STAC Briefing Report – FISH datasets
Nearshore Ecological Data Atlas
March 1, 2012

Oregon is currently engaged in a marine spatial planning process that will lead to the identification of areas within the territorial sea suitable for ocean energy development. During this process ODFW is responsible for providing pertinent ecological information and identifying the most important ecological areas, relative to goal 19, which should be protected from future development. ODFW's ecological information being used in the marine spatial planning process is a portion of the data in the Nearshore Ecological Data Atlas (NEDA), a collection of spatially represented data sets. NEDA will be an important resource for use in current and future statewide planning and management efforts. While the current planning process is expected to be completed in 2012, we (ODFW) intend to continue work on NEDA, adding datasets and analyses for years to come.

More specifically, NEDA is a collection of ecological data sets (biological, oceanographic, habitat) that are displayed and analyzed in a spatially explicit way. The NEDA datasets that are part of the Territorial Sea Plan (TSP) Part 5 planning process and that can be shared (all non-confidential data) are displayed on Oregon Marine Map, and many are available for download on Oregon Ocean Info. As a planning resource, NEDA will serve the following purposes:

1. Identify existing information relevant for Goal 19 protection and CMSP
2. Make existing information accessible to public and managers in a spatially explicit format
3. Prioritize areas in territorial sea that are important for ecological resources (based on current best available science)

How ODFW approached building NEDA

The inspiration for developing the NEDA was TSP Part 5; however, it was built to be a general resource, useful for ODFW and the public for any management need. Because the TSP Part 5 timeline (2010-2012) did not allow for original data collection, NEDA is currently restricted to existing data, provided by ODFW and other data sources. In identifying existing information relevant for this planning exercise we looked for three defining dataset characteristics:

- Provides coastwide data (preferably as a continuous response surface)
- Data are available (no new data collection; data are already in a spatial format or can be readily converted to a spatial format)
- Data have differing values across the territorial sea (i.e. the data can not have equal value across the planning area – not useful for planning process)

As part of identifying existing information, we have identified data gaps but we have not conducted an in-depth gaps analysis. We also documented data we considered for inclusion in NEDA but did not use, along with the reasons for not including the data at this time. The data gaps and data not used lists are available on the STAC section of the oregonocean.info website.

We organized marine resource data into categories as follows:

- Habitat/Oceanographic
- Fish & Invertebrates (rock habitat as a proxy for rock-associated species)
- Seabirds
- Marine Mammals

Marine resource data are presented in two forms on Oregon Marine Map:

- Basic data (primarily data mapped directly from original sampling, which may have been summarized, but not expanded through modeling)
- Modeled data (original data modeled to develop a continuous response surface over an area defined by the original studies) – includes some datasets within Fish, Seabirds, Marine Mammals

Both forms of the data were then analyzed using Marxan, a software program that has been used worldwide to identify conservation areas based on integration (or summarization) of many input datasets. Marxan provides a spatial solution where a threshold resource value is returned for each target identified – the result is a footprint on a map within which threshold levels (or higher) are present for all targets identified. A "target" may be a resource map (e.g. rock reef habitat) or may be a stratified resource map (e.g. rock reef habitat north coast, rock reef south coast). In this case, we chose stratification for some datasets as we identified targets. We set all target thresholds equally, so that each resource target had equal weight relative to all others (60% target level for all targets). Lastly, we chose to have Marxan run multiple times to produce a "sum run" solution, which approximates an average Marxan output for a given set of targets (and thresholds). Initially, we ran Marxan on each of the 4 categories separately (Habitat, Fish & Inverts, Seabirds, Marine Mammals). Ultimately, ODFW used results from an all-target Marxan sum run, which included targets from all 4 categories analyzed together.

Fish-specific methods

There are 4 analytical approaches used in developing fish data layers and analyses that require additional explanation:

1. Modeling methods used by
    a. NOAA Biogeography Branch on fishery-independent data
    b. TNC (The Nature Conservancy) on fishery-dependent data
2. Proxy datasets (kelp proxy for rock; rock proxy for rock-reef species)
3. Indices (for estuary importance: salmon, ESA, nursery habitat)
4. Stratification (for Marxan; rock reef inputs by depth, N/S, TS/EEZ)

1. Modeling methods

ODFW worked with collaborators over the past year to develop models for fish distribution from survey data (NMFS trawl surveys and ODFW flatfish trawl surveys; modeling done by NOAA Biogeography Branch) and from fishery data (ODFW logbook programs for groundfish trawl and Dungeness crab fisheries; modeling done by TNC). For each of these modeling exercises, we focused on developing distribution information for adult stages of species sampled by the trawl surveys, based on a suite of predictor variables. The trawl surveys sample primarily soft bottom associated species; to account for rock reef associated species, we used rock substrate from seafloor mapping surveys as a proxy for reef fish (see below). We selected complementary and partially overlapping species for each of the two modeling exercises.

NOAA Biogeography Branch modeling: Random Forest

The original trawl survey data had rigorous estimates of species relative abundance, and all individuals were identified to the species level for all catch. Due to the limited time availability of the NOAA Biogeography Branch staff, models were developed for community metrics (species richness, diversity) and for a nearshore species group (selected by ODFW), rather than attempting to model individual species distributions. NOAA used a Random Forest modeling approach described further in **Appendix A**. The metrics they modeled included total fish count, total fish biomass, species richness, species diversity, and total count and biomass for the nearshore species group. The nearshore species group consisting of the following (chosen by ODFW due to relative abundance or importance in waters shallower than 60m): Sand sole, English sole, Pacific sanddab, Speckled sanddab, Starry flounder (included as a species of potential concern), Petrale sole, and Butter sole.

NOAA used 42 environmental predictors (for full list, see **Appendix A**). Briefly, the predictors included information on benthic habitat, distance from shore & shelf, chlorophyll, upwelling, depth, bathymetry (including several derivatives to capture habitat complexity).

Of the metrics modeled by NOAA, four (total species count, biomass, species richness, and nearshore species biomass) were included in NEDA (mapped in marine map, included in Marxan analysis). The species diversity metric was not used on advice from NOAA modelers, due to poor model performance. Nearshore species count was not used on advice from the science workshop because it provided no additional information from the biomass model output.

The map results of NOAA's modeling effort (6 map layers, total) may be found on Oregon Marine Map. Look for these maps in the following file path:

Biology > Fish > Fish Distribution Models (NOAA, 2011)

1. **Species Richness**
2. **Biomass (Weight)**
3. **Abundance**
4. **Biomass for Nearshore Flatfish Species**
5. **Sampling locations**

Biology > Fish > Ecological Analysis using Marxan (TNC, ODFW, 2011)

6. **Fish Marxan Results**

TNC modeling: MaxEnt

Commercial fishery logbook data were more limited in their utility for modeling exercises because not all species were identified and catch estimates were not useful as abundance measures due to several factors (incomplete logbook records, market variation over years and seasons dictate how much is landed, species retained change due to regulatory constraints, etc.).

These data were converted into a location and "presence only" data set, which required a different modeling approach called MaxEnt which is appropriate for use on "presence only" data.

The TNC modeling focused on the following individual species, based on their abundance in nearshore fisheries: Petrale sole, Lingcod, Dover sole, Sand sole, Sanddab (not differentiated to species), Starry Flounder (included as a species of potential concern), and Dungeness crab.

TNC used the same predictor variables as NOAA in their modeling, with some minor differences in how they were calculated

The map results of TNC's modeling effort are incorporated within the fish Marxan run, which may be found on Oregon Marine Map in the following file path:

Biology > Fish > Ecological Analysis using Marxan (TNC, ODFW, 2011)

1. **Fish Marxan Results**

Note: individual species model outputs are confidential, due to the confidentiality of the logbook data used, and are not displayed in Marine Map.

2. Proxy datasets

Rock reef comprises approximately 7% of the territorial sea habitat. Because of its rarity, we wanted to include this habitat type in its entirety in our protected class, for spatial planning. The Active Tectonic and Seafloor Mapping Lab produced a new seafloor map for Oregon state waters in 2011, which, among other resources, utilized ODFW's kelp survey data from the 1990's to help fine-tuned and update the seafloor map. Kelp survey data acquired in 2010 was not available in time to be included in the new seafloor map, so we added to the map for our analyses. Although this resulted in a very small addition of rock habitat to the existing map, it was conceptually important to confirm. The resulting rock reef dataset was then used in the NOAA and TNC modeling, and was used in the Marxan analysis as a proxy for rock reef associated communities (both fish and invertebrates).

3. Indices

There were a number of comments during the science workshop and during public meetings about the need to represent the ecological importance of estuaries to the nearshore ocean. While the Territorial Sea Plan does not include estuaries in its planning area, we recognize the contributions of estuaries to the nearshore environment and visa-versa. For the Marxan analysis, we developed data targets that expressed the relative importance of estuaries using three indices: 1) salmon ranking, 2) nursery area proxy, 3) non-salmonid fish ESA critical habitat score. Since the Marxan analysis required data to be located in the Territorial Sea, it was also necessary to develop a consistent method to attribute the estuary indices to nearshore areas outside the estuary mouths. **Appendix B** describes the estuary indices and method for attributing the indices to the Territorial Sea

4. Stratification

Because Marxan is designed to return a spatial solution that captures your defined targets, it is important to carefully design each target. For NEDA, stratification was used to develop targets within data layers so that additional resource components were accounted for. Stratification of datasets into multiple targets prior to Marxan analysis guarantees that Marxan will select values from within each of the strata in the resource map. For hypothetical example, Marxan might choose a solution where all the rock reef habitat was in shallow water (just by chance). By stratifying rock into strata by depth, Marxan will return rock reef habitat in each of the strata thus returning a solution that has a variety of rock reef habitat depth types. For some resources, we want to protect the variation found in Oregon's resources, rather than simply a percentage of the resource. During the science workshop in September 2011, we received strong guidance to have Marxan return a variety of habitat types in its solution. Stratification helps ensure this.

Stratification applied in Marxan that is relevant to fish distribution:

- Rock reef
    - Territorial Sea vs. Exclusive Economic Zone (east-west stratification)
    - North coast vs. South coast (divided at Coos Bay)
    - By depth (0-30 m, 30-60 m, 60-200 m, 200-700 m, >700 m)
- Fish model outputs
    - Territorial Sea vs. Exclusive Economic Zone (east-west stratification)
- Estuary indices
    - Columbia River vs. South of Columbia River (note: all areas are within the Territorial Sea)

Appendix A:
NOAA Biogeography Branch Random Forest Modeling Methods and Results

Summary

Spatial predictive models for six fish assemblage metrics were developed for waters offshore of Oregon. Analysis was completed using regional fishery-independent trawl datasets. An accuracy assessment using 20% of data removed prior to analysis was completed to assess the validity of model results in nearshore (<3nm) and offshore areas (>3nm).

Predictive models were developed for:
1. All species – biomass
2. All species – count / abundance
3. All species – number of species / species richness
4. All species – diversity
5. Nearshore species – biomass
6. Nearshore species –count / abundance

All models provide reasonably reliable results, except for diversity. The diversity model should NOT be used for further analyses. Overall accuracy ranged from 64% to 87%.

Nearshore Group included: Sand Sole, English Sole, Pacific Sanddab, Speckled Sanddab, Petrale Sole, Starry Flounder, and Butter Sole

Models used 8 distinct spatial layers to derive 42 environmental predictors. Spatial layers and measurements used to develop predictors are defined in **Table 1**.

Table 1. Datasets used to create environmental predictors.

Basic Spatial Layer	Measurement	Criteria
Distance from shore	Average distance to grid cell centroid	Shore identified as 0 m on Coastal Relief Model
Distance from shelf edge	Average distance to grid cell centroid	Shelf edge was 200m isobath)
Benthic Habitat Richness*	Number of distinct habitat classes in grid cell	Habitat class derived from
Benthic Habitat Majority *	Single unique habitat class with greatest area in grid cell	
Chlorophyll hotspots	Presence of highest concentration in grid cell	Provided by TNC. Chlorophyll-a concentrations into two categories, high and low, using >=2 SD above the mean for the high, 1 to 2 SD above the mean for low concentrations
Upwelling hotspots	Presence of longest upwelling persistence in grid cell	Provided by TNC. Upwelling split into two categories, high and low persistence, using >=1.5 SD below the mean for the high persistence and 0.5 to 1.5 SD below the mean for low persistence
Depth**	Mean, minimum, maximum, range at 1600, 4800, 8000 spatial scales	Coastal Relief Model (CRM) used
Bathymetric complexity**	Mean, minimum, maximum, range at 1600, 4800, 8000 spatial scales	Derived from CRM using ESRI spatial analyst extension and 3 X 3 CRM cell neighborhood
Bathymetric slope**	Mean, minimum, maximum, range at 1600, 4800, 8000 spatial scales	Derived from CRM using ESRI spatial analyst extension and 3 X 3 CRM cell neighborhood
Bathymetric Slope of Slope**	Mean, minimum, maximum, range at 1600, 4800, 8000 spatial scales	Derived from CRM using ESRI spatial analyst extension and 3 X 3 CRM cell neighborhood
Bathymetric Aspect **	Mean, minimum, maximum, range at 1600, 4800, 8000 spatial scales	Derived from CRM using ESRI spatial analyst extension and 3 X 3 CRM cell neighborhood

*Benthic habitat maps developed using best information available. Merged v3.1 and v3.5 maps provided by ODFW. Merge defaulted with 3.1, unless 3.5 was available, which occurred within State waters. These datasets were later determined to have errors. After careful deliberation and analysis of predictor contributions we determined that the contribution of these derived layers would not impact predicted spatial patterns.
**Min, mean and max metrics used to characterize bathymetry and derived bathymetric variables (layers 5-8) were calculated for 3 spatial scales. Spatial scales were equivalent to a 1600m X 1600 m area, 4800 m X 4800 m area and a 8000m X 8000m area. These areas correspond to the lowest resolution possible given the spatial framework we were working in and from there.

Data Sources

Abundance, biomass and taxonomic data were obtained from three data sources, all of which collected data using benthic trawls:

1. ODFW /OSU trawl data
2. NWFSC shelf and slope surveys
3. AFSC shelf and slope surveys

Data preparation and modeling methods

All metrics were standardized by area swept (distance X trawl width) to allow distinct trawl datasets to be merged.

Water hauls identified following Zimmerman et al. 2011 (collected before 1995 and with less than 1kg/10000m^2) were removed from analysis. Water hauls are abnormal data.

All data in the ODFW/OSU and AFSC were analyzed. Only data with haul type =3 and performance >=0 from NWFSC were analyzed.

Spatial analysis was undertaken using a contiguous grid developed by joint efforts between the Biogeography Branch, ODFW and TNC. The grid was derived from BOEMRE lease blocks (http://www.boemre.gov/offshore/mapping/) and covered shelf and slope waters. Grid cells measured 1600m by 1600m, such that 9 grid cells fit within 1 lease block. The resolution of grid cells determined the resolution of output.

Fish metrics were predicted to all grid cells in the spatial analysis grid using fish-environment relationships modeled using boosted regression trees for continuous variables and binary logistic regression trees for categorical data (High/Low classes). High values (hotspots) were defined as the top 10% of values (=>90th percentile). For all assemblage level fish metrics, the binary classification outperformed continuous predictions. Due to poor performance and high uncertainty/error for continuous metrics, only hotspot maps were produced and assessed with an independent data set (random 20% of nearshore & offshore trawl samples held back n=1,534). See example in Fig 1 below.

The independent map accuracy assessment provided an error matrix with the percentage of correctly classified grid cells for each class. The continuous variables were assessed by the strength of correlation between predicted and actual values.

Upon request we mapped both continuous abundance and biomass and hotspots for the nearshore fish group.

Overall, the best performing models based on independent map accuracy assessment were logistic regression trees grown independently by the algorithm without cross-validation or any other form of testing. These models were computed using the maximum allowable number of trees (usually 1000 trees) and in most of the models all environmental predictors contributed to the final predictions, although the first three contributed most.

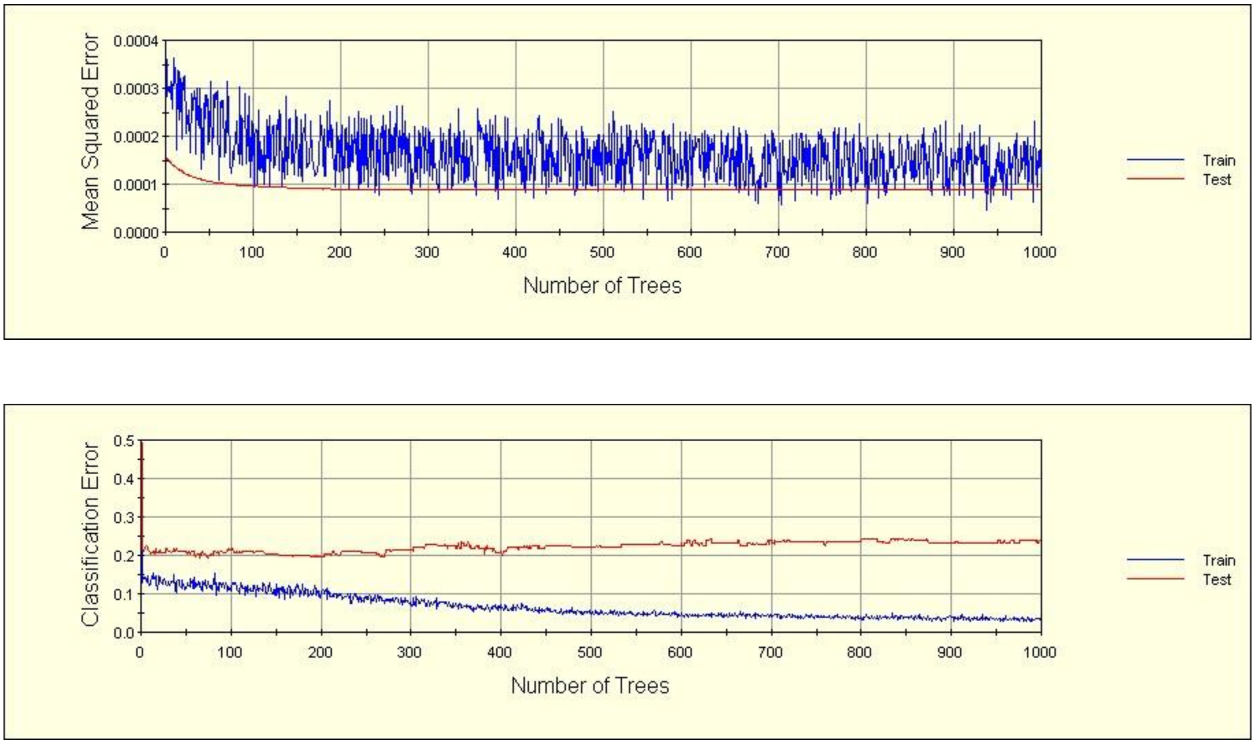

Figure 1. Example of model output showing (A) high variability in error from trees predicting continuous biomass; and (B) smaller error in classification of categories (High/low) of biomass:

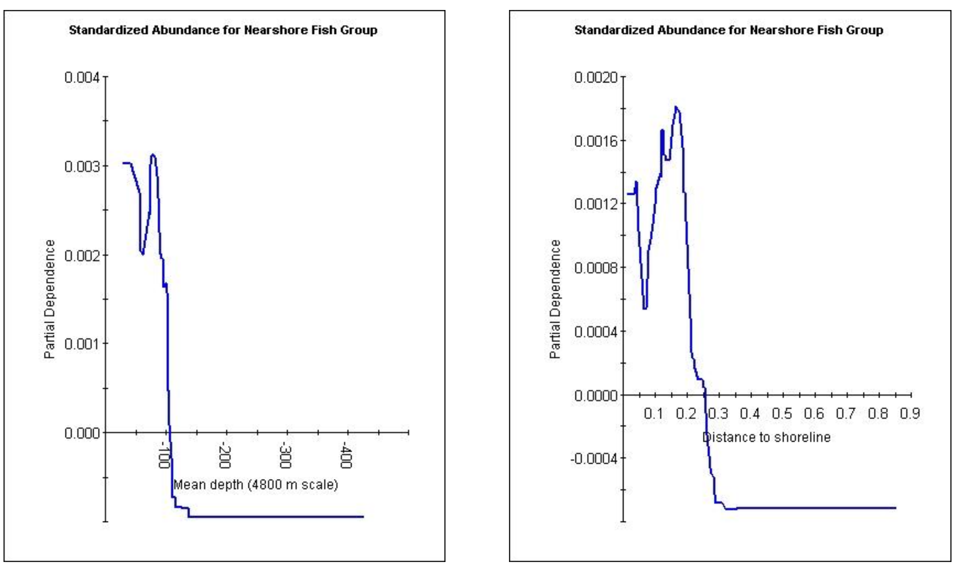

Figure 2. Example of output from plots of relationship of std abundance of nearshore fish group with top two environmental predictors (mean depth and distance to shoreline). Environmental thresholds exist that constrain the abundance of 6 chosen species. Note positive influence on abundance in waters shallower than 100 and closer to shore than approximately 0.25.

Maps

Two different predictions are provided a categorical prediction and a probability. The probability is derived from the response variable output from the binary logistic regression. Areas with greater probability are more likely to be hotspots. The categorical prediction is derived from this probability. Probabilities > 50% are classified as a hotspot or “High” and probabilities <= 50% are classified as background or “Low”.

Accuracy Assessment

Metric	Overall accuracy	% high misclassified	% high misclassified (only nearshore)	% low misclassified	% low misclassified (only nearshore)
Biomass	87%	82%	?	7%	73%
Diversity	23%	82%	3%	2%	13%
Abundance	74%	65%	100%	22%	21%
Richness	64%	46%	?	35%	53%
Near. Count	83%	18%	7%	17%	45%
Near. Biomass	63%	80%	5%	16%	30%

<u>Appendix B:</u>
Estuary Indices

The following discussion describes the estuary data layers used as targets in the 12/2011 Marxan analysis performed by TNC for ODFW. The description includes:

1) how we developed the indices used for the three data layers
2) how we defined which analysis units cells the indices would be attributed to.

Data Layers:

<u>1. Salmon Ranking:</u>
All freshwater outlets to the ocean (i.e., streams, rivers, estuaries) that harbor significant runs of salmonid species were ranked according to their perceived importance relative to both species diversity and abundance (species-specific or cumulative). The following criteria were considered during the ranking process: watershed acreage, number of salmonid populations, and whether the freshwater system had a "lake" component to it or not. Watershed acreage was used as a proxy for water volume and flow rate of the system. This proxy was considered a good indicator because precipitation rates are relatively consistent coast-wide (although southern Oregon is slightly drier than central and northern Oregon). The number of salmonid populations was defined as the total number of distinct runs (e.g., fall run, spring run) of chinook salmon, coho salmon, pink salmon, chum salmon, and steelhead; cutthroat trout were not considered. While the number of salmonid populations per watershed is discrete, some populations are not self-sustaining and must depend on source populations in nearby watersheds; however, we did not have access to this type of information for all species and watersheds under consideration, so this was not taken into account during the ranking. Further, all species/run combinations were treated equally; ESA-listed runs were not given more weight. Data for watershed acreage and number of salmonid populations were obtained from The Nature Conservancy, which acquired the data from the Wild Salmon Center (both Non-Governmental Organizations) for use in their 2011 assessment of West Coast estuaries (Gleason et al., 2011). Determination of whether a riverine system had a "lake" component (loosely defined as a significant slow-moving body of water through which the water flows that persists through the summer months) was made by ODFW biologists. This criterion was deemed important because lake habitat offers opportunities for additional juvenile rearing habitat and productivity (i.e. food). All systems in which lake habitat was present were increased in rank by one unit. Ranking was from 1 (low) to 5 (high), and generally fell into the following categories: 1) very large river systems (i.e. the Columbia River), medium-large coastal rivers (i.e., Rogue and Umpqua Rivers), medium-sized coastal rivers (e.g., Yaquina River) and large creeks with lake components (e.g., Tahkenitch Lake), small rivers (e.g., Salmon River) and medium-sized coastal creeks (e.g., Hunter Creek), and small coastal creeks (e.g., Beaver Creek). All decisions were made based on consensus professional opinion of ODFW MRP salmon staff and ODFW coastal district biologists.

Gleason MG, S Newkirk, MS Merrifield, J Howard, R Cox, M Webb, J Koepcke, B Stranko, B Taylor, MW Beck, R Fuller, P Dye, D Vander Schaaf, J. Carter 2011. *A Conservation Assessment of West Coast (USA) Estuaries*. The Nature

Conservancy, Arlington VA. 65pp.

2. Nursery Area Proxy:
Estuarine aquatic bed and intertidal habitat are considered important nursery areas for Dungeness crab and several marine fish species. The total surface area of these habitats within estuaries provides a proxy for the relative importance of each estuary to nursery area functions. The surface area values for larger estuaries were derived from DLCD (1987) (P. 33, table entitled Estuarine Habitat Class Distribution by Estuary) and represent the sum of areas of the following habitats: subtidal aquatic bed, intertidal shore, intertidal flat, and intertidal aquatic bed. Surface area values for estuaries not listed in the Oregon Estuary Plan book were derived from National Wetland Inventory maps (aquatic bed and intertidal polygons - wetland codes E1AB, E2AB, E2US) (US Fish and Wildlife Service 2011). Surface area units are in acres.

DLCD. 1987. The Oregon estuary plan book. Salem, OR: Department of Land Conservation and Development. 126 pp.

US Fish and Wildlife Service. 2011. Wetlands and Deepwater Habitats of the Conterminous United States, National Wetland Inventory Maps. Washington, DC: US Fish and Wildlife Service.).

3. Non-salmonid fish ESA critical habitat score:
These data indicate estuaries and/or their river systems with designated ESA critical habitat for eulachon or green sturgeon (southern DPS). A score of "1" is given to estuaries with critical habitat for one of the species, and a score of "2" is given to estuaries with critical habitat for both of the species. The Columbia, Coos, Umpqua, Yaquina, and Siuslaw have critical habitat for green sturgeon, and the Columbia, Upmqua, and Tenmile Creek have critical habitat for eulachon. The information source is the Federal Register notices designating critical habitat for these species.

Determination of analysis unit cells outside the mouths of estuaries to attribute estuary metrics:
The purpose of developing estuary metrics in the Territorial Sea planning effort was to map the influence of the estuary on the nearshore ocean. To accomplish this, it was necessary to attribute the estuary metrics to an ocean area outside of the mouth of each estuary. Since there is no standard, scientifically-accepted way of doing this, we developed an “attribution area” based on ranking each estuary in terms of its total surface area and the surface area of its watershed. The rationale for choosing this method is based on the notion that the amount of water moving between the estuaries and ocean due to both tides and river flow can provide a relative measure of the influence of the estuary on the nearshore ocean. Estuary area was used as a proxy for tidal prism and watershed area was used as a proxy for river flow. The estuary areas were derived from DLCD (1987) and Lee and Brown, eds. (2011), and watershed areas were derived from US Department of Agriculture, Natural Resources Conservation Service Watershed Boundary Dataset.

Estuaries were ranked from 1 through 5 (smallest to largest) for both estuary and watershed surface area based on natural breaks in the surface area data. Total estuary surface area was averaged for each size category and used to compute a radius of a semicircle of those averaged areas. These radii were rounded to the nearest 1000 m and were used to define the ocean areas within which to attribute the estuary metrics (radius drawn from the mouth of each estuary). The radii are as follows: rank 1 – 1,000 m, rank 2 – 2,000 m, rank 3 – 3,000 m, rank 4 – 5,000 m, rank 5 – 20,000 m (Columbia River only). For all systems except the highly river-dominated Rogue and Chetco estuaries, the estuary area rank was greater than or equal to the watershed area rank, and the radii were based on the estuary area rank. For the Rogue and Chetco, the radius of influence was based on the watershed area rank rather than the estuary area rank.

The metrics were then attributed to analysis unit grid cells based on the radii. For 1,000 and 2,000 meter radii, analysis unit cells were selected that touch the semicircle drawn by the radius. For larger radii, analysis unit cells were selected whose centroid falls within the semicircle drawn by the radius.

DLCD. 1987. The Oregon estuary plan book. Salem, OR: Department of Land Conservation and Development. 126 pp.

Lee II, H. and Brown, C.A. (eds.) 2009. Classification of regional patterns of environmental drivers and benthic habitats in Pacific Northwest estuaries. U.S. EPA, Office of Research and Development, National Health and Environmental Effects Research Laboratory, Western Ecology Division. EPA/600/R-09/140.