Paper for consideration by DQWG

Modeling Approximate Areas

Submitted by:	Jeppesen/BSH/NIPWG
Executive Summary:	Describes approaches to the representation of approximate areas in nautical publications information, maritime safety information, and other domains.
Related Documents:	(1) NIPWG2-6.3 Annex A / S-100WG1-10.7A; (2) NIPWG1-21.3
Related Projects:	(1) S-100; (2) S-122, S-123, S-124, S-126, etc.

1 Introduction

The work on the S-122 (Marine Protected Areas) data sample revealed shortcomings in the ability to represent approximate areas (or "fuzzy areas") in S-100. Approximate areas are relevant to several NIPWG data products. Further, similar issues arise in S-124 (Marine Safety Information), and are likely to arise in other product specifications. The NIPWG2 paper cited in "Related documents" describes several solutions, some of which use ideas similar to data quality work in DQWG. Since the issues and solution approaches appear to be similar, this paper describes those approaches and requests DQWG assistance with defining solutions.

The scope of this paper is limited to approximate areas of the types which are encountered in maritime information domains and which it would be useful to represent in S-100-based data sets. It includes only areas - approximate points or curves are not addressed.

1.1 Discussions in NIPWG and S-100 WG

The original version of this paper concluded that different aspects of the problem needed to be addressed by the S-100WG, NIPWG, and DQWG and discussed a combined solution approach, consisting of using modeling structures allowed by S-100 Edition 2.0.0, extensions to S-100 Edition 2.0.0, and formalization of terminology used for describing approximate areas (the last, with the intention of making it easier to convert these descriptions to an object/attribute model as used in S-100). The current position of the S-100 WG is that the existing framework in S-100 edition 2.0.0 should be sufficient and extensions to S-100 are not needed.

1.2 Background

The basic question is: **How will the mariner assess the nautical information in an S-100 environment?**

For ships sailing under the SOLAS requirements, there exists an information entity which combines charts and publication. That means no mariner can navigate a ship safely without having consulted the associated nautical publications. Both charts and publications can be available in digital or printed format. Some HOs provide their publications in digital format but the majority of HOs is issuing their nautical publications in printed format.

In an S-100 based ECDIS system and with the existence of Data Products which either interact with each other or which overlay each other, the Mariner will have access to the information only by the system. The intuitive calculation of the presence of an effect or service is no longer possible. The areas where such effect might occur or a service is available must be indicated visually or by a characteristic which indicated the "fuzziness" of the existence of an effect or service.

Publications provide information which is currently not available from the chart, e.g. traffic regulations, effects on the passage during different environmental conditions. The mariner has to compare the charted situation with the environment and the associated publication information.

Publications are often not able to determine the exact location where one particular effect might occur. Rather they provide more abstract locations, e.g. "in front of the harbour entrance". In theory the area "in front of the harbour entrance" can be extended thousands of miles. In practice the mariner calculates intuitively to which distance off the harbour entrance the effect might occur.

In addition to the hydrographic environment, sailing directions for the passage of islands or headlands are often described in a way that mariners have to take their ships condition into account. That may have effects on the area where a ship can navigate. The navigable area (recommended passage) may have specific limits to the area where the water depth is less than allowed but has an uncertain extend (a fuzziness) to the open sea. This spatial extent to the open sea cannot be determined by fixed values. Further, some areas might be passed by vessels carrying specific cargo in a defined distance. The outer limits of the passage are often unspecified. Mariners have to create a route which does not touch the inner limits and which keeps a certain distance to this defined inner limits. The "definition" of the outer limits, which in fact don't exist, depends on the mariner's interpretation. A certain level of fuzziness would be helpful in defining such uncertain limits and would avoid unnecessary deviations.

2 Analysis/Discussion

The discussion in this paper is confined to issues that are similar to data quality issues. A more complete discussion can be found in the NIPWG2 paper (see "Related documents").

2.1 Examples

Some examples of approximate areas follow:

1) North American right whale sighting areas off the U.S. Atlantic coast are described in general terms as, variously "year round in the north-eastern region from Cape Code to Nova Scotia", "during winter and early spring, calving area in the coastal waters off the Southeast Atlantic coast," and "during fall, the migration route from the northern zone to the calving area, which runs through near-shore waters along the mid-Atlantic coastline". The restricted areas designated in regulations as seasonal management and critical habitat areas are at best substitutes for the approximate areas described above. The approximate areas are depicted on fact sheets from NOAA Fisheries [FS1, FS2].

2) Radio services may be received in a certain distance from ashore for sure and with decreasing reliability at increasing distances beyond this distance. Encoding this in a dataset would require a graduation of the service reliability which could fuzzy out at the outer edge.

   The limitation of the portrayal to only one area results in an insufficient access to the information. The reality will show the mariner that the information is either valid or not. Ideally, the chart should depict an area where the service is 100% available and an extended area where the presence of the service is likely but not guaranteed. This extended area cannot be represented with currently existing or planned methods of encoding uncertainties (CATZOC or the planned replacement of CATZOC), in large part because the factors and magnitudes are very different.

   The example below shows the hypothetical decrease in reliability of a radio service broadcasted from the island. The arrows are only placed for presentation purposes. The lighter shade indicates lower reliability of reception.

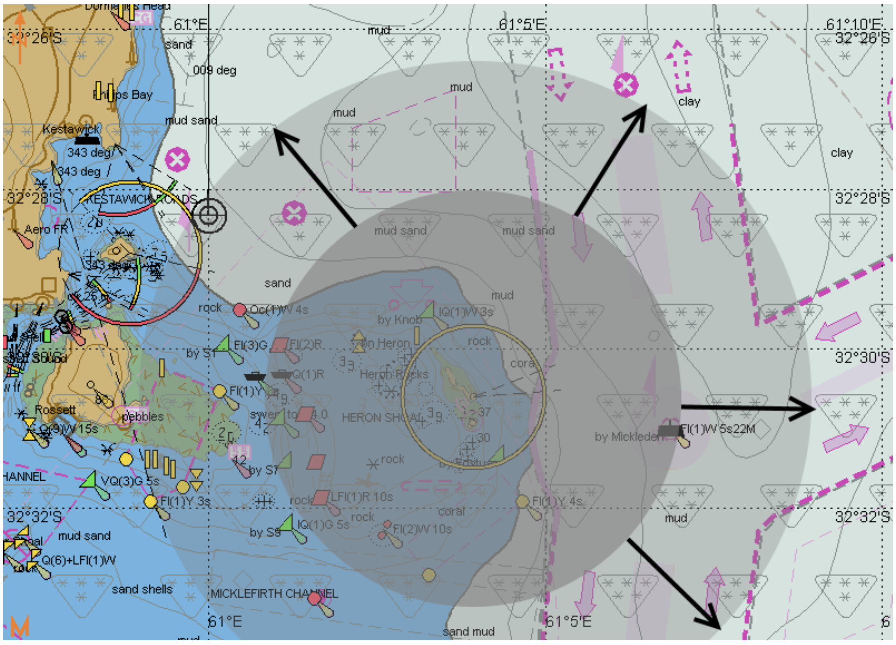

Figure 1. Example of radio service with areas of definite and uncertain reception

NIPWG2-6.3 Annex A contains other examples.

2.2 Methods of modeling approximate areas

Several methods of modeling approximate areas are described in the NIPWG2 paper. To begin with, the NIPWG intends to consider the following approaches:

1) Likelihood zones: Likelihood-based methods represent the likelihood of a specific point location being in or out of the area by a numerical value. The likelihood zones method defines a set of enclosing polygons representing stepwise likelihood of a location belonging to the region, ranging from (say) 100% (definite) to 5% (negligible). The number of steps would have to be determined, but 2-5 intermediate levels may be appropriate.

2) Direct use of data quality information: Applying the attribute *quality of position* to the relevant boundary spatial attribute with value 4 (approximate) indicates that the boundary is approximate but provides no indication of the magnitude or extent of the uncertainty.

3) Positional uncertainty value attached to the boundary curve: The magnitudes of uncertainties would need to be appropriate and the semantics of uncertainty for curves need to be defined (e.g., does it mean displacement of the curve as a whole, the radial uncertainties for each individual point in the segments, or something else?)

The figure below shows hypothetical ways of using a zone-based model to define fuzzy areas around a generic feature, by defining a generic “FuzzyZone” feature with a single attribute for the strength of the association to the core (either a percentage or an ordered list of predefined intervals of strength). Instances of this feature are associated with the core feature to define the less likely outlying zones. The fuzziness can be expressed numerically (Alternative 1) or stepwise (Alternative 2). E.g., for the radio service example the lighter annulus would be a FuzzyZone with attribute fuzzyPercentage = 50 defining the reliability as 50% to 99%.

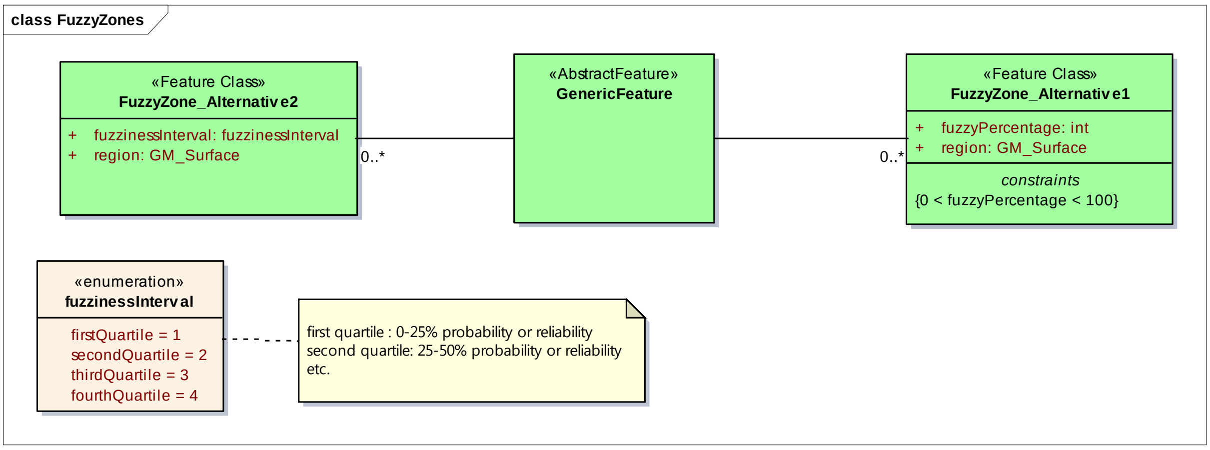

Figure 2. Generic zone features carrying attributes for the degree of fuzziness

2.3 Fuzzy terms and zones of likelihood

Approximate areas are often described using imprecise terms. Such fuzzy terms include, example "in the vicinity of", "near", etc., relative to a specific feature, landmark, or location coordinates. Formalization of a restricted vocabulary of such terms and what they mean in terms of fuzziness is needed in order to convert the areas in question into a likelihood/zone-based model that can be processed and portrayed by software.

To avoid a proliferation of slightly different definitions of fuzziness, it is necessary to define a system for categorizing positional uncertainty similar to the CATZOC system, but limited to positional uncertainty (CATZOC considers depths too), and with break points at values that are more realistic for the types of fuzziness encountered in nautical information, as in the examples above. The current CATZOC system appears too fine-grained in distinguishing categories by meters and 10's of meters. Magnitudes in the current model of uncertainty are too small for the approximations needed. For the existing system of Zones of Confidence, even CATZOC "C" corresponds to positional uncertainty of ±500m, all uncertainties larger than that are classified as D

Possible criteria and examples are in the table below (BB = bounding box; ZOF = zone of fuzziness). It is a starting point based on the CATZOC table but the factors are different from CATZOC.

Table 1. Fuzziness zones

ZOF	Location Accuracy	Boundary uncertainty	Temporal variability	Other variability	Reliability or Probability	Terms
X	±1NM to ±2NM	5% of BB diagonal	(dependency on time of day, season, etc.)	(dependency on wind, water level, tides, etc.)	90%	in the vicinity of
Y	±2NM to ±5NM	10% of BB diagonal			75%	near

Zones might be simplified into a three-tier model (e.g., definitely within / probably within / possibly within) – but this has not been discussed in NIPWG and it may be an over-simplification with no benefits from conforming to the three-tier data quality model that closely.

2.4 Portrayal

Portrayal of fuzzy areas should be addressed in detail later, but in general depiction of a likelihood shadow, e.g., by a thicker line, shadows, solid or gradient colours, or thinner or grey line or lines at specific likelihood contours.

Likelihood zones could be treated as just polygons of different symbol / hash densities, color, etc. Likelihood contours might be depicted as curves either with or without a visual association to the 'core' (on the same principle as limits of territorial seas and the relevant coastline or baseline).

Other options include a decorated boundary e.g. hachures, or broken zigzag lines, etc.

The processing power needed for portrayal should be considered when developing portrayal rules and symbols.

The portrayal proposals will be developed by the NIPWG and forwarded to the NCWG for further consideration when the quality description of zones of fuzziness has been defined by the DQWG.

3 Recommendations

1) The initial implementation should be focused on likelihood zones/contours and on the "fuzzy terms" association class approach.
2) Define a system for categorizing fuzziness. This might best be done as a joint effort with NIPWG members can be expected to be more familiar with the significance of terminology and magnitudes of approximations.

4 Impacts

Developing the classification will probably involve preparation and group evaluation spread over 2-3 meetings of the DQWG and NIPWG. Impacts on product specifications should be low since existing application schemas are not affected.

5 References

[FS1] North Atlantic Right Whale Fact Sheet, NOAA Fisheries, Southeast Regional Office. Location: <http://sero.nmfs.noaa.gov/protected_resources/outreach_and_education/index.html> (retrieved 12 Jan 2016).

[FS2] North Atlantic Right Whale Seasonal Distribution and Habitat Use (Fact Sheet), NOAA Fisheries, Southeast Regional Office. Location: <http://sero.nmfs.noaa.gov/protected_resources/outreach_and_education/index.html> (retrieved 12 Jan 2016).

6 Actions Requested

The DQWG is invited to:

- advise on the modeling of approximate areas, drawing on experience in modeling data quality;
- advise on, and participate in, the development of a system for classifying the degree of fuzziness of spatial objects as described by imprecise terminology;