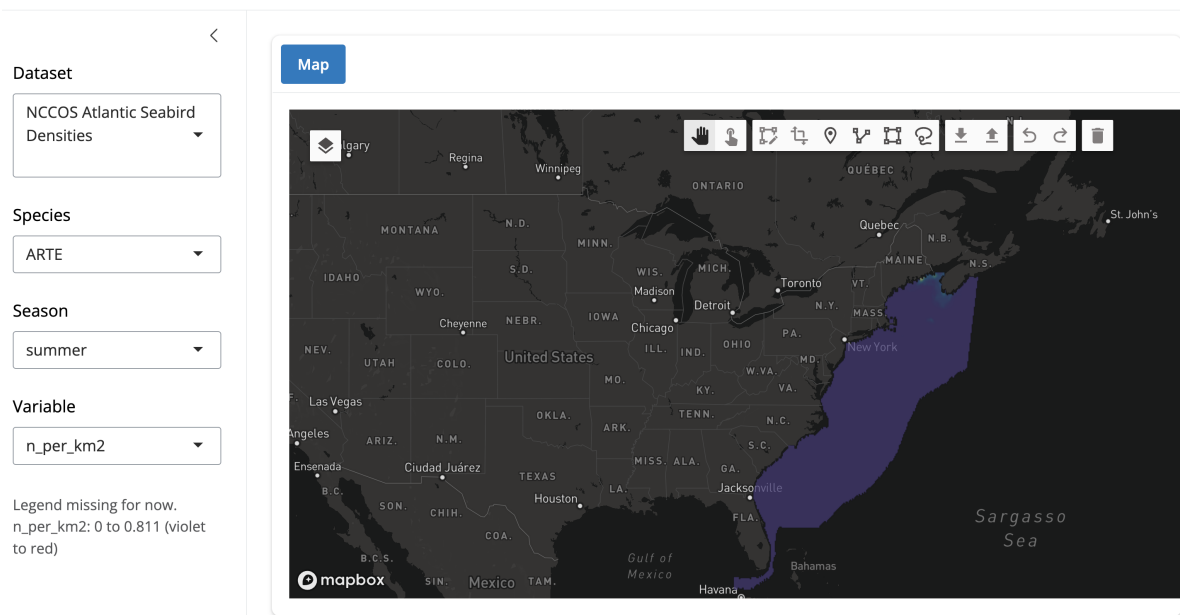


Distributions, Raster

Show species distributions with high resolution rasters as cloud-optimized GeoTIFFs (COGs).

- [website](#)
- [code](#)

SDM Raster Explorer



Vulnerability Mapper

Combine species distribution models (raster) and vulnerability metrics (tables) to identify areas of high conservation concern.

- [website](#)
- [code](#)

