NOAA Data Quality Guide

The following spatial data quality issues should be considered when evaluating how data can be effectively used in a geographic information system (GIS). The usage caveats derived from these data issues should be carefully considered when interpreting any analyses based on data layers.

Topics:

- General data issues
- Coastline Artifacts
- Data Aggregation Errors

General data issues

▲ Georeferencing Errors

Errors in the processing of GIS data can sometimes create inaccuracies in the spatial representations of geographic features. While these types of errors can be insignificant when creating large-scale maps, they become more problematic when relying on data to provide accurate information at very specific locations.



Above: Spatial errors visible in a 1:2,000,000 scale polygon representing Treasure Island in the San Francisco Bay.

▲ Spatial Resolution

All GIS data is created with accuracy and detail appropriate for use at specific spatial scales. When low-resolution data is used in a high-resolution context it may not adequately describe the complexity of the features it is meant to represent.



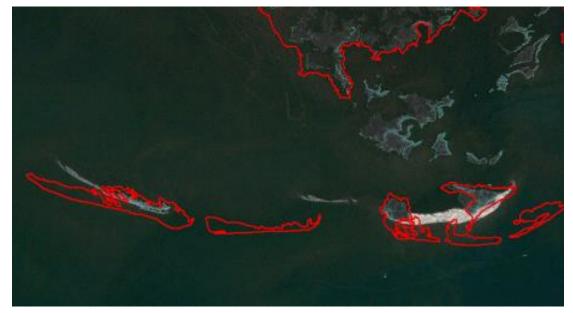


Left: A island off the coast of Maine represented by a low-resolution (1:3,000,000 scale) coastline.

Right: The same island represented with a higher resolution (1:100,000 scale) coastline.

▲ Out-of-date information

Geographic and environmental features can change significantly over time, particularly in the marine and coastal environment. As ground conditions change, GIS data representing conditions at a specific point in time will becomes increasingly unreliable.



Above: Out-of-date spatial data representing barrier islands off the coast of Louisiana overlaid on more recent imagery.

▲ Inadequate Metadata

The metadata describing a GIS dataset is a critical component in evaluating how it can and should be used. If metadata does not exist or if it is incomplete, it can be very difficult to accurately interpret the information that the data provides.

▲ Overly Exclusive Data

Attempts to refine the spatial extents of EFH using high-resolution habitat or coastline data can result in a data set that has a high probability providing false negative EFH information (i.e. indicates that EFH is not present when it actually might be). In situations where a conservative (or precautionary) analysis is required, highly refined datasets may be too exclusive to appropriately account for the uncertainty involved in spatially defining marine habitat.

Coastline Artifacts

Marine GIS data layers that extend to the shoreline are generally cropped using one of the many available shoreline polygons. If the GIS data is going to be used to return location-specific information in near-shore or estuarine areas the following potential artifacts inherent in shoreline data should be considered:

▲ Spatial Resolution of Shoreline

If the resolution of the shoreline data used to define the boundaries of EFH is too low to conform to the complexity of actual geographic shoreline features, it will not be reliable in near-shore areas.



Left: An aerial image covering an approximately 40 sq mile area of the coast of South Carolina showing Port Royal Sound branching into the Broad, Beaufort, and Chechessee Rivers. **Right:** A 1:3,000,000 scale coastline representing the same area.

▲ Inland Extents

GIS shorelines generally describe the interface between open water and intertidal land areas. EFH data clipped to a shoreline may limit the inland extent of aquatic species to the open water and falsely exclude what might actually be important habitat areas.



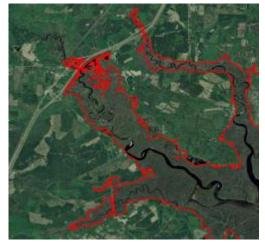
Left: An aerial image covering an approximately 7.5 sq mile area of coastal South Carolina showing a section of the Broad River with the land-water boundary defined by a standard 1:2,000,000 scale coastline.

Right: The same area with the with an inland boundary defined by the limits of the estuarine system as mapped by the National Wetlands Inventory.

▲ Upstream Extents

The upstream ranges of marine and anadromous species depend on many physical and geographic factors specific to the area of interest. Often the polygons that are used to represent species distributions were not created with any consideration of the factors that would determine the actual upstream extents of a species, but rather incorporate the arbitrary delineation between a river and a bay that are inherent in the coastline data used.





Left: A 5 square mile area of Coastal South Carolina showing imagery of the branches of the Tultifinny River with a 1:10,000 scale coastline overlaid in red. Note the arbitrary upstream cutoff points of the river branches.

Right: The same area with the upstream and inland extents defined by the limits of the estuarine system as mapped by the National Wetlands Inventory.

\land Land Overlap

In an attempt to maintain a conservatively inclusive boundary for EFH, data may not be clipped to a shoreline at all, but rather allowed to overlap land. Data of this type can be overly inclusive and may not be appropriate for analyses that require some more refined mapping of upstream or inland extents.



Above: EFH polygons for bonnethead shark shown overlapping land.

Data Aggregation Errors

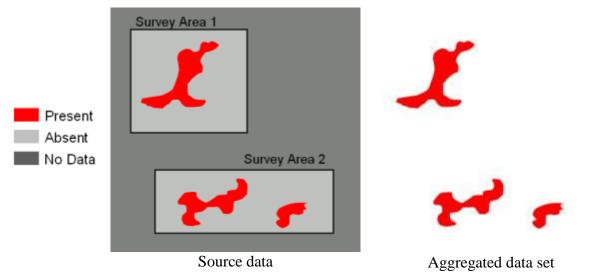
Habitat mapping in the marine environment is rarely done on a large scale. It is therefore common that regional-scale EFH maps rely on the aggregation of numerous small-scale habitat datasets. GIS datasets created in this way should be used cautiously due to the following common issues:

▲ Inconsistent Methodologies

Marine habitat datasets from different sources rarely use the same methodologies or classification systems. The process of combining them, therefore, often must reduce the classification scheme to the lowest common denominator of the two datasets. In this process, important distinctions in habitat type, density, or patchiness can be reduced to simple presence absence. This process can also create edge matching artifacts where adjacent polygons from different data sets do not line up correctly due to variations in sampling methods or changes in the ground conditions at the different times.

▲ Loss of "no data" information

In the process of assembling an aggregated dataset, the information about the bounding areas of the individual datasets is often lost. When this happens, areas where there was no data collected become intermingled with the portions of the surveyed areas where there was no habitat found. If the absences of habitat and the absence of data are indistinguishable, the entire dataset becomes unreliable and tends to falsely imply absence in areas where, in fact, it contains no information.



Left: An illustration of source data maintaining the distinctions between areas of presence, absence, and no data.

Right: An illustration of how the distinction between absence and not data can be confused in an aggregated dataset.

▲ Loss of source metadata

In an aggregated dataset each polygon can potentially be derived from a different source data set. If there is no attribute associated with each polygon to indicate its source, this important information can be lost.

Source: https://www.habitat.noaa.gov/application/efhinventory/

Resources and Help:

Data Quality