Whales and Ships: Shall they never meet.

Data Source Descriptions and Metadata

The following documentation describes each of the data sources provided to teams for the CANDEV challenge. All sources have been preprocessed to facilitate quick usage and reduce the need for data cleaning and filtering. All spatial data is referenced using the WGS 84 (EPSG:4326) CRS.

Ship Track AIS Data

Data Format	Tabular (CSV)
File organization	One file per month of 2020
Total file size	466.3 MB (61.6 MB zipped)
Source	MarineCadastre – Vessel Traffic Data
Description	A dataset containing one year (2020) of AIS transmissions along the Atlantic coast of the United States. Each record in the dataset corresponds to a transmission sent by a ship. See Appendix A for a description of the dataset columns.
Additional details	- See Appendix A for the data dictionary - Cleaning has been applied to ensure that every record has a value in the columns of MMSI, UTC_Timestamp, Lat and Lon. All other columns may contain null values in the absence of data. Notably, the columns for Cargo, Draft and Status have many null values. - The following filters have been applied to reduce the total file size: - Maximum of one transmission per 30 minutes for each ship - All pleasure craft ships removed - Spatial extent limited to UTM 19-T - See https://coast.noaa.gov/data/marinecadastre/ais/faq.pdf for a detailed FAQ pertaining to the original unfiltered dataset

Pseudo Habitat Suitability Models

Data Format	Raster (TIF)
File organization	One file per month
Total file size	4.7 MB (146 KB zipped)
Source	Marine-life Data and Analysis Team – Marine Mammal Abundance
Description	Raster representations of pseudo habitat suitability for North Atlantic Right Whales along the Atlantic coast of the United States. Each raster cell is encoded

	with a numeric value in the range [0,1] representing the habitat suitability within the spatial extent covered by the cell. Low values indicate poor suitability while high values indicate good suitability.
Additional details	• Originally a set of abundance rasters, cell values have been mapped to pseudo habitat suitability values using a normalization process

Slow Zones

Data Format	Vector (GeoJSON)
File organization	One file per zone
Total file size	23 KB (4 KB zipped)
Source	NOAA Fisheries – Right Whale Seasonal Management Areas
Description	Files encoding the geographic boundaries of zones within which mandatory regulations have been put in place to limit the speed of ships in order to reduce the likelihood of injurious collisions with whales. Each zone is defined as a sequence of geographic coordinates outlining the perimeter of a polygon.

Additional details	• Only the zones within the spatial extent of UTM 19-T are provided
	• Ships 19.8 meters or longer must travel 10 knots or less in slow zones
	• The speed restrictions are limited to the following date ranges:
	    o Cape Cod Bay: Jan 1 – May 15
	    o Off Race Point: Mar 1 – Apr 30
	    o Great South Channel: Apr 1 – Jul 31
	    o Block Island Sound: Nov 1 – Apr 30

Marine Protected Areas

Data Format	Vector (GeoJSON)
File organization	One file with all zones
Total file size	8 KB (2 KB zipped)
Source	Protected Planet – World Database on Protected Areas
Description	A file encoding the geographic boundaries of zones within which policies of marine conservation and protection have been implemented. Most commonly, the policies place restrictions on fishing activities. Each protected area is defined in the file as a sequence of geographic coordinates outlining the perimeter of a polygon.
Additional details	The spatial extent has been limited to UTM 19-T

Collision Lethality Model

Data Format	Python script (Py)
File organization	Single file
Total file size	1 KB
Source	Model source: https://onlinelibrary.wiley.com/doi/full/10.1111/j.1748-7692.2006.00098.x Simplified model creation: DFO Mentor – Jim Theriault Script author: DFO Mentor – Lee Croft
Description	A simplified lethality model for collisions between ships and whales. The model is a function of ship speed (in knots) to probability of lethality for a whale in the event of a collision. The script provides both a set of discretized points for the function as well as a Python function to give linear interpolations between the data points.
Additional details	Usage instructions: • Save the script in the same folder as the rest of your code • Load the script using "import lethality_model" • Access the discretized data points using "lethality_model.data_points" • Access linear interpolations between the data points using "lethality_model.leth_fnc(speed)" where speed is the ship speed in knots

Appendix A – AIS Data Dictionary

Column Name	Column Description
MMSI	Maritime Mobile Service Identity – A unique numerical identifier for the ship.
UTC_Timestamp	UTC date and time of transmission.
Lat	Latitude of ship at transmission.
Lon	Longitude of ship at transmission.
SOG	Speed in knots of ship at transmission.
COG	Course in degrees of ship at transmission.
Heading	Heading in degrees of ship at transmission.
VesselName	Name of ship. May be misspelt of missing in some cases.
VesselType	Numerical code indicating the ship type. See: https://coast.noaa.gov/data/marinecadastre/ais/VesselTypeCodes2018.pdf
Status	Navigational status of ship. See: https://help.fleetmon.com/en/articles/4476744-ais-navigational-status
Length	Length of ship in meters.
Width	Width of ship in meters.
Draft	Draft depth of ship in meters.
Cargo	Cargo type of ship. See: https://coast.noaa.gov/data/marinecadastre/ais/VesselTypeCodes2018.pdf
TransceiverClass	Class of AIS transceiver. A: Mandated transceivers. B: Recreational craft transceivers.