# Automated and operational access to environmental data for Alaska’s management areas

Jordan T Watson and Matthew W Callahan

# Abstract

The proliferation of operational satellite data has facilitated downstream data products catered towards specific fisheries applications in near real-time. We utilized this data accessibility to connect a suite of fishery-dependent data with spatially-explicit environmental information in the backend of the Alaska Region database environment at the Alaska Fisheries Information Network (AKFIN). For example, sea surface temperature (SST) data were linked to all fish tickets and observer in the Oracle backend from 2002 - present (more than one million records), and new data are automatically matched each day. We further extended the utility of satellite data products through customized spatial clipping of gridded satellite data extents to regions of interest for Alaska fisheries management. Full gridded data sets stratified by Alaska management and research shapefile polygons can be queried from the AKFIN database. Alternatively, aggregated data products (e.g., time series of SST for individual NMFS regions or ecosystem areas) can be accessed via custom web services, or URL-based data queries. We demonstrate several queries of the web service and illustrate how this product can yield seamless integration with downstream analyses by detecting marine heatwaves in the Eastern Bering Sea ecosystem region.

# Introduction

The role of oceanographic and other environmental or ecosystem parameters on the productivity of the world’s fish stocks has long been established. Increasingly, such parameters are explicitly incorporated into fisheries stock assessments ([**Holsman2016?**](#ref-Holsman2016), [**Marshall2019?**](#ref-Marshall2019)), risk assessments ([**Gaichas2014?**](#ref-Gaichas2014)); ecosystem reports ([**Ferriss2020?**](#ref-Ferriss2020), [**Ortiz2020?**](#ref-Ortiz2020), [**Siddon2020?**](#ref-Siddon2020)), or other documents used by the U.S. Regional Fishery Management Councils to guide decision making. Meanwhile, a growing trend in the development of dynamic ocean management tools seeks to incorporate environmental information in near real-time to inform stakeholders for bycatch avoidance ([**Hazen2018?**](#ref-Hazen2018), [**Breece2021?**](#ref-Breece2021)), harmful algal blooms ([**Harley2020?**](#ref-Harley2020)), avoiding interactions with protected species (<https://oceanview.pfeg.noaa.gov/whale_indices/>), and more. Thus, as NOAA moves towards a broader adoption of ecosystem-based fisheries management and dynamic ocean management, the accessibility of ecosystem information becomes increasingly critical.

One of the most fundamental ecosystem parameters considered in fisheries is water temperature. Temperature regulates the timing and intensity of primary production, which has ripple effects on secondary producers and on to higher trophic levels. Temperature directly impacts fish growth and other metabolic processes in addition to regulating the location and abundance of prey. Thus, for most mobile fish species, temperature often defines the habitat of the species, and subsequently, the location of the fishing fleets that target them ([**Haynie2012?**](#ref-Haynie2012), [**Watson2018?**](#ref-Watson2018), [**Rogers2019?**](#ref-Rogers2019)).

As global climate changes, water temperatures have been among the most easily measured metrics by which to understand how ocean ecosystems are responding. Broad warming trends are leading to poleward shifts in the distributions of fish species and the fleets that target them ([**Kotwicki2013?**](#ref-Kotwicki2013), [**Rogers2019?**](#ref-Rogers2019), [**Pinsky2020?**](#ref-Pinsky2020), [**Fredston2021?**](#ref-Fredston2021)), while anomalously warm periods or marine heatwaves are driving protracted impacts on ecosystems ([**Suryan2021?**](#ref-Suryan2021)) and commercial fish stocks ([**Barbeaux2020?**](#ref-Barbeaux2020)). Such dynamics underscore the need for reliable access to near real-time water temperature data.

Satellite-derived sea surface temperature data have been available since the early 1980s and a proliferation of new technologies, sensors, and data products have led to increasingly frequent and spatially resolved information with latencies as little as one day (Liu et al., 2015; Maturi et al., 2017; Minnett et al., 2019). Moreover, the development of programs like NOAA’s CoastWatch and data technologies like Environmental Research Division’s data access program (ERDDAP) servers (Simons 2020) have facilitated easier access to these data worldwide in near real-time and via a suite of data formats. While such technologies have improved data access, challenges still exist for some end users due to the large file sizes of high spatial and temporal resolution data sets, difficulty subsetting data within irregular polygons (custom spatial strata), and the need for data infrastructure that supports operationalization and automation of data ingestion (Welch et al., 2019).

After assessing the needs of a suite of fisheries biology, stock assessment, and socio-ecological modeling efforts at the Alaska Fisheries Science Center (NMFS-NOAA), we developed an automated and operational framework for serving satellite environmental data products for a suite of spatial strata used for fisheries management and research in Alaska. The framework we present uses daily sea surface temperature data but can easily be extended to other environmental data products like chlorophyll, wind, ROMS model extractions, or other data identified by stakeholders. We describe the data used, the process for joining the data to spatial strata, backend database merges with fishery dependent data (e.g. observer and fish ticket data), and data access through customized web services (data queries via URL).

# Methods & Results

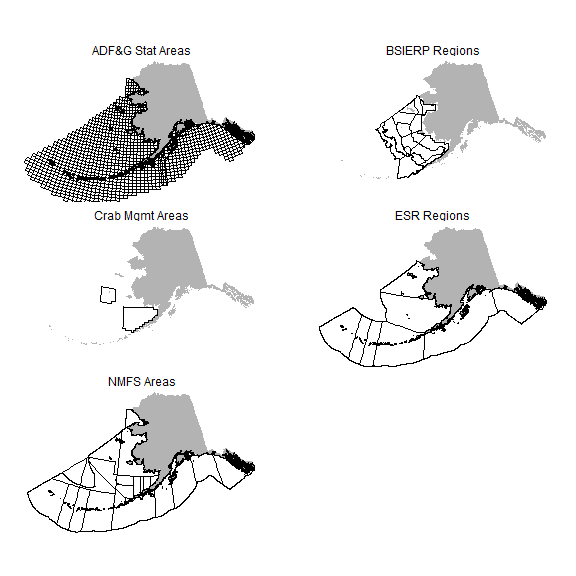
*Satellite data*  
For this study, two daily satellite sea surface temperature (SST) products were used. In both cases, data were accessed via NOAA ERDDAP servers ([Simons 2020](#ref-Simons2020)) and downloaded as netCDF files within the Oracle database backend at the Alaska Fisheries Information Network (AKFIN), maintained by the Pacific States Marine Fisheries Commission. The SST data are publicly available but by ingesting them into the AKFIN backend, they can be seamlessly merged, behind the NOAA firewall, with confidential fishery-dependent data sets like observer data, vessel monitoring system (VMS) data, and fish tickets. Both of the SST products provide gap-free data each day. The MUR SST data set is provided by JPL NASA (JPL MUR MEaSUREs Project. 2015) and is available from June 2002 - present and are accessed via the NOAA CoastWatch West Coast Node ERDDAP server (<https://coastwatch.pfeg.noaa.gov/erddap/griddap/jplMURSST41.html>). These data are provided across a 0.01° x 0.01° (1 km) spatial grid. Meanwhile, the CRW SST data set covers a 0.05° x 0.05° (5 km) spatial grid and these data are obtained from the NOAA Coral Reef Watch Program (<https://coralreefwatch.noaa.gov/product/5km/index_5km_sst.php>) through the NOAA PacIOOS Program ERDDAP server (pae-paha.pacioos.hawaii.edu/erddap/griddap/dhw\_5km.html) from April 1985 – present. Additional data (January – March 1985) were downloaded via public ftp link from NESDIS (<ftp://ftp.star.nesdis.noaa.gov/pub/socd/mecb/crw/data/coraltemp/v1.0/nc/1985/>). Both the MUR and CRW data sets typically have a 1-2 day latency period.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Data set | Spatial resolution | Duration | Gap-free | AKFIN spatial strata | AKFIN Web service | Number of daily data points within Alaska grid | Fishery-dependent data sets linked in AKFIN data base | Citation | Online source |
| Multi-ultra resolution (MUR) | 0.01 degrees | 2002 - present | X | X |  | ~1 million | Observer, VMS, Fish tickets | JPL MUR MEaSUREs Project. 2015 | [Link](https://coastwatch.pfeg.noaa.gov/erddap/griddap/jplMURSST41.html) |
| CRW SST | 0.05 degrees | 1985 - present | X | X | X | 212k | Not sure | Liu et al., 2015 | [Link](https://pae-paha.pacioos.hawaii.edu/erddap/griddap/dhw_5km.html) |
| OI SST | 0.25 degrees | 1982 - present | X | X |  | ~42k |  | Reynolds et al. 2007 | [Link](https://coastwatch.pfeg.noaa.gov/erddap/griddap/ncdcOisst2Agg.html) |

Both the MUR and the CRW data sets have native formats with longitudes ranging from -180 to +180. Because the spatial extent for Alaska waters includes the International Date Line, the daily data are downloaded via two separate operations each day. One operation downloads the negative longitude data from 46°N to 68.8°N and -180°E to -130°E and the second operation downloads the positive longitude data from 47.5°N to 60.0°N and 164°E to 180°E. These downloads are merged and then clipped to spatial regions of interest within the exclusive economic zone surrounding Alaska, yielding 212,813 SST records per day.

*Spatial strata* State and Federal waters of Alaska include numerous spatial strata that are relevant to fisheries management, ecology, and individual species distributions. For example, the Alaska Department of Fish & Game (ADF&G) divides Alaskan waters into nearly 1,800 statistical areas, many of which are 0.5° latitude by 1.0° longitude boxes. Meanwhile, the National Marine Fisheries Service (NMFS) divides the same waters into only 25 management areas. However, these regulatory strata are inconsistent with ecological stratifications (Eastern Bering Sea, Gulf of Alaska, and the Aleutian Islands) identified for the same waters. These ecosystem regions, even when subdivided, do not necessarily align with spatial strata identified for individual fish or crab stocks, so stock assessment scientists and fishery managers are often interested in yet further customized spatial boundaries. Thus it is not surprising that different users of environmental information like SST may want those data aggregated or clipped to a different (or multiple) spatial boundaries.

To develop operational data products across Alaska’s suite of spatial strata, we undertook extensive point-in-polygon geoprocessing operations to apportion the individual latitude-longitude coordinates for both the MUR and CRW SST spatial grids to each of the polygons from a suite of shapefiles (ADF&G management areas, NMFS management areas, Ecosystem regions [from NMFS Ecosystem Status Reports], Bering Sea Integrated Ecosystem Research Program [BSIERP] regions, Bristol Bay red king crab management areas, St. Matthew blue king crab management areas) (Fig. 1). The spatial extent of Alaska includes more than 200,000 data records daily for the CRW data set and more than 1 million records daily for the MUR data set. To avoid repeating the computationally intensive point-in-polygon operations, we created spatial lookup tables that are stored in the backend of the AKFIN Oracle database system. Thus, as data are downloaded daily from ERDDAP servers across the spatial extent of Alaska’s waters, each SST record is matched via a database join on latitude and longitude to the spatial strata in which it falls instead of via repeated point-in-polygon operations.

 Figure 1. Spatial strata in Alaska for which sea surface temperature data have been clipped and aggregated within the AKFIN database backend. SST data for these strata can be queried and accessed several ways.

*Accessing the data* Data ingested into AKFIN can be accessed and used for operational workflows via serveral different methods (Figure 2). We demonstrate two general methods for accessing the data stored in AKFIN. The first method, customized web services (web APIs), is ideal for accessing time series of aggregated data (e.g. daily SST averaged across a spatial stratum or multiple spatial strata) and for queries less than about 100,000 records. This approach leverages a simplified data access point (url) that is outside of the AKFIN firewall and requires no user login. The second method, direct database access, requires a NOAA network connection (local or VPN) and a login to the AKFIN database backend. This approach relies on SQL to extract either aggregated data summaries or larger gridded data sets (e.g., millions of data records). In the sections that follow, we demonstrate data queries using custom web services and by using direct SQL and R access. For each case, we illustrate the utility of operational data workflows by piping these data into R functions for calculating marine heatwaves (MHWs).

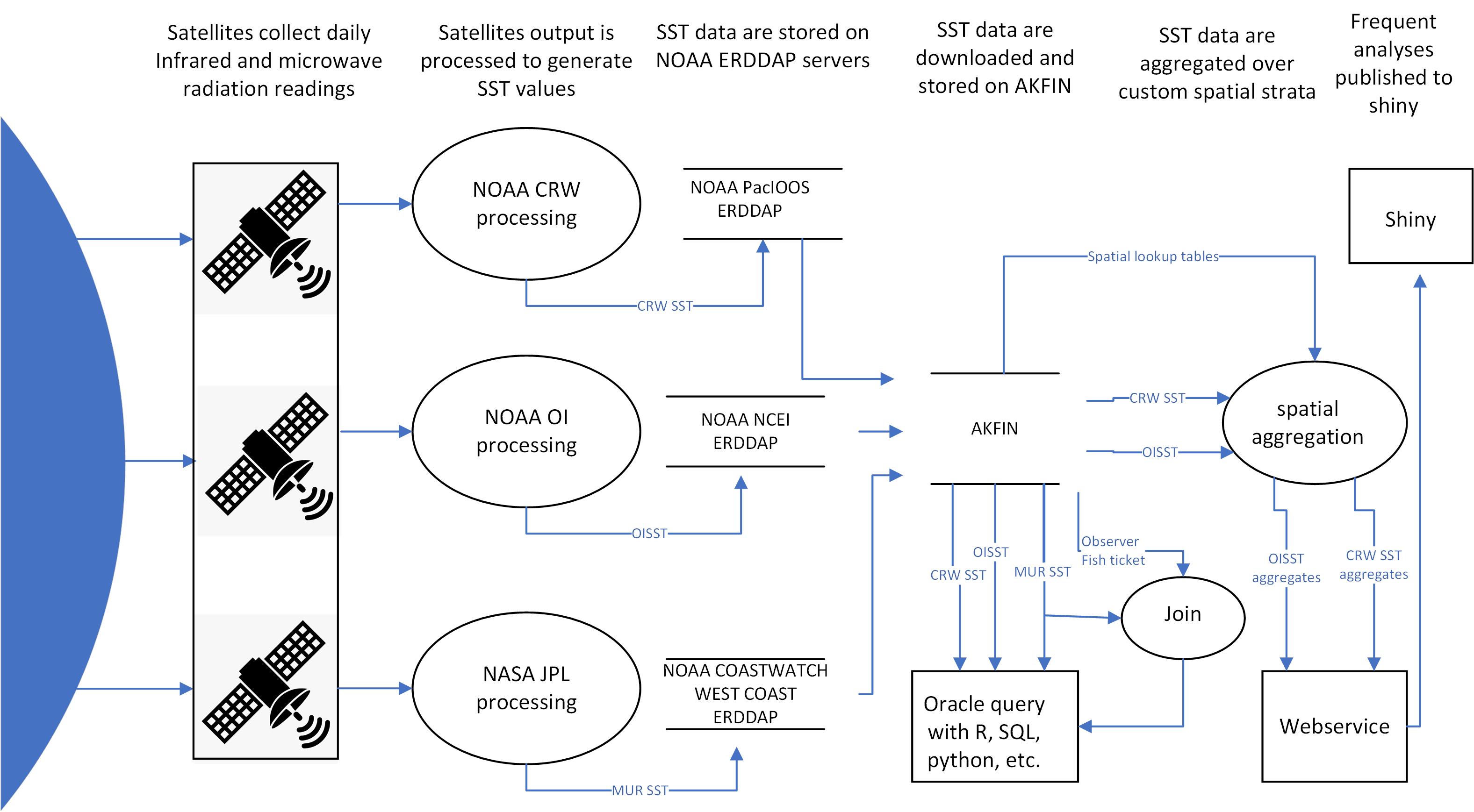


Figure 2: Data flow diagram for the ingestion, processing, and extraction of satellite sea surface temperature data within AKFIN.

*Customized Web Service (Web API)* For queries that are likely to be repeated often or to become part of an automated process, customized web services offer a particularly efficient data access option. These web services require no accounts, no passwords, no VPN - just internet. A web service URL can be readily embedded into programming applications (e.g., R, Python) and for multiple or different spatial strata as well as for specified time series. An additional convenience is that web services allow users to query time series without storing data locally which is particularly helpful for operations that would typically append data to existing files (e.g., adding a new day of data to a time series). With this tool, users can easily incorporate SST data into stock assessments and other processes.

The AKFIN web service web service enables a query of CRW SST using a URL, where the URL contains the query parameters (Fig. 3). For example, in the URL “<https://apex.psmfc.org/akfin/data_marts/akmp/nmfs_area_crw_avg_sst?nmfs_area=640>,” where “nmfs\_area\_crw\_avg\_sst?” is the name of the dataset. This is the daily SST data averaged by **nmfs\_area**. A “?” separates the data set name from the query criteria. To query multiple areas, separate the values by a comma. Large Marine Ecosystem subregions are assessed in AFSC Ecosystem Status Reports (ESRs). The **ecosystem\_sub** fields available for query include the regions within the Eastern Bering Sea, Aleutian Islands, and Gulf of Alaska. Spaces in region names are filled with “%20.” For example, to query the data for the “Southeastern Bering Sea,” for example, add “ecosystem\_sub=Southeastern%20Bering%20Sea.” For the Bering Sea and Gulf of Alaska, the query filters only data where water depth is between 10 and 200m. For the Aleutian Islands, a depth filter is not implemented. Analysts that are interested in data for different depth ranges, custom spatial bounds, or aggregated NMFS areas can contact the authors and we will arrange for your request.

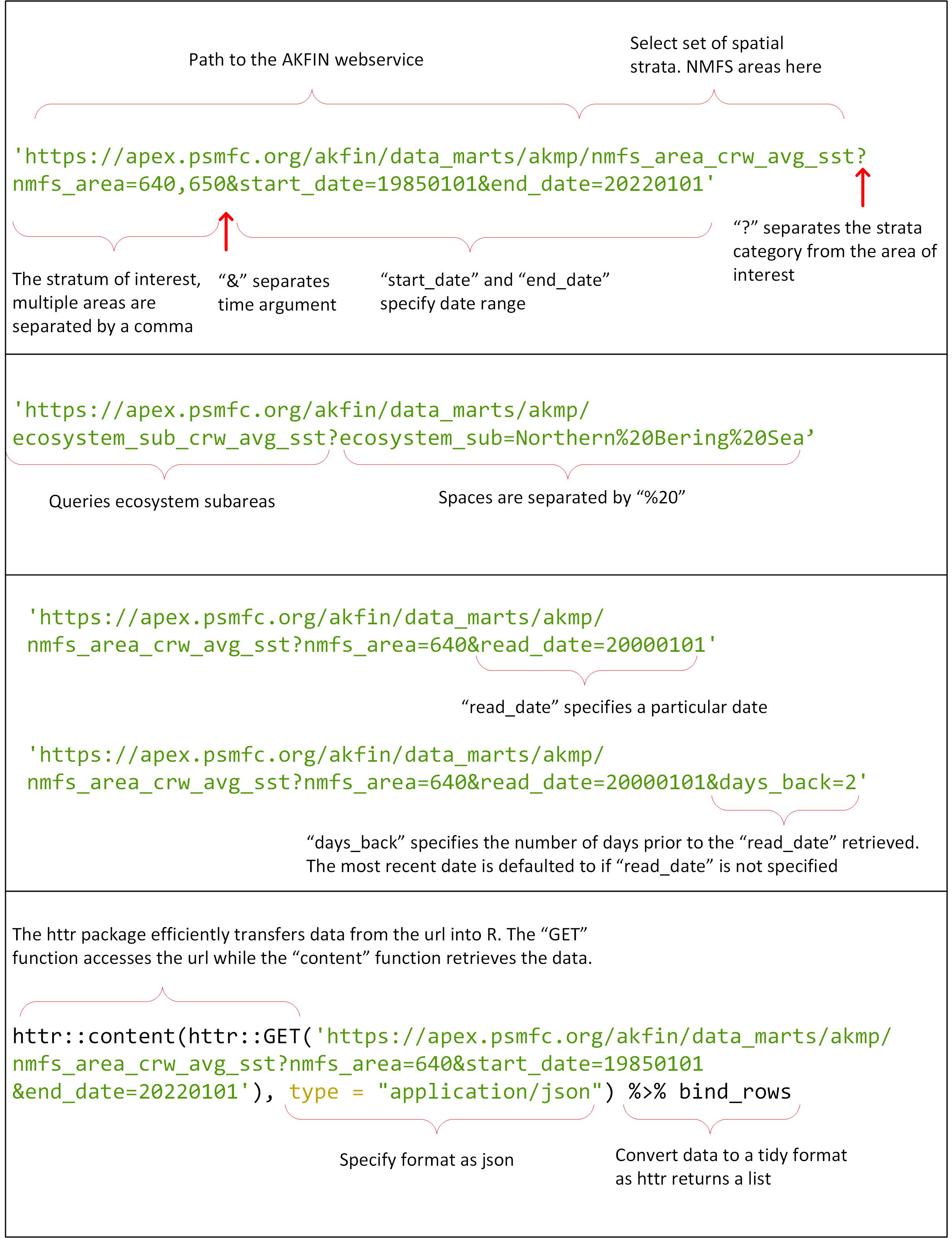


Figure 3: Configuration of AKFIN web service URL parameters... (will elaborate)

To add a time component to a query, specify “start\_date” and “end\_date,” “read\_date,” or “dates\_back” parameters. If no time argument is included, the default behavior is to pull the single most recent datum record. Time parameters should be included after spatial parameters and separated by an “&.” Most users will want the entire time series, which starts on 1985-01-01. To query the entire time series, specify “start\_date” & “end\_date.” “end\_date” must be included, but if you do not know the most recent date of the time series, you can choose an end date some time in the future and it will query all of the data that exist. The full time series yields more than 13,000 rows of data per area (i.e., daily data from 1985-01-01 to present). Dates are queried in the “yyyymmdd” format, for example “19871214” queries December 14, 1987. The “read\_date” argument retrieves data from any date in the time series, however, it is necessary to query the date after the desired day. The web service date format contains a time component, which is set to 12:00:00Z for each day, SST records were created after that time stamp, causing queries to return values for the previous date. Finally, a “days\_back” parameter specification allows users to query any number of days prior a date of interest. If “read\_date” is not specified, “days\_back” returns the most recent SSTs. *Plan to inlcude the EGOA and WAI time series image or something*

To access web services using R Statistical Software, we recommend the R package **httr** to pull data from a URL (Fig. 3), and the data format should be specified as json. The data can be saved as an object for manipulation or piped directly into downstream functions. Additional packages **tidyverse** and **lubridate** are recommended for plotting and manipulation but the object retrieved using **httr** could be manipulated using base R instead. Simply pasting the URL into a web browser would also display fetched data in that browser.

**Piping web service queries into marine heatwave calculations** *Plan to include the 2019 NBS marine heatwave figure*

With the daily time series for a spatial stratum, implementation of the heatwaveR package ([**W.Schlegel2018?**](#ref-W.Schlegel2018)) is straightforward. We encourage readers to explore the functionality described in the heatwaveR vignettes, from which the following examples are generated . We demonstrate a few simple examples below using web services to query the SST time series (Supplement). To better illustrate the heatwave categories here, we cherry-pick an example from the Northern Bering Sea from 2019-01-01 to 2019-12-31 (Figure 5).

**Oracle database queries** *Haven’t worked this section yet* Some data users prefer the flexibility and transparency of querying raw gridded data directly from the Oracle server. This is particularly useful for larger queries (e.g., millions of records) or for exploring data across a suite of different spatial extents (e.g., custom depth ranges, shapefiles, etc.). As we note in the web services section, such custom queries can also be automated by working with Alaska Fisheries Science Center (AFSC) and AKFIN staff. However, for users that prefer to code their own queries, we provide several examples here. Notably, to query directly from the database, users will need an AKFIN database account, which can be provided by contacting the authors of this document.

This section is not meant to serve as a SQL tutorial. Rather, its purpose is to orient users to the structure of the database related to the SST data and lookup tables. We assume that users interested in querying the database directly via Oracle (e.g., SQL Developer) or through odbc connections from R or Python are already acquainted with the coding and configuration settings. However, interested users can contact the authors for assistance establishing such connections or custom SQL queries.

The gridded SST data (Fig. BB) are stored within the AFSC schema on the AKFIN Oracle database and the primary key linking the lookup tables (Fig. AA) with the gridded data is the ID field. In the lookup table, it is simply “ID” and in the data table it is “CRW\_ID.”

CRW SST query within SQL. Several columns reveal ‘NA’ because the particular latitude - longitude coordinates shown do not fall within any spatial strata represented by those columns.

The following query demonstrates the primary key relationship between the data and lookup tables. In this case, we query SST (“TEMP”) data that fall within a crab management area and we add a field for “Year” (Fig. CC).

CRW SST query for records that fall within a crab management area.

Plotted query of Bristol Bay crab management area SST data averaged daily and plotted with default smoothing.

**Matching SST data with fishery-dependent data** *I think this is fine in methods/results. This is MUR SST right? We should probably have MUR somewhere later on in the methods* The above sections demonstrate access to gridded or raw SST data that are updated daily within AKFIN. In addition to raw SST data access, the daily SST data are also integrated within the AKFIN back-end to observer and fish ticket data. For both of these fishery-dependent data sets, the MUR SST data have been used, and users with access to these confidential data sets through AKFIN can find an SST field in the comprehensive\_ft and comprehensive\_haul (check table names) tables.

Observer data include latitudes and longitudes of gear deployment and retrieval locations, which are matched with the nearest gridded SST data for a given date. The temperature data are then averaged across the retrieval and deploy points to yield a single SST datum for each observed fishing event.

Fish ticket spatial data are recorded at the scale of ADF&G statistical areas (typically 0.5 degree latitude x 1.0 degree longitude), so gridded SST data cannot be matched as directly. Instead, daily SST data for all gridded locations within each statistical area (N=1758) are averaged, to yield a single daily datum for each of the statistical areas. These daily average data are then matched with the reported statistical areas on fish tickets based on the date that fishing was reported to have begun within a particular statistical area.

# Discussion *didn’t go over this*

The ability to integrate environmental and fishery data sets in near real-time is fundamental to an increasing number of fishery management priorities. However, creating automated database infrastructure is beyond the expertise of most users of such data. Working with AKFIN programmers, we developed a back-end database infrastructure that automatically clips SST data to areas of interest identified by a suite of end users at the AFSC. These data can then be accessed either in gridded form, using direct database queries, or in aggregate form, using customized web services, or APIs.

The options we present each have advantages and disadvantages. The web services allow users simple and seamless access to data through a URL, which requires no login or password. As we have demonstrated, these web services can be easily incorporated into workflows to support operational data applications, like R Shiny Apps. However, each web service URL is based on a backend SQL query that must be pre-meditated and coded by programmers. So, while the end-users do not need to code any database queries, a programmer does. Meanwhile, direct database access requires a VPN connection and a login to the AKFIN database, but once users have established this connection, they can customize any SQL queries they want using either direct Oracle access or ODBC connections through R, Python, or other data access points. This puts total control into the hands of the end-user. The goal with these combined approaches is to serve a suite of users and applications across a range of complexities of data tasks.

This document is meant to serve two primary purposes. The first is to demonstrate the functionality and access to existing environmental data products within AKFIN. The second is to give end users a sense of the types of data products and access approaches that can be requested and implemented within AKFIN. The spatial extents, satellite data sets, and web service queries demonstrated here were chosen based on previous requests or needs from individual data users at the AFSC. While additional data product development is underway for satellite-based chlorophyll, ROMS-based bottom temperatures and heatwaves, Pacific-wide vessel monitoring data, machine learning model outputs, and more, the authors of this study are keen to work with end-users and AKFIN staff to connect additional data needs with AFSC end-users. Thus, we encourage data users to contact us to discuss data access, automation, and operationalization needs and interests.

# References

Simons B. 2020. ERDDAP.