



Fish Inventory at Cape Hatteras National Seashore in 2003 - Data Release Report

Issac Quevedo ()

NPS Inventory and Monitoring Division, 1201 Oakridge Dr., Suite 150, Fort Collins, Colorado

The Southeast Coast Network consists of 20 parks within the states of North and South Carolina, Georgia and Florida, all of which are responsible for a large percentage of the fish diversity in the United States. Collectively, these parks would be expected to harbor a high diversity of fishes due to the wide range of habitats represented. At the time of the present study (2001-2005), many basic species inventories for fishes were lacking for many of the parks in the network. The purpose of the study was to sample fishes in 12 parks to supplement existing data or to provide preliminary records of fishes where no records existed. From these data an expected list of fishes for all parks was generated. In 2022, the data was then processed for dissemination by the Inventory and Monitoring Division (IMD), complying with Executive Order 13642, Making Open and Machine Readable the New Default for Government Information. Processed tabular datasets have been created using the data provided from the parks, following Darwin Core naming standards and introducing data quality flagging where data was missing or unclear. The dataset contains 27 columns, 179 observations, and 55 unique species.

Data Records

Data Inputs - Raw Data

Fish inventories were conducted by Carol Johnston and others in 2003 as a part of the National Park Service's Species Inventory program. Investigators performed a combination of field studies and a review of museum records to determine what species were present at CAHA. Original data from their study are available from the NPS DataStore (Johnston et al., 2023).

Summary of Datasets Created

This report refers to the data package Fish Inventory at Cape Hatteras National Seashore in 2003 - Open Format Dataset which contains a metadata file and one data file. The data was processed for dissemination by the Inventory and Monitoring Division (IMD), drawing from raw data provided by the park (cited above). The data file in the package is in CSV format containing observations from a fish survey conducted within the park, along with associated taxonomic information, geographic and temporal coverage, and contextual notes from the study. This package is available at <https://doi.org/10.57830/2295255> (see [Table 1](#)). Column names in the data file largely follow Darwin Core naming standards, with the exception of “custom” names that do not have a corresponding Darwin Core equivalent. Additional fields to hold custom NPS data and data quality flags have been added.

Table 1. Fish Inventory at Cape Hatteras National Seashore in 2003 - Open Format Dataset: List of data files.

File Name	Size	Description
CAHA_Fish_Cleaned_PUBLIC.csv	69 KB	CAHA Fish Observations

Data Quality Evaluation

Data quality was evaluated for fields pertaining to geographic coordinates, taxonomy, and sampling event dates (Table 2). Data quality fields have been appended to the dataset to inform data use and communicate any uncertainty that may exist regarding the quality of data surrounding geographic, taxonomic, and temporal data, respectively. Where possible, data that not meeting data quality standards were corrected.

The data within the data records listed above have been reviewed by staff in the NPS IMD to ensure accuracy, completeness, and consistency with documented data quality standards, as well as for usability and reproducibility (Table 3). Of the data that were evaluated for quality, 99% of fields met data quality standards.

The Fish Inventory at Cape Hatteras National Seashore in 2003 - Open Format Dataset is suitable for its intended use as of the date of processing (2023-04-17).

Table 2. Acceptance criteria for data evaluated.

Field	Acceptance Criteria	Mitigation
eventDate	Sampling event date within the start and end dates of the project	Data rejected
scientificName	Taxon exists within Integrated Taxonomic Information System	Data rejected
coordinate	Sampling location is within the park unit boundaries	Data rejected

Table 3. Summary of data quality flags for each column. [A – Accepted; AE – Accepted but Estimated; P – Provisional; R – Rejected.]

Measure	Number of Records	Accepted		Not Accepted		% Accepted
		A	AE	P	R	
eventDate	179	174	0	0	5	97
scientificName	179	179	0	0	0	100
coordinate	179	179	0	0	0	100

Usage Notes

Users of the data package described in this report will find that column/attribute names follow Simple Darwin Core guidelines (Darwin Core Task Group, 2014). In processing the data provided, an attempt was made to match column names to corresponding Darwin Core terms, defined in the Darwin Core quick reference guide (Darwin Core Maintenance Group, 2021). Those column names that did not fall within a Darwin Core category were named using the format **custom_[original column name]**. In addition, three flagging columns were appended to the dataset at the end using the format **[Fieldname]_flag**.

When known, the circumscriptions of identified taxa were recorded based on publications used to make species identifications. Based on the final project report, three publications were used by Johnston et al. (2006) to identify fish to the species level (Carpenter 2002, Eschmeyer 1998, Nelson et al. 2004, Johnston et al. 2006). However, because the original report did not specifically identify which circumscription was used for any given identification, all three references are noted in **custom_TaxonomicNotes** field. The **namePublishedIn** column was left as “unknown.”

Acquiring the Data Package

This data package is available for download from the NPS DataStore at <https://irma.nps.gov/DataStore/Reference/Profile/2295255>.

This data package can also be directly imported into R data frames using the following code:

```
devtools::install_github("nationalparkservice/NPSutils")
library(NPSutils)
get_data_packages(2295255)
dat <- load_data_package(2295255)
list2env(dat, envir=.GlobalEnv)
```

Methods

Data Collection and Sample Processing Methods

Cape Hatteras is a large (> 14,000 ha) park with a variety of marine and freshwater habitats included within its boundaries. The park was surveyed for fishes in 2003 at 23 sites. All representative habitats were sampled, and fish were identified to the species level. For a detailed description of site locations and sampling methodologies, please see Johnston et al. (2006). Methods varied according to habitat and included boat and backpack electrofishing, snorkeling, and seining (10 - 100' nets, depending on habitat). Most specimens were anesthetized using MS 222, fixed in 10% formalin and transported to Auburn University for identification. Very large specimens were photographed and released. On some occasions species were observed via snorkeling but specimens were not vouchered. At each sampling location, latitude and longitude were taken with a GPS unit along with common locality data.

Additional Data Sources

Threatened and Endangered species were determined using an approved NPS dataset, helping to establish the appropriate release of the data (DeVivo 2023).

Data Processing

Details pertaining to the data processing and quality control cleanup process can be found in the published standard operating procedures (Quevedo and Sherman 2023).

- The original raw files submitted at the completion of the CAHA fish inventory were converted from a data matrix into a Microsoft Excel spreadsheet. Location data were converted to decimal degrees.

- Data were then read into R, where the existing field names were renamed to fit Simple Darwin Core standards (Darwin Core Maintenance Group, 2021).
- Additional Darwin Core columns were appended to fit the Simple Darwin Core Guidelines (Darwin Core Task Group 2014) and to add context to the source data.
- Data quality was evaluated for the coordinate, taxonomy, and date fields and quality flags were added to the end of each dataset (refer to the Data Quality Evaluation section of above).
- To determine whether any data warranted protection from public release, the dataset was evaluated to determine whether any records were observations of threatened or endangered species. No records were identified.
- Finally, scientific names were validated against the Integrated Taxonomic Information System database (?).

Code Availability

The scripts used to create this report and associated datasets are available on the NPS DataStore repository at <https://doi.org/10.57830/2295430> (Quevedo and Patterson 2022).

References

- Carpenter, K. E., ed. 2002. The living marine resources of the Western Central Atlantic. Volumes 1-3. FAO Species Identification Guide for Fishery Purposes and American Society of Ichthyologists and Herpetologists Special Publication No.5. FAO, Rome.
- Darwin Core Maintenance Group. 2021. Darwin Core quick reference guide. Biodiversity Information Standards (TDWG). Available at <https://dwc.tdwg.org/terms/>.
- Darwin Core Task Group. 2014. Simple Darwin Core. Biodiversity Information Standards (TDWG). Available at <https://rs.tdwg.org/dwc/terms/simple/>.
- DeVivo, J. 2023. Federal Listing Status of Taxa in National Park Service Units for use in Determining Appropriate Release of Data (Approved Dataset). <https://doi.org/10.57830/2272462>
- Eschmeyer, W. N., ed. 1998. Catalog of Fishes. Special Publication No.1 of the Center for Biodiversity Research and Information. California Academy of Sciences, San Francisco, California.
- Johnston, C. E., M. Castro, L. Casten, A. Henderson, and A. Kennon. 2006. Fish Inventories at Southeast Coast Network Parks. Auburn University. Auburn, AL. <https://doi.org/10.57830/617584>
- Johnston, C. E., M. Castro, L. Casten, A. Henderson, and A. Kennon 2023. Fish Inventory at Cape Hatteras National Seashore in 2003 - Raw Data. <https://doi.org/10.57830/2295401>
- Nelson, J. S., E. J. Crossman, H. Espinosa-Perez, L. T. Findley, C. R. Gilbert, R. N. Lea, and J. D. Williams. 2004. Common and scientific names of fishes from the United States, Canada, and Mexico. American Fisheries Society, Special Publication 29, Bethesda, Maryland.
- Quevedo, I. and J. Patterson. 2022. Fish Inventory at Cape Lookout National Seashore in 2003-2004 - Processing Script. <https://doi.org/10.57830/2295430>
- Quevedo, I. and A. Sherman. 2023. IMD Inventory Data Cleanup Standard Operating Procedure. <https://doi.org/10.36967/2298851>