CHAPTER 1

Cloudy with a Chance of Fish: ArcGIS for Server and Cloud-Based Fisheries Oceanography Applications

Tiffany C. Vance, Stephen Sontag, and Kyle Wilcox

Abstract

Wright et al. (2013) defined cyber-GIS as "GIS detached from the desktop and deployed on the web." This chapter describes two cyber-GIS applications—the Ecosystems Fisheries Oceanography Coordinated Investigations Data Access and Analysis Tool and LarvaMap—developed for data integration and particle modeling in support of fisheries oceanography research and fisheries management. Both were developed to make it easier for researchers to locate, link, analyze, and display in situ and model data. Because geospatial/temporal relationships are critical in understanding the early life history of commercially valuable fish species such as walleye pollock (*Gadus chalcogrammus*), integrating the generation of model results and in situ environmental data using GIS provides critically needed tools.

The Ecosystems Fisheries Oceanography Coordinated Investigations Data Access and Analysis Tool uses an ArcGIS for Server front end for selection, display, and analysis of data stored in an Oracle database. LarvaMap is a cloud-based tool for running particle dispersion models for fish and invertebrate larvae. It supports scenario testing and makes it easy for nonmodelers to configure and run models. The front end is a web-based map interface, and running the tracking model in a cloud resource allows scaling of computer resources to meet computational needs. Both applications provide basic display and analysis of data and can be integrated with desktop GIS for advanced analyses and two-, three-, and four-dimensional visualizations. The results from these projects can be translated to other types of particle tracking, other data management needs for environmental studies, and used as case studies of deployment of GIS tools to cyber-GIS computing resources.

Problem and Challenge

Knowledge of the complex relationships between the targeted fish species and its food sources, predators, and environment will aid the new emphasis on ecosystem-based approaches to fisheries management (Link 2002). Understanding these relationships can guide the management of a fishery and the setting of quotas, timing of fishing periods, and management of related fisheries or species that depend on the target species. Investigating these relationships requires a combination of field research and computer modeling. Field research gathers large amounts of disparate data that needs to be managed, integrated, analyzed, and visualized. Many times, the only common characteristic of this data is the fact that it can be geolocated. Spatial analyses and geospatial measures are critical in understanding this data. Models of ocean circulation, fish life histories, or complicated ecological interactions can require huge amounts of computing resources, generate massive output files, and be best understood using advanced visualization and analysis tools.

Providing the computational and analytical environment to support these studies is challenging. Local desktop resources are frequently insufficient. Tightly focused databases can make integrating information almost impossible, and visualization tools may be hard to use. New developments in cyber-GIS, specifically the expansion of ArcGIS for Server and increased access to cloud computing resources, have enabled us to develop and test a number of applications to address solutions for integrating disparate data for easy retrieval and analysis using ArcGIS for Server, either locally or in the cloud, and explore the possibilities of running models in the cloud and storing and displaying their results in a geospatially enabled application.

Introduction and Background

Parallel technological advances in computing and marine fisheries research have contributed greatly to truly implementing ecosystem approaches to fisheries management. Technologies first developed during World War II have been used both to improve sampling in marine systems and support modeling and geospatial analyses. Military surplus vessels formed the core of the oceanographic and fisheries research fleet in the 1950s and onward (Rainger 2000; Rozwadowski and Van Keuren 2004; Hamblin 2005); sound navigation and ranging (SONAR) techniques were adapted to become the core of fisheries hydroacoustics (Fornshell and Tesei 2013); and biological sampling became increasingly electronic (Dickey 2003; Wiebe and Benfield 2003). Computing resources evolved from large mainframe computers to minicomputers and personal computers, to computers on a chip embedded in deployed instruments, and finally back to large centralized, elastic computing resources in the cloud. Geospatial analysis has gone from the early days of GIS in SyMAP and the Canada Geographic Information System (Coppock and Rhind 1991) to ArcInfo to open-source GIS, cyber-GIS, and ArcGIS for Server providing GIS in the cloud.

GIS and other spatial analysis tools provide important tools to understand natural systems. The advent of cyber-GIS has made these tools even more powerful by removing constraints on the size and complexity of analyses through elastic computing resources and cloud-based storage for terabytes of data. Of course, what has been called the "grand challenges" remain, of data intensity, computing intensity, concurrent intensity, and spatiotemporal intensity (Yang et al. 2011). The data integration and complex system modeling tools we have created with the Ecosystems Fisheries Oceanography Coordinated Investigations (EcoFOCI) Data Access and Analysis Tool (EcoDAAT) and LarvaMap take advantage of early answers to these challenges.

Wright et al. (2013) have written of cyber-GIS as GIS removed from the desktop to centralized, usually cloud-based resources. Wang et al. (2013) describe cyber-GIS as "cyberinfrastructure-based GIS" and a subset of the broader realm of "spatial cyberinfrastructure" as defined in Wright and Wang (2011). Yang et al. (2013) have identified a number of disciplines in which cyber-GIS can support critical advances. These include topics related to fisheries such as climate science, ecology, environmental health, and disaster management. They call for advancements in system architectures, which could directly support modeling and visualization and are key to understanding the outputs of complicated ecosystem models; data storage, which would support both the storage of field data and outputs of models; and data and process colocation, which may be less critical because of developments such as the grid caching we developed for LarvaMap which is described later in this chapter (Yang et al. 2013).

Nyerges (2010) specifically addresses the application of cyber-GIS to marine systems by looking at uses of cyber-GIS for modeling coastal marine policy and governance. Although his focus is on human communities, many of the elements he cites are equally applicable to marine ecosystems that also have a "large number of interacting components," creating systems that are "computationally complex and intense" with spatiotemporal elements that "require tremendous amounts of data and computational support." EcoDAAT uses the cyber-GIS concepts of moving processing and data storage from the desktop to a centralized resource with a web-based thin client providing the user interface. The results can be used either for analyses via ArcGIS for Server or for local GIS or statistical analyses using ArcGIS or tools such as R. LarvaMap takes advantage of cloud-based computing resources by running its models on a cloud computing resource. The ability to launch additional model runs in the cloud provides easily scalable computer resources, and data can be stored either in the cloud or in local Open-Source Project for a Network Data Access Protocol (OPeNDAP)/Thematic Real-Time Environmental Distributed Data Services (THREDDS) servers. Both of these tools help meet the ecosystem-based management needs of integrating field data with models, integrating physical and biological data, and providing information on the early life history of commercially important species.

History and Status of the Walleye Pollock Fishery in Alaska

EcoDAAT and LarvaMap are first being applied to studies of the early life of walleye pollock (*Gadus chalcogrammus*), hereafter simply *pollock* (figure 1.1). Pollock are a groundfish found primarily in the north Pacific and Bering Sea that supports the largest commercial fishery in the United States by weight (Kendall et al. 1996; Alaska Fisheries Science Center 2013). The primary human uses of pollock are in making fish sticks and fish sandwiches, and as surimi to make imitation crab. Juvenile pollock are a critical food source for flatfish (Lang et al. 2000), and adult seabirds serve as food for Pacific cod, Pacific halibut, sea lions, and other marine mammals.



Figure 1.1. Walleye pollock larvae, adults, the products made from pollock, and major predators of pollock. By Ingrid Spies and other photographers from the Alaska Fisheries Science Center, NOAA Fisheries Service; additional photograph by Nazila Merati.

In the Gulf of Alaska, pollock start as eggs spawned in March and April. The eggs are spawned at a depth of 100–250 m and hatch in two to three weeks. Larvae occur at an average depth of 40 m and develop over the next few months. At this stage, they are still at the mercy of oceanic currents and transported as though they were inanimate particles. Once the larvae are about 60 days old, they develop rudimentary swimming skills. Though not able to swim long distances, they are able to swim in bursts to consume prey and avoid potential predators. After one hundred days, they are competent swimmers and can move up and down thorough the water column. At the age of three or four years, they are large enough to be commercially harvested and reach sexual maturity at age four or five. They can live for more than 20 years, but fish that old are unusual (AFSC 2010).

The Gulf of Alaska and Bering Sea pollock fishery started with a small trawl fishery in the 1950s, and catches increased rapidly in the 1970s and 1980s as the foreign fishing fleet targeted the population. The discovery of a spawning population of the fish in Shelikof Strait in 1981 led to the development of a roe fishery. Intense fishing in the international waters of the Donut Hole area of the western Bering Sea in the 1980s led to the collapse of that pollock population (Bailey 2013).

By 1988, changes in fisheries regulations excluded the foreign fleet from American waters and transformed the pollock fishery into a domestic US fishery with the primary consumer products being frozen filets and surimi. In the 1990s, the rise of ecosystem-based management led to consideration of related species in managing the fishery, and in 1999, the Aleutian Islands portion of the fishery was closed in response to declines in Stellar sea lion populations, with the thought that fishing was outcompeting sea lions for fish (National Research Council 2003). This fishery was reopened in 2005 (AFSC 2010). A complete history of the fishery, both domestic and international, can be found in Bailey (2013).

Management of the fishery depends on, among other things, an understanding of the early life history of pollock. Success in the early stages of life is one factor determining how many fish are eventually available to be harvested while preserving a sustainable population. The EcoFOCI program at the National Oceanic and Atmospheric Administration (NOAA) studies the factors that affect the variability in recruitment of pollock and other commercially harvested species in the Gulf of Alaska, north Pacific, and Bering Sea. Oceanographic and biological factors are studied using a combination of field and model studies. Field research includes physical oceanographic surveys and biological sampling to look at life stages from eggs to juveniles and post-age-1 fish. A particular emphasis of the sampling is studying pollock early life history. The zooplankton that support the fish populations are also sampled and described. The samples collected also support age and growth, genetic, and food habit studies. The field data supports an understanding of the biological and physical factors affecting recruitment. The data also provides initial conditions and validation for model studies.

Understanding the Early Life History of Pollock Using Individual-Based Models

One of the "grand challenges" in fisheries research is understanding the fluctuations in the size and structure of fish stocks. The classic work by Hjort (1914) described two hypotheses to explain the variability: The first, the Critical Period hypothesis, posits that the size and strength of a year class of fish is determined very early in life, when the fish have exhausted their yolk sack and started feeding, and is a function of the availability and quality of the plankton they are feeding on. The second, the Aberrant Drift hypothesis, examines the role of transport in taking fish larvae so far from place of birth that they are unable to return to support future populations. A variety of hypotheses have developed in the century since Hjort's work, including the Match-Mismatch hypothesis of Cushing (1974), which relates the timing of larvae hatching and bloom of plankton to feed them; the effect of vertical stratification in the Stable Ocean hypothesis of Lasker (1978); and a number of other hypotheses described and related in Houde (2008) and Hare (2014).

In studies of marine fish recruitment, early studies to test hypotheses were primarily descriptive and considered aspects such as the scale of distributions of plankton (Haury et al. 1978). More recently, advances in the collection of quasi-synoptic in situ and satellite data, the rise of ecosystem approaches to studying marine systems (Levin 2006; Vance and Doel 2010), and rapid and striking advances in computers and data storage (Gentleman 2002; Yang et al. 2011) have radically changed research on recruitment. Hydrodynamic models are now able to resolve details at the scales that affect recruitment (tens of kilometers), and they can be run for weeks to months of model time to cover the early life history of slowly developing species. Biological and physical processes can now be coupled and modeled, and advances in visualization tools have made visualizing and interpreting large model outputs easier (Megrey et al. 2002; Vance et al. 2006; Vance 2008).

One requirement these various hypotheses share is a need to know where fish larvae are located at various stages of life. This need has led to the development of techniques to model the dispersion of larvae based on hydrodynamic models of the ocean. At their simplest, these models distribute larvae as though they were inanimate particles drifting in ocean currents. More detailed models include larval behaviors such as the ability to swim, motion toward or away from light, and avoidance of warm or cool water (Hinckley et al. 1996; Hermann et al. 2001; Miller 2007; Hinckley et al. 2009). These are the simplest of what are called *individual-based models* (IBM), which are called *agent-based models* in other disciplines. An example is the DisMELS model (Dispersal Model for Early Life History Stages), which incorporates behavior in early life stages to study the dispersion of eggs and larvae (Stockhausen 2006). More complex models describe both larval distributions and the distributions of prey for the larvae. These tools, often called *nutrient-phytoplankton-zooplankton* (NPZ) models, describe spatial and temporal interactions (Franks 2002). Upper-trophic level models such as FEAST (Forage and Euphausiid Abundance in Space and Time) add predator and bioenergetic parameters (Aydin et al. 2010; Ortiz 2014).

Integrating and Analyzing Field Data using EcoDAAT

Web-based GIS applications for data discovery and marine planning encompass a wide range of tools. SeaSketch (http://mcclintock.msi.ucsb.edu/projects/seasketch) supports collaborative design but does not support data storage or analysis. SIMoN (http://sanctuarysimon.org/regional_sections/maps /mpaviewer/) and similar tools provide mapping of the location of data collection sites and links to project websites but do not allow for analysis of the data. GIS-based data discovery/download-only tools, such as many of those listed for ecosystem-based management (EBM) on the EBM Tools Network (http://www.ebmtools.org/?q=ebm_tools_taxonomy#), acquire and manage data but do not focus on geoanalysis. The strength of EcoDAAT is that it allows location and downloading of raw data as well as simple geoanalyses.

Because the field sampling for EcoFOCI work on pollock and other forage species is multidisciplinary, the datasets gathered are in disparate formats. The datasets for ichthyoplankton, zooplankton, and environmental measurements were originally stored in separate Microsoft Access databases. This made it hard to answer questions such as "What are the water temperatures where we find larger than average pollock larvae?" in a timely manner. Although the datasets were not enormous—the largest had about 250,000 records—network response time made it impractical to store the data in a single shared database on a network drive. Rather, it was necessary to have the complete databases on the desktop system of each user. This was a maintenance challenge and caused databases to get out of sync if updates failed or were delayed. The physical oceanographic datasets of EcoFOCI collaborators contain up to six million conductivity-temperature-depth (CTD) records, making it impossible to deploy them as part of the existing databases.

The project to create EcoDAAT had a number of goals: to integrate the databases so that crossdata-type queries were possible, move the integrated database to a larger and more robust database application, and make the data selection and display more geospatial by providing a map-based interface and ArcGIS for Server-based analytical tools. To meet these goals, an ArcGIS for Server/ Flex front end connected to an Oracle database was chosen. The system layout is shown in figure 1.2. The datasets are stored in a single Oracle database designed to contain, and link, the various types of data being collected. The data cascade is started at sea, where a stand-alone Java/SQLite application called the Cruise Operations Database (MasterCOD) is used to input metadata and collection information about every operation. The information includes the latitude and longitude of the operation, time and date, type(s) of equipment deployed, types of sample(s) gathered, and any notes about the operations. These records are used to track samples as they are analyzed and also to geolocate and cross-reference operations and samples. These datasets are used to populate the Haul tables of the Oracle database with core information and also to populate tables in two associated applications, ZooPPSI and IchPPSI. These two Java/SQLite applications are used by researchers at the Polish Plankton Sorting and Identification Center to record identifications and measurements of zooplankton and ichthyoplankton in samples sent to them for processing. Data from completed ZooPPSI and IchPPSI tables is loaded into the EcoDAAT Oracle tables using a series of Standard Query Language (SQL) scripts. These datasets are used to populate sample and specimen tables and associated tables to hold species and diet information. Information on larger age-0 and older fish collected during trawling operations is stored in Microsoft Excel spreadsheets at sea. The datasets are loaded into the Oracle database in a similar process to the loading of ichthyoplankton data. Laboratory analyses of stomach contents and other diet and condition-related information are loaded from text files and spreadsheets. Physical oceanographic datasets, including CTD and Seacat measurements, are processed using Seabird programs, reformatted into comma-separated value (CSV) files, and loaded using SQL scripts. Both CTD and bottle datasets are stored and can be linked to sample data.

The EcoDAAT front end uses the Flex viewer available with ArcGIS for Server (figure 1.2). This was chosen for ease of implementation and enhanced functionality. The front end allows for data selection by type of data (e.g., ichthyoplankton or oceanographic data), geographic location using graphical or text-based searches, time of collection, and combined criteria. The interfaces for data selection are dynamic, and the available selection criteria vary based on the type of selection being made. Because some of the data in EcoDAAT is still undergoing quality assurance/quality control (QA/QC), the front end uses Lightweight Directory Access Protocol (LDAP) authentication for secure access. In the future, a tiered access scheme, in which data can be made fully public, shared with collaborators only, or limited to program scientists, will be implemented.



Figure 1.2. EcoDAAT system architecture showing integration of Flex, Oracle, and ArcGIS for Server. Esri.

The results of a search are available both in summary form on the map portion of the front end (with data locations highlighted and the standard GIS Identify tool implemented to query data) and as a table that can be output as a shapefile or .csv file for further analysis (figure 1.3). The results of a search for haul information can also be linked to searches for associated data to create complex queries.



Figure 1.3. Example of the map-based selection tool in the Ecosystems Data Access and Analysis tool (EcoDAAT) and a table of data. By NOAA Alaska Fisheries Science Center; data sources: NOAA Alaska Fisheries Science Center, Esri, DeLorme, GEBCO, NOAA NGDC, National Geographic, HERE, Geonames.org, and other contributors.

Simple geoanalyses such as creating and plotting a bathymetric cross section are available (figure 1.4). These types of tools will be expanded in the future by taking advantage of the geoprocessing tools available in ArcGIS for Server.



Figure 1.4. Example of the cross section tool in EcoDAAT. By NOAA Alaska Fisheries Science Center; data sources: NOAA Alaska Fisheries Science Center, Esri, DeLorme, GEBCO, NOAA NGDC, National Geographic, HERE, Geonames.org, and other contributors.

Tracking the Transport and Dispersion of Larvae Using LarvaMap

Particle-tracking models enable the study of the motion of particles in the atmosphere or the ocean, from oil globules from an oil spill to plumes of smoke or sediment to even the ash from a volcanic eruption. Many models run locally and require setting up complicated parameter files for a model run. Examples include Ichthyop for Lagrangian modeling of the Bay of Biscay and Mediterranean Sea

(Lett et al. 2008) and the Lagrangian Transport Model (LTRANS), which was originally created to model the dispersion of oyster larvae in Chesapeake Bay (http://northweb.hpl.umces.edu/LTRANS .htm). Model runs can take hours to days to complete, depending on the number of particles released, duration of the run, and power of the underlying computer(s). There are an increasing number of web-based particle-tracking models, including the HYbrid Single-Particle Lagrangian Integrated Trajectory (HYSPLIT) Volcanic Ash Model (Draxler and Rolph 2003), which allows users to model ash transport from a list of volcanoes using one of five common atmospheric models. Connie2 (http://www.csiro.au/connie2/) allows the user to set up and run particle-tracking models for the seas around Australia and the Mediterranean Sea. Connie2 allows for either passive/ inanimate particles or particles showing limited behaviors such as the ability to swim.

Particle tracking is a critical element of modeling the early life history of pollock and other commercially important species. The models can be as simple as one that tracks inanimate particles and as complex as ones that describe the interactions at many trophic levels in an ecosystem. Lagrangian models follow an individual through space and time by calculating a path line and tracking the motion of the particle. Eulerian models use positions on a fixed spatial grid and streamlines to describe transport through the location. LarvaMap is a cloud-hosted implementation of a Lagrangian particle-tracking model supporting a variety of behaviors in a target species. It is an individual-based model capable of using output from a variety of ocean circulation models to drive the dispersion and movement of particles. The front end is a series of web pages for setting up larval behaviors, choosing the circulation or hydrodynamic model, configuring the model run, and launching the run. The model itself is run on a cloud computing resource (currently the Amazon cloud), but it could also be configured to run on a local resource or private cloud. The results are visualized as a track line in the web client, and results are available as an Esri shapefile and network Common Data Format (netCDF) file for further analysis.

LarvaMap is a combination of four pieces of technology working together (figure 1.5), consisting of:

- A behavior library to create, catalog, and share larval behaviors: http://bit.ly/WFiZjK
- A fate and transport model written as a Python library: http://bit.ly/16dFk9x
- A model service implementing the fate and transport model in a cloud architecture
- A web client for interacting with the Model Service through a representational state transfer application programming interface (REST API): http://bit.ly/14aZinD



Figure 1.5. LarvaMap architecture showing the interconnections between the various elements and the process of creating a model run. RPS/ASA.

Setting Larval Behaviors with the Behavior Library

The first step in running LarvaMap is discovering or creating a set of larval behaviors for an organism or species of interest using the Larval Behavior Library (figure 1.6). The library allows the creation of new behavior files, searching for existing files, and sharing files. By allowing behaviors to be exchanged between researchers, modelers, and students, it supports creating behaviors as a social process, with the ability to see how others have defined behaviors, to run models using both your own and community behavior files, and to work collaboratively on building behaviors. With the ease of running the models, it is possible to quickly and easily compare the results for different behaviors.



Figure 1.6. The interface to the Larval Behavior Library is used to define the behaviors of a specific

species. By NOAA Alaska Fisheries Science Center; data sources: NOAA Alaska Fisheries Science Center, Esri, DeLorme, GEBCO, NOAA NGDC, National Geographic, HERE, Geonames.org, and other contributors.

Behaviors that can be defined include:

- The geographic range of the organism and any keywords to be used for regional searches
- Adding life stages (e.g., egg, larval, juvenile) with associated duration in days or by temperature range
- Adding behaviors to life stages, including capabilities such as swimming speed and vertical migration, reactions such as avoidance behaviors, and sensory reactions such as avoiding salinities or temperatures
- The timing of the transition out of the plankton stage and where the organism will end up—settling on the bottom for benthic species and where in the water column pelagic species will reside

Configuring the Model Run Using the Web Client: Setting Parameters and Choosing Circulation Models

Configuring the model requires setting the location, date, time, and depth of the release and specifying the number of particles to be released. The behavior model is specified, usually by species. Vertical and horizontal dispersion coefficients can be set, and the underlying hydrodynamic model is specified (figure 1.7). Finally, the time step for the model runs is specified. Model durations of months and years can be supported, but the time step for the model must be selected to be reasonably small to provide the level of detail needed without requiring the calculation of an unnecessary number of time steps.



Figure 1.7. The interface for configuring and running a LarvaMap model. By NOAA Alaska Fisheries Science Center; data sources: NOAA Alaska Fisheries Science Center, Esri, DeLorme, GEBCO, NOAA NGDC, National Geographic, HERE,

Geonames.org, and other contributors.

Hydrodynamic models include those provided via THREDDS/OPeNDAP server or available as a local file. OPeNDAP and file-based datasets can be added to the LarvaMap system as they are discovered. An administrator of the LarvaMap system can add a dataset through an interface, and datasets must be registered in LarvaMap by an administrator of the LarvaMap system before they can be used for particle forcing. Both local and OPeNDAP datasets can be registered in LarvaMap. A Python library called paegan (https://github.com /kwilcox/paegan) was developed as part of the LarvaMap project to facilitate consistent access to many formats of hydrodynamic model data. If the dataset opens in paegan as a CommonDataset object, it is usable in LarvaMap. Currently, paegan supports regular grids (delta y == delta x), curvilinear grids (sometimes called *i/j grids*), and static triangular meshes. Dynamically changing grids and nontriangular meshes are not yet supported. LarvaMap does not support uploading of any ocean circulation data files for internal storage. File-based datasets must be made accessible to all model runs in the LarvaMap architecture before being used. Installations of LarvaMap using the distributed file system GlusterFS can achieve this.

The Regional Ocean Modeling System (ROMS) and US Navy Coastal Ocean Model (NCOM) models have been used to look at transport of pollock and herring in the north Pacific. Lowry (2013) used LarvaMap linked to an ROMS model for Prince William Sound to model drift and recruitment of herring. The NCOM region 7 model has been used to look at transport of sea urchin larvae off the California coast. The NCOM AmSeas model has been implemented for use in the Gulf of Mexico.

The Heart of LarvaMap: Fate and Transport Models

During a model run, the transport process is broken into smaller fragments called *models*. Each model implements a *move* method that moves a particle based on environmental and behavioral properties. Each model has access to environmental conditions from the hydrodynamic model (u, v, w); temperature; and salinity) at the current location and time of the particle. In a few specific scenarios, environmental data for a particle's location and time may be unknown: particle has left the domain of the hydrodynamic dataset, or there is a data gap in the hydrodynamic dataset. Under these conditions, the last known environmental conditions the particle experienced are used in the model.

LarvaMap implements three main models. The models are run sequentially, and the results are checked against the shoreline and bathymetry at the end (figure 1.8). A typical time step in a model run looks like this:

- Run transport (moves particle).
- Check shoreline.
- Check bathymetry.
- Run dispersion (moves particle).
- Check shoreline.
- Check bathymetry.
- Run behavior (moves particle).
- Check shoreline.
- Check bathymetry.



Figure 1.8. Elements of the transport process during a single time step in a model run. Variables *u*, *v*, and *w* are the circulation model velocity components. RPS/ASA.

Transport

The transport model for LarvaMap is a parallelized Lagrangian transport model using hydrodynamic circulation data stored as netCDF/OPeNDAP files. The model is written in Python and uses the paegan library, which provides a data model for ocean and meteorological data stored in netCDF files and the paegan-transport libraries for the model itself. It moves a particle based on the current velocity components (u [zonal or east–west], v [meridional or north–south], and w [vertical]) from the hydrodynamic model. The vertical component (w) is optional. If the transport model receives no u, v, or w data for the place and time of the particle, the last known conditions are used.

Dispersion

Each model run is configurable with a horizontal and vertical dispersion coefficient (in m/s). These coefficients are used to compute the u, v, and w dispersion.

Behavior

The behavior model is set up using a configuration from the Larval Behavior Library. A configuration consists of many life stages, each with its own behaviors. The behavior model is responsible for calling the correct life stage model based on such factors as a particle's development, age, and mass, and adding a "dead" life stage to every behavior configuration. This model does not actually move the particle, the life stage model does. It can be thought of as a *container* model.

Life Stage

The life stage model grows each particle based on time or as a function of temperature. It also calls the correct behavioral models associated with the particle's life stage. For example, if there are four diel behaviors configured for a single life stage, this life stage model determines which diel behavior to call based on the particle's time. The behavioral models described below work a bit differently from the transport and life stage models. They do not return a new location for the particle, they only return the u, v, and w vectors. The vectors are summed from all behaviors by this life stage model, and the particle is only moved once, after all behaviors are completed. If the life stage model receives no temperature or salinity data for the place and time of the particle, the last known conditions are used.

Life Stage: Diel

The diel model moves a particle up and down in the water column based on the sun's position in the sky. Each life stage of a larva has a configured static vertical swimming speed that is used here to determine how fast the particle can move to its desired depth range. If the particle's vertical swimming speed will put the particle beyond the desired depth range, the swimming speed is recomputed for a single model time step to land the particle in the middle of the desired depth range. If the desired depth range will not be reached, or the particle will end up inside the depth range, the vertical swimming speed configured in the behavior is used. A particle below desired depth will swim up, a particle below desired depth will swim down, and a particle at the desired depth will do nothing.

Life Stage: Settlement

The settlement model decides whether a particle should settle under its current conditions. As described in the Larval Behavior Library section, there are two types of settlement:

- Benthic: settles particle at 1 m above the bathymetry if bathymetry is within configured settlement depth range
- Pelagic: settles particle at its current depth if it is within configured depth range

The settlement model does not return u, v, or w vectors. It directly settles or moves the particle to the desired settlement location.

Life Stage: Dead

The dead model's purpose is to kill a particle that has not settled and has aged beyond its settlement life stage. It continues to track environmental conditions around the dead particle, and the particle will continue to be forced by the other models (transport/dispersion), but the dead model always returns 0 m/s u, v, and w.

Under the Hood: Circulation, Shoreline, and Bathymetric Data for Model Runs

Hydrodynamic Caching Strategy

One of the challenges in running a particle tracking model is that the particles are sparsely distributed. The transport model supports long-running models on the order of months and years, and there is actually no upper time limit imposed by the trajectory model. Although the circulation

model data needs to cover a large area to support any possible particle locations, at any given time step the particles occupy a small subset of this total area. Downloading the entire circulation/ hydrodynamic model data for each time step is prohibitive. But because LarvaMap allows for remote data access over the data access protocol (DAP), the entire hydrodynamic dataset for the area of interest never has to be downloaded and stored locally. The data can be accessed on the fly, and the model will download only the data it needs to run the model (figure 1.9). Data that is downloaded by the model is reused from particle to particle, so the same data is only accessed and downloaded once. Particles use the locally cached data in forcing algorithms.





Shoreline and Bathymetry

The fate and transport model uses land polygon objects to test for shoreline collisions. The land polygons can be stored in an Esri shapefile or accessed from an Open Geospatial Consortium (OGC) web feature service (WFS) server. Every particle loads the shoreline on creation and

indexes the shoreline around its current location. Every time a particle tests its location against the shoreline file, it first makes sure it has the appropriate data from the source shoreline dataset to do its calculation (figure 1.10). If it does not, it accesses the raw shoreline data, whether it is a shapefile or WFS server, and pulls a chunk of the data around its current location. This acts as a caching mechanism so the particles are not reading from the shoreline data at every detection check. A similar process is used with bathymetric data to constrain particles to the water column.



Figure 1.10. Shoreline indexing scheme used to locate the appropriate shoreline segments in the vicinity of a particle. Avoidance behavior when a particle nears a shoreline is shown in the last two parts of the figure. RPS/ASA.

The Model Service: Implementing LarvaMap in the Cloud

The model service queues model run requests received from the web client and distributes them within a network of "worker" servers. Worker servers can be running in many different places, including the Amazon cloud and local data centers. Each central processing unit (CPU) core on a worker server can force a single particle for the entire duration of its model. If one hundred worker cores are available, one hundred particles can be run in parallel. This can be a single model run with one hundred particles, or 10 model runs with 10 particles each. The model service will automatically

create new worker servers when the load demands it. Exactly when to spawn a new worker server is determined by how many particles are in the queue waiting to be run. If the number of particles in the queue is three times more than the number of CPU cores available on all worker servers, another worker server is initialized. Worker servers are shut down as the number of particles in the queue decreases. When no particles are in the queue, there is a minimum of one worker server running. Cloud computing technology allows LarvaMap to scale the number of worker servers dynamically as they are needed. In the near future, the model service will also be able to distribute a single particle's model run onto multiple servers, which can reduce model run times to one-tenth of the current times.

The model service collects the results from all the worker servers and pieces them back together into the final results for each model run. It creates the final output formats and uploads the result files to a common location for the web client to access. Utilities are available to monitor the status of the model service currently running jobs and details about any failed jobs.

Outputs from Model Runs

Output from LarvaMap is available in a number of formats. A GeoJSON track line is created for display in the web client. A shapefile, a netCDF file, and a Hierarchical Data Format 5 (HDF5) file are produced for display and analysis in ArcGIS and other tools (figure 1.11). The output files contain the date, latitude, longitude, and depth of the particle; the temperature and salinity values at the point if they are available from the hydrodynamic model; and the u, v, and w speed components at the grid point. A log file is also produced for troubleshooting and system analysis. TopoJSON is being considered in the future to allow a web client to show the track lines for each individual particle instead of using the single centroid track line GeoJSON output format currently available. An experimental TopoJSON output format is available at https://github.com/kwilcox/geojson _examples/blob/master/particles/huge_particle_tracklines.topojson.

Post Processing and Product Generation: Gridded Products

LarvaMap contains some utilities for post-processing the netCDF output into gridded products. These are typically used for stochastically representing the probability of a Lagrangian element passing through a grid cell or, in the case of larvae, settlement occurring in a grid cell based on an ensemble of model runs. The gridded products can be output as GeoTIFFs for use in GIS applications such as ArcGIS ArcMap. This workflow is not automated as part of the model implemented in the cloud, but could be easily added to the cloud-based system.



Figure 1.11. Results from a LarvaMap run shown in (a) the LarvaMap interface and the track lines shown in the (b) ArcMap and (c) ArcScene applications in ArcGIS. By NOAA Alaska Fisheries Science Center; data sources: NOAA Alaska Fisheries Science Center, Esri, DeLorme, GEBCO, NOAA NGDC, National Geographic, HERE, Geonames.org, and other contributors.

Role of Cloud and ArcGIS for Server, Future Needs

Creating EcoDAAT and LarvaMap has allowed us to explore practical applications of cyber-GIS advancements. Although EcoDAAT is not currently a cloud solution, the architecture of ArcGIS for Server means it could be redeployed using cloud storage and computation. The web client can use a variety of back-end architectures. LarvaMap is a more classic implementation of cyber-GIS. It uses cloud storage, optimized data transfer schemes, cloud-based computation, and local analysis of model results. In this way it requires solutions to Yang et al.'s (2011) challenges of data intensity, computing intensity, concurrent intensity, and spatiotemporal intensity.

Implementation of these tools has produced some challenges, and some unexpected opportunities. The EcoDAAT integration of data and ability to make data easily accessible has also raised challenges concerning data sharing and data use. Although the recent White House memorandum on open data (President Barack Obama 2013) calls for free and open distribution of data, this must be balanced with the need to make sure the data being distributed is fully quality controlled and the right of the researchers to their intellectual property is respected. Tiered access and other technical solutions supported by cyberinfrastructure will allow us to balance these needs.

Deploying LarvaMap raises challenges in procuring cloud resources for government use. Full deployment would require the security accreditation of cloud storage and computing and improved procedures for paying for cloud time. Most or all of these challenges could be solved by the creation of private clouds within the federal system or deploying LarvaMap in local high-performance computing resources.

One totally unexpected use of LarvaMap is as a teaching tool. Because of its ease of use and computational capabilities, it can be used by students to explore larval transport. A recent biological oceanography course in the joint Massachusetts Institute of Technology–Woods Hole Oceanographic Institution graduate program had the students using the tool to create and run scenarios for their final projects (L. Mulineaux, personal communication). This also provided us the benefit of extended stress testing of the system as the students created and ran numerous models.

These tools have proven that "GIS detached from the desktop and deployed on the web" can provide critical resources to support fisheries research and management. We plan to continue taking advantage of future cyber-GIS developments to improve and expand our capabilities.

Acknowledgments

This chapter was greatly improved by reviews from Christine Baier, Annette Dougherty, Ann Matarese, and Nazila Merati and an anonymous reviewer. This research was supported, in part, with funds from the NOAA High Performance Computing and Communications (HPCC) program and the National Marine Fisheries Service Ecosystems and Fisheries Oceanography Coordinated Investigations and funding to Kyle Wilcox from the Oil Spill Recovery Institute. This is contribution EcoFOCI-N819 to the NOAA North Pacific Climate Regimes and Ecosystem Productivity research program. The findings and conclusions in this paper are those of the authors and do not necessarily represent the views of the NOAA National Marine Fisheries Service. Reference to trade names does not imply endorsement by the NOAA National Marine Fisheries Service.

References

- AFSC (Alaska Fisheries Science Center). 2010. Walleye Pollock. http://www.afsc.noaa.gov/Education/factsheets /10_Wpoll_FS.pdf. Last accessed February 15, 2014.
- ------. 2013. Walleye Pollock Research. http://www.afsc.noaa.gov/species/pollock.php. Last accessed February 11, 2014.
- Aydin, K., N. Bond, E. N. Curchitser, M. Dalton, G. A. Gibson, K. Hedström, A. J. Hermann, E. Moffitt, J. Murphy, I. Ortiz, A. Punt, and M. Wan. 2010. "Integrating Data, Fieldwork, and Models into an Ecosystem-Level Forecasting Synthesis: The Forage-Euphausiid Abundance in Space and Time (FEAST) Model for the Bering Sea Integrated Research Program." *ICES CM 2010* L:21. http://info.ices.dk/products/CMdocs/CM-2010/L/L2110.pdf.
- Bailey, K. M. 2013. Billion-Dollar Fish: The Untold Story of Alaska Pollock. Chicago: University of Chicago Press.
- Coppock, J. T., and D. W. Rhind. 1991. "The History of GIS." In *Geographical Information Systems: Principles and Applications*, edited by D. J. Maguire, M. F. Goodchild, and D. Rhind, 21-43. Vol. 1. New York: Longman Scientific & Technical; New York: Wiley.
- Cushing, D. H. 1974. "The Natural Regulation of Fish Populations." In *Sea Fisheries Research*, edited by F. R. Harden-Jones, 399–412. New York: John Wiley and Sons.
- Dickey, T. D. 2003. "Emerging Ocean Observations for Interdisciplinary Data Assimilation Systems." *Journal of Marine Systems* 40/41: 5–48.
- Draxler, R. R., and G. D. Rolph. 2003. HYSPLIT (HYbrid Single-Particle Lagrangian Integrated Trajectory) Model. http://www.arl.noaa.gov/HYSPLIT.php. Last accessed June 4, 2014.
- Fornshell, J. A., and A. Tesei. 2013. "The Development of SONAR as a Tool in Marine Biological Research in the Twentieth Century." *International Journal of Oceanography* 3:1–9.
- Franks, P. J. S. 2002. "NPZ Models of Plankton Dynamics: Their Construction, Coupling to Physics, and Application." *Journal of Oceanography* 58 (2): 379–87.
- Gentleman, W. 2002. "A Chronology of Plankton Dynamics in silico: How Computer Models Have Been Used to Study Marine Ecosystems." *Hydrobiologia* 480 (1–3): 69–85.
- Hamblin, J. D. 2005. Oceanographers and the Cold War: Disciples of Marine Science. Seattle: University of Washington Press.
- Hare, J. A. 2014. "The Future of Fisheries Oceanography Lies in the Pursuit of Multiple Hypotheses." *ICES Journal* of Marine Science. doi:10.1093/icesjms/fsu018.
- Haury, L. R., J. A. McGowan, and P. H. Wiebe. 1978. "Patterns and Processes in the Time-Space Scales of Plankton Distributions." In Spatial Pattern in Plankton Communities, edited by J. H. Steele, 277–327. New York: Plenum Press.
- Hermann, A. J., S. Hinckley, B. A. Megrey, and J. M. Napp. 2001. "Applied and Theoretical Considerations for Constructing Spatially Explicit Individual-Based Models of Marine Larval Fish that Include Multiple Trophic Levels." *ICES Journal of Marine Science* 58: 1030–41.
- Hinckley, S., A. J. Hermann, and B. A. Megrey. 1996. "Development of a Spatially Explicit, Individual-Based Model of Marine Fish Early Life History." *Marine Ecology Progress Series* 137:47–68.
- Hinckley, S., J. M. Napp, A. J. Hermann, and C. Parada. 2009. "Simulation of Physically Mediated Variability in Prey Resources of a Larval Fish: A Three-Dimensional NPZ Model." *Fisheries Oceanography* 18 (4): 201–23.
- Hjort, J. 1914. "Fluctuations in the Great Fisheries of Northern Europe in the Light of Biological Research." *Rapports* et Proces-verbaux des Réunions. Conseil International pour l'Éxploration de la Mer 20:1–228.
- Houde, E. D. 2008. "Emerging from Hjort's Shadow." Journal of Northwest Atlantic Fishery Science 41:53-70.
- Kendall, A. W., J. D. Schumacher, and S. Kim. 1996. "Walleye Pollock Recruitment in Shelikof Strait: Applied Fisheries Oceanography." *Fisheries Oceanography* 5:4–18.
- Lang, G., R. Brodeur, J. Napp, and R. Schabetsberger. 2000. "Variation in Groundfish Predation on Juvenile Walleye Pollock Relative to Hydrographic Structure near the Pribilof Islands, Alaska." *ICES Journal of Marine Science* 57 (2): 265–71.
- Lasker, R. 1978. "The Relation between Oceanographic Conditions, and Larval Anchovy Food in the California Current: Identification of Factors Contributing to Recruitment Failure." *Rapports et Proces-verbaux des Réunions. Conseil International pour l'Éxploration de la Mer* 173:212–30.

- Lett, C., P. Verkey, C. Mullon, C. Parada, T. Brochier, P. Penven, and B. Blanke. 2008. "A Lagrangian Tool for Modelling Ichthyoplankton Dynamics." *Environmental Modelling & Software* 23:1210–14.
- Levin, L. A. 2006. "Recent Progress in Understanding Larval Dispersal: New Directions and Digressions." *Integrative and Comparative Biology* 46 (3): 282–97.
- Link, J. S. 2002. "What Does Ecosystem-Based Fisheries Management Mean?" Fisheries 27 (4): 18-21.
- Lowry, N. 2013. Larva Map: Final report to the Oil Spill Response Institute. Seattle, WA: Research4D.
- Megrey, B., S. Hinckley, and E. Dobbins. 2002. "Using Scientific Visualization Tools to Facilitate Analysis of Multidimensional data from a Spatially Explicit, Biophysical, Individual-Based Model of Marine Fish Early Life History." *ICES Journal of Marine Science* 59 (1): 203–15.
- Miller, T. J. 2007. "Contribution of Individual-Based Coupled Physical-Biological Models to Understanding Recruitment in Marine Fish Populations." *Marine Ecology Progress Series* 347:127–38.
- National Research Council. 2003. The Decline of the Stellar Sea Lion in Alaskan Waters: Untangling Food Webs and Fishing Nets. Washington, DC: The National Academies Press.
- Nyerges, T. 2010. *Cyber-Enabled Platforms for Regional Ocean and Coastal Governance*. National Science Foundation TeraGrid Workshop on Cyber-GIS. Washington, DC: National Science Foundation. http://www.cigi.illinois .edu/cybergis/docs/Nyerges_Position_Paper.pdf.
- Ortiz, I., K. Aydin, A. J. Hermann, and G. Gibson. In review, 2014. "Climate to Fisheries: A Vertically Integrated Model for the Eastern Bering Sea." *Deep Sea Research II*.
- President Barack Obama. 2013. Memorandum on Open Data Policy–Managing Information as an Asset (May 9, 2013). http://www.whitehouse.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf. Last accessed June 4, 2014.
- Rainger, R. 2000. "Patronage and Science: Roger Revelle, the US Navy, and Oceanography at the Scripps Institution." *Earth Sciences History* 19 (1): 58–89.
- Rozwadowski, H. M., and D. K. Van Keuren. 2004. *The Machine in Neptune's Garden: Historical Perspectives on Technology and the Marine Environment*. Canton, NJ: Science History Publications.
- Stockhausen, B. 2006. DisMELS: A Dispersal Model for Early Life History Stages. http://www.afsc.noaa.gov /quarterly/jas2006/divrptsREFM6.htm. Last accessed February 23, 2014.
- Vance, T. C. 2008. "If You Build It, Will They Come? Evolution towards the Application of Multi-dimensional GIS to Fisheries-Oceanography." PhD diss. Corvallis, OR: Oregon State University.
- Vance, T. C., and R. E. Doel. 2010. "Graphical Methods and Cold War Scientific Practice: The Stommel Diagram's Intriguing Journey from the Physical to the Biological Environmental Sciences." *Historical Studies in the Natural Sciences* 40 (1): 1–47.
- Vance, T. C., B. A. Megrey, and C. W. Moore. 2006. "GeoModeler: Integration of a Nutrient-Phytoplankton-Zooplankton (NPZ) Model and an Individual-Based Model (IBM) with a Geographic Information System (GIS)." Paper presented at the ICES Annual Science Conference, Masstricht, Netherlands.
- Wang, S., L. Anselin, B. Bhaduri, C. Crosby, M. F. Goodchild, Y. Liu, and T. L. Nyerges. 2013. "CyberGIS Software: A Synthetic Review and Integration Roadmap." *International Journal of Geographical Information Science* 27 (11): 2122–45.
- Wiebe, P. H., and M. C. Benfield. 2003. "From the Hensen Net toward Four-Dimensional Biological Oceanography." Progress in Oceanography 56 (1): 7–136.
- Wright, D. J., and S. Wang. 2011. "The Emergence of Spatial Cyberinfrastructure." In Proceedings of the National Academy of Sciences 108 (14): 5488–91.
- Wright, D., S. Kopp, and C. Brown. 2013. "What Is Cyber GIS?" *Esri Insider*. http://blogs.esri.com/esri /arcgis/2013/10/01/what-is-cybergis/. Last accessed January 14, 2014.
- Yang, C., M. Goodchild, Q. Huang, D. Nebert, R. Raskin, Y. Xu, M. Bambacus, and D. Fay. 2011. "Spatial Cloud Computing: How Can the Geospatial Sciences Use and Help Shape Cloud Computing?" *International Journal* of Digital Earth 4 (4): 305–29.
- Yang, C., Y. Xu, and D. Nebert. 2013. "Redefining the Possibility of Digital Earth and Geosciences with Spatial Cloud Computing." *International Journal of Digital Earth* 6 (4): 297–312.

Supplemental Resources

URLs and QR codes are provided for the digital content that goes with this chapter.

Hyperlinks are also available on the Esri Press "Book Resources" webpage for chapter 1. Go to esripress.esri.com/bookresources. Then, in the list of Esri Press books, click *Ocean Solutions, Earth Solutions*. On the *Ocean Solutions, Earth Solutions* resource page, click the chapter 1 link to access that webpage and the hyperlinks listed there.

LarvaMap-Related Resources

• Larval Behavior Library to create, catalog, and share larval behaviors, at http://behavior.larvamap.axiomdatascience.com/



• LarvaMap Transport Model written as a Python library, at http://github.com/asascience-open/paegan-transport



• LarvaMap web client for interacting with a web service through a REST API, at http://services.asascience.com/mapapp/larvamap



• Paegan Common Access Library, at https://github.com/asascience-open/paegan



 Online archive of particle-tracking models, at http://www.nefsc.noaa.gov/drifter/particles.html



• OceanNOMADS for access to hydrodynamic models, at http://ecowatch.ncddc.noaa.gov



• Access to high-resolution operational hydrodynamic models, at http://www.opc.ncep.noaa.gov/newNCOM/NCOM_currents.shtml



EcoDAAT-Related Resources

 Ichthyoplankton Information System for information on the early life histories of fish species in the Bering Sea and Gulf of Alaska, at http://access.afsc.noaa.gov/ichthyo/index.php



• Survey data on the occurrence of adult pollock and other species, at http://www.afsc.noaa.gov/RACE/groundfish/survey_data/default.htm



• Mapping essential fish habitat for a variety of species, including pollock, at http://www.habitat.noaa.gov/protection/efh/efhmapper/index.html

