GAP Production Data Documentation

Emily Markowitz[1],*, Zack Oyafuso[2],* and Sarah Friedman[2],*

1. NOAA Fisheres Alaska Fisheries Science Center, Groundfish Assessment Program, Bering Sea Survey Team
2. NOAA Fisheries Alaska Fisheries Science Center, Groundfish Assessment Program, Gulf of Alaska and Aleutian Island Survey Team

* *Correspondence:* Emily Markowitz emily.markowitz@noaa.gov * *Correspondence:* Zack Oyafuso zack.oyafuso@noaa.gov * *Correspondence:* Sarah Friedman sarah.friedman@noaa.gov

Table of contents

List of Figures

List of Tables

Part I

Welcome

Please consider this resource to be a **Living Document**. The code in this repository is regularly being updated and improved. Please refer to releases for finalized products and project milestones.

Our Objective

As part of our commitment to open science and transparency, we provide this interactive metadata guide to compliment our public-domain data. Please refer to our Draft Data Changes Brief. Once finalized, this language will be included here.

User Resources

- Groundfish Assessment Program Bottom Trawl Surveys
- AFSC's Resource Assessment and Conservation Engineering Division
- Survey code books
- Publications and Data Reports
- Research Surveys conducted at AFSC

Background of the gap_products repo

This work is the result of the massive efforts of three concurrent GAP working groups:

1) Index Computation Working Group: consolidation of the methods used to produced design-based estimates of abundance and size/age composition between the Bering Sea and AIGOA survey regions.

2) Data Processes Working Group: reorganization of the Oracle data infrastructure that houses the standard data products produced by GAP.

3) Gulf of Alaska Survey Restratification Working Group: implementation of a new stratified random survey design in the Gulf of Alaska bottom trawl survey.

We began this effort in collaboration with the Status of Stocks team (SSMA) to present both the orientation and opportunity to interact with Gulf of Alaska data from the restratified survey design that we will be implementing in the 2025 field season. As that

part of the project evolved, the Data Processes Working Group identified the opportunity and need for gaining efficiencies by redesigning and consolidating the Oracle objects (tables and materialized views) that have historically served these data. The Index Computation Working Group also identified an opportunity to gain efficiencies by consolidating the various scripts that were developed independently by both survey region groups into a workflow that was more accessible and documented.

The Index Computation Working Group developed the gapindex R package, a code repository that consolidates the code that calculates the various standard GAP products (e.g., CPUE, total biomass, size/age composition) for both the Bering Sea and AIGOA survey regions. The Data Processes Working Group was responsible for compiling the data structures needed to support data product tables that were consistent across all of the AFSC GAP survey regions as well as the creation of the GAP_PRODUCTS oracle schema that will house these consolidated products in the future.

This gap_products GitHub repository houses the code that will conduct the "standard production run" that produces the new data tables via the gapindex R package and upload those tables to the GAP_PRODUCTS Oracle schema.

Major Advantages

- Consolidated production tables include all standard data products for all surveys. Data will be provided in the same format, with the same units, and created using the same mathematical methodology. This should limit data pulls, reduce complexity for data access, and reduce complicated secondary data wrangling.

- Consistent naming conventions for schemata, tables, and column metadata. Columns across all tables will use the same naming conventions, units, and data types. Restricting standard data product table content to absolutely necessary columns.

- Removal of redundant data columns that can be acquired by joining to reference tables is key for providing consistent and up-to-date data while limiting data table sizes.

- Consolidation and repurposing of Oracle schemata. This will help the GAP team limit unnecessary access to unprocessed or problematic data by outside users.

- Vetted data methods. All code and data inclusion decisions and wrangling are documented in the {gapindex} R package. Streamlined and rapid data production. Improved and consolidated data creation and documentation provide data

Major Advantages

creators and users with greater confidence in the data products and enhanced ability to share the data. 3-04/750x500-bottom-trawl-survey-afsc.jpg)