### Water Quality

**Layer type(s):** Pressure, Clean Waters goal layer

**Data description:** The water quality layer is derived from the US Environmental Protection Agency [National Coastal Condition Assessment (NCCA)](https://www.epa.gov/national-aquatic-resource-surveys/ncca) which has data available for 2010, 2005/2006 and 2001. The report includes information across four indices for randomly sampled sites to represent the condition of all coastal waters. The Report consists of four indices: Biological Benthic Index, Water Quality Index (Phosphorus, Nitrogen, Dissolved Oxygen, Water Clarity, Chlorophyll a), Sediment Quality Index (contaminants & toxicity), Fish Quality Index. For our water quality layer we use only the Water Quality Index component of the report which measures Phosphorus, Nitrogen, Dissolved Oxygen, Water Clarity, and Chlorophyll a for each monitoring site. The Water Quality Index (WQI) provides data per site in categories of “Good”, “Fair” and “Poor” according to thresholds established by the EPA. Some sites were assigned the category “Missing”. These were removed before analysis.

WQI data was downloaded from the NCCA data portal for the 2010 assessment year. Data for 2005/2006 was not available in a usable format on the portal. Personal communication with people at the EPA led to retrieval of the 2005/2006 information which was emailed to us on August 27, 2018.

**Methods summary:** The data was filtered for sites in the Northeast region, identified by postal code. Each sampling site had an associated weight assigned by the EPA intended to account for how representative the site is for the region. Code was provided by contacts at the EPA to accurately account for these weights. Using the spsurvey R package we applied a categorical data analysis across all sampled sites. This returned an estimate (p) for the percent of the region that site represents, and their assigned category. We then assigned scores to each category (Good = 100, Fair = 50 and Poor = 0), multiplied them by the estimate (p) per site, and then summed up the score per region to get a full picture of water quality in each region.

**Water quality pressure layer** The water quality pressure layer had values for each region and year set equal to the inverse of the clean waters goal layer so that a higher value indicated a lower water quality and therefore a higher pressure.

**Clean Waters goal layer** The water quality layer for the Clean Waters goal had values for each region and year set equal to the sum of site-weighted scores per region.

**Gapfilling:** This dataset was spatially comprehensive covering our entire region, but only contains two time points, 2005/2006 and 2010. The most recent field season was conducted in 2015 but unfortunately the data from that season has not yet been made available. This necessitated temporal gapfilling. Scores for each layer were the same for 2005 and 2006, and then gapfilled linearly by region for the years between 2006 and 2010. Scores for 2011-2017 were set equal to the 2010 scores.

**References:** U.S. Environmental Protection Agency. 2016. National Aquatic Resource Surveys. National Coastal Condition Assessment 2010 (data and metadata files). Available from U.S. EPA web page:<https://www.epa.gov/national-aquatic-resource-surveys/data-national-aquatic-resource-surveys>. Data from the National Aquatic Resource Surveys. Date accessed: 2018-07-31.