



Jul 1st, 12:00 AM

Particles in the Cloud – running GIS-based particle tracking models for fish larval dissemination in traditional and cloud computing architectures

Tiffany C. Vance

Kyle Wilcox

Follow this and additional works at: <https://scholarsarchive.byu.edu/iemssconference>

Vance, Tiffany C. and Wilcox, Kyle, "Particles in the Cloud – running GIS-based particle tracking models for fish larval dissemination in traditional and cloud computing architectures" (2012). *International Congress on Environmental Modelling and Software*. 170.
<https://scholarsarchive.byu.edu/iemssconference/2012/Stream-B/170>

This Event is brought to you for free and open access by the Civil and Environmental Engineering at BYU ScholarsArchive. It has been accepted for inclusion in International Congress on Environmental Modelling and Software by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.

Particles in the Cloud – running GIS-based particle tracking models for fish larval dissemination in traditional and cloud computing architectures

Tiffany C. Vance¹ and Kyle Wilcox²

¹ San Francisco State University [tcvance@sfsu.edu]

² Applied Science Associates [kwilcox@asascience]

Abstract: Models provide a way to understand the complex interactions of biological and physical process. As an example, the movement of dust plumes generated on continents over the adjoining oceans can be modeled for both the physical process of particle transport and for the geochemical effects of the dust deposition on life in the oceans. Models can also describe and link two physical processes; a model can be used to simulate the complex 3-dimensional movement of ocean currents and these model results can be linked to models that predict the movement of many waterborne objects such as fish and larvae, oil from seeps or accidental spills, and marine debris.

One of the more common uses of particle tracking models is to study the dispersion of the larvae of marine fishes. Many marine fishes spawn in areas where their larvae are subject to dispersion due to currents after hatching. This dispersion can affect both their survival and their growth. The motion of particles, or organisms too small to swim, can be calculated by using a combination of general circulation models and particle tracking models to describe the physics and biology of the particles being studied. A GIS can be used to set up the parameters for a model run and also to display the results.

LarvaMap uses a map based display for set up and also is tightly linked to ArcGIS for analysis of the output. The GIS can be used to integrate the particle tracking model output with *in situ* measurements of oceanographic and environmental parameters. LarvaMap can use both traditional local computing resources and can be deployed on the Amazon cloud to test the speed and costs of cloud deployment. The transition from a local to a cloud deployment is relatively simple and the performance improvements can be considerable.

Keywords: GIS; particle tracking; environmental modelling; fisheries

1 INTRODUCTION AND BACKGROUND

1.1 Introduction

The motion of particles, or organisms too small to swim, can be calculated by using a combination of general circulation models and particle tracking models to describe the physics and biology of the particles being studied. These linked models use an underlying circulation model to describe motion in one or more dimensions, and an overlying particle-tracking model describes the physical behavior of the organism or particle in a fluid or gaseous medium. A model of the biological behavior and development of the organism can be layered on top of the particle tracking model to better simulate life histories of organisms. Setting up the parameters for these model runs is a complex process and the output can be

voluminous and complicated. Visual techniques for both the input of model parameters and the output of visualizations can make it much easier to use these models, and to understand the results of model runs.

These models can be computed in one-, two-, three-, or four-dimensions. One-dimensional models describe the motion of particles at a point in an imaginary stream or other linear feature. Adding a second dimension allows the model to show the evolution of patterns in arbitrary vertical or horizontal slices of space or planes. Creating animations with these would provide a third temporal dimension. Spatially three-dimensional models show snapshots of the positions of particles in space and a series of these snapshots can provide a rough animation to show the evolution of trajectories and patterns through time. Four-dimensional models provide both spatial and temporal understanding, and can be used to study the evolution of patterns with time.

1.2 Examples of Particle Tracking Models in Fisheries Research

In fisheries research, particle-tracking models are used to study the dispersion and survival of a variety of species. Ribergaard et al. (2004) have used particle-tracking models based on finite element circulation models to study the transport of shrimp larvae west of Greenland. The particle-tracking model uses Lagrangian advection and uses diffusion to represent the effects of small-scale turbulence. The Quoddy circulation model drives the particle tracker. Grzechnik and Noye [1998] used a similar model on top of the ted circulation model to calculate the spread of western king prawn larvae in Gulf St. Vincent, Australia. Brickman et al. [2007] used the METACOD circulation model and an advective and random displacement particle-tracking algorithm to study the dispersion of cod larvae near Iceland. While the particle tracks were mapped, there was no spatial analysis of the results. In the north Pacific, models have been used to study the transport of Walleye pollock in Shelikof Strait [Hermann et al., 1996] and Lanksbury et al. [2007] looked at the dispersion and transport of rock sole larvae.

An example of a particle-tracking model linked to a GIS is the Chesapeake Bay Oyster Larvae Tracker (CBOLT) [<http://ches.communitymodeling.org/models.php>] CBOLT used an Internet Map Server interface to set up the parameters (start location for larvae, duration of run, run of a hindcast or a forecast etc.) for submission to a model controller running the Chesapeake Bay 3-D Physical Oceanographic (C3PO) Model. This finite element hydrodynamic model was developed by the NOAA Coast Survey Development Lab to simulate circulation in the Bay. The output of the hydrodynamic model drove a particle-tracking model and the results were stored in an ArcSDE database. The results were displayed back to the user using the same ArcIMS interface used to set up the parameters to run the model. The results were two-dimensional and static. CBOLT did not take into account the biological development of the larvae. This behavior is being developed in a more recent project described in North et al. [2008].

GeoFish, described in Vance et al. [2008] was a prototype of linking a particle tracking model with a GIS for analysis and display of the results. The particle tracking model from CBOLT was linked to an ArcIMS and VTK-based interface. The interface was used to set the release location for the larvae, the release data and time, the run duration and direction, the number of larvae to release and the type of spatial output. The results were returned as a netCDF file of a three-dimensional grid of the trajectories of the larvae over time. The features can be shown as a two-dimensional map in an ArcIMS and as a three-dimensional visualization in GeoFish. The ArcIMS renderer provided analyses such as spatial queries and buffering. The GeoFish renderer allowed for color-coding of the tracks, overlays of other parameters stored in shapefiles and netCDF files, and zoom, pan and rotation of the visualization for user exploration [Figure 1].

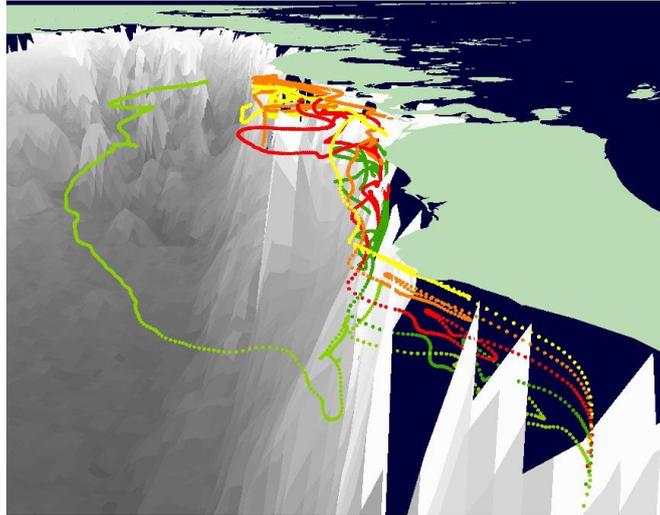


Figure 1. GeoFish visualization of rock sole larval trajectories. The larvae were seeded in a vertical column near the shore. Most drift along shore close to the shoreline but one is carried offshore into deeper water.

2 PARTICLES IN THE CLOUD: LARVAMAP

LarvaMap is a tool for scientists and resource managers to use to create scenarios for larval dispersion (species, start location, start depth, duration of dispersion), to run the particle tracking model, and to display the results for visualization and analysis. LarvaMap (Figure 2) uses a map based display for set up and also is tightly linked to ArcGIS for analysis of the output. The GIS can be used to integrate the particle tracking model output with *in situ* measurements of oceanographic and environmental parameters. LarvaMap can use both traditional local computing resources and can be deployed in the cloud to test the speed and costs of cloud deployment. The transition from a local to a cloud deployment is relatively simple and the performance improvements can be considerable.

The motivation for developing LarvaMap was a desire to make it easy for researchers to create and run their own scenarios. The advent of high performance computing resources and the availability of circulation model data via OPeNDAP and other servers makes it possible to run this kind of model on the user's desktop via a web interface. This rapid turn around and easy visualization makes it possible to use these models for exploratory analyses.

2.1 Model Architecture

The front-end client is built with Adobe Flex and the OpenScales Mapping application programming interface to provide an exploratory GIS map interface. Users define larval release parameters, including temporal extent, release duration, number of particles, and species, all from within a Web browser. These values are passed via XML to a Java-based REST Web service running on external servers that process the parameters, generate a unique run ID (which can be used to share model runs among users) and add the parameters onto the end of a queuing system. When the Java-based backend model runner sees a message waiting on the queue, it processes the message and initiates the larval-transport model.

While the model is running, the service returns progress messages to clients, so users can monitor progress. After the model run completes, the service returns an XML representation of the computed trackline to the client, which is displayed on the mapping interface. Users also can download a netCDF file of the model results from the client for use in external applications such as ArcGIS.

The particle-based larval model is driven by physical ocean conditions such as temperature and water speed (currents), and LarvaMap can use many diverse hydrodynamic data sources to drive the transport model. Leveraging the capabilities of the Unidata Java NetCDF library, it is possible to use local files supported by the library and remote datasets via the OPeNDAP protocol.

2.2 Traditional Deployment and Use

The interface for LarvaMap is a web page where the user can set the starting points(s) for the particles as a point location or as a polygonal area. The user can also specify the number of particles to release, the depth(s) at which they are released and the duration of the model run. A variety of underlying circulation model outputs can be used and the horizontal and vertical diffusivity can also be specified. The input parameters are verified before the model run is submitted. Once the user approves all of the parameters the model run is launched. When the run is complete, a summary map of the locations of the centroid of the particle “cloud” is drawn, the total model run time is given and the full results are available for download as a netCDF file.

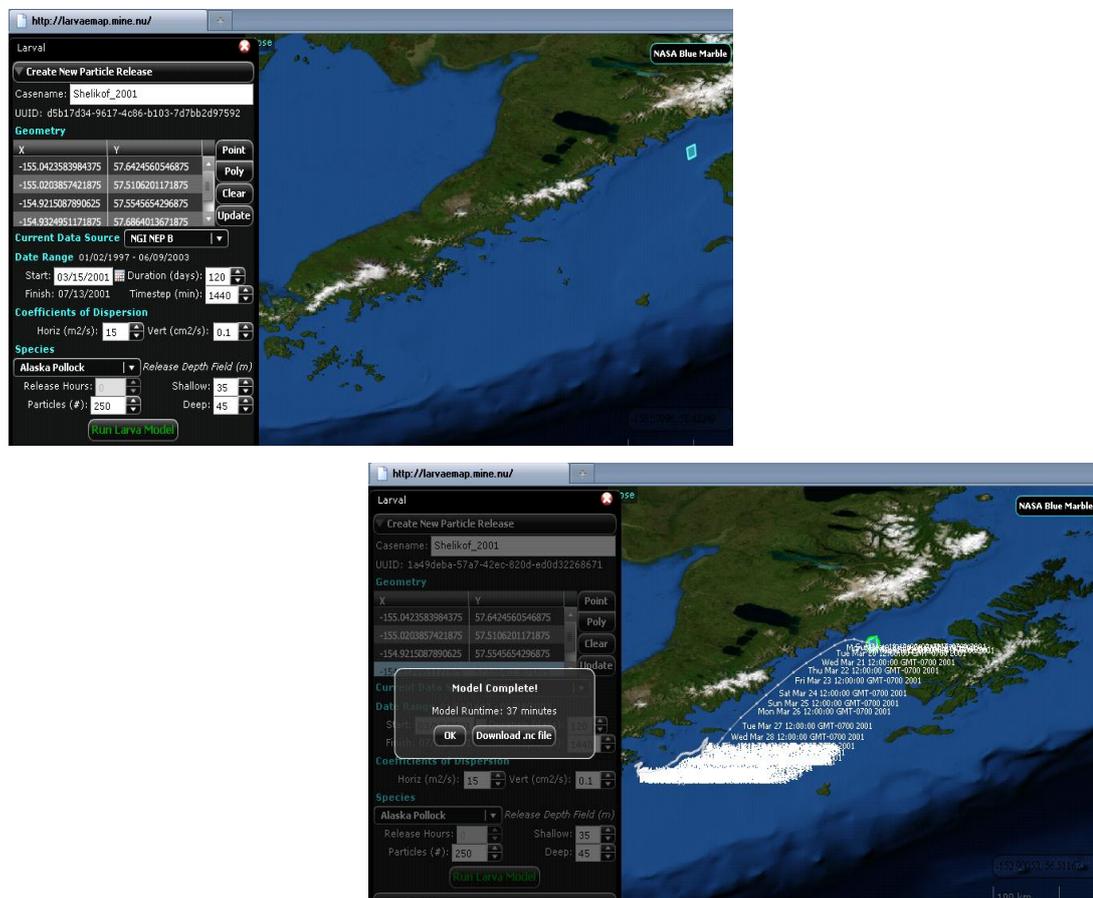


Figure 2. The LarvaMap interface. The left-hand image shows the setup of model parameters. The right-hand image shows the results with the daily centroids of the particle cloud plotted on a map and the full results available for download as a netCDF file.

Since the results are in a netCDF file, they can easily be ingested and analysed in a GIS such as ArcGIS. Figure 3 shows the output of two model runs as both a 3-D perspective plot and as a 2-D map which can also be displayed as an animation. Analyses of the tracks, such as the location of the centroid of a daily or hourly “cloud” of particles or along path plots of temperatures (from the underlying circulation models) or path lengths can be made using the tools intrinsic to the GIS.

The results could also be integrated with *in situ* environmental or biological data gathered in the same geographic locations or the results of other types of models such as wind and wave forecasts.

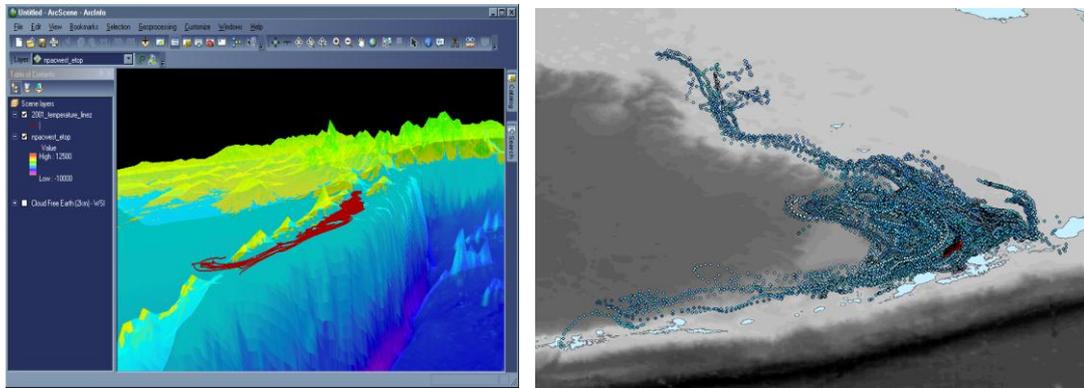


Figure 3. Visualizations of the LarvaMap output. The left-hand view is a 3-D display of the particle tracks. The right-hand display shows the positions of all the particles at all time steps of the model run. The red symbols show the particle locations for a single day and can be animated to show the travel of the particles for the entire model run.

2.3 Deploying LarvaMap in the Cloud

The original LarvaMap deployment was a standalone application, utilizing resources on a single piece of hardware to power all system aspects. The time required to complete the model was relative to the number and speed of CPUs available. RAM use wasn't as important, but to efficiently compute a large-scale particle run, 512MB of free RAM was required. Because there was only one model runner, multiple users would be required to complete any existing model run before moving on to the next. Using a cloud based architecture described in Wilcox and Vance [2011] solved a number of challenges by providing the ability to scale out the application horizontally to provide computational resources on demand. LarvaMap was re-architected for a cloud-based infrastructure and redeployed on the Amazon cloud.

Front-end clients and middleware Web services were redesigned to be "hot deployable" to multiple instances and fronted with Amazon load balancers. This allowed the client and Web services to support an almost unlimited number of simultaneous users. To support multiple simultaneous model runs, the system required multiple model runners, and this required a way to manage communication among the Web service and runners. Using the Amazon Simple Queue Service, a system was developed that allowed a model run request to be serviced by any available model runner.

Cloud resources improved the computation time of the model runners significantly. Modern laptop and desktop computers typically have anywhere from one to four CPU cores, meaning they can simultaneously perform one to four calculations at the same time. The particle model makes heavy use of all available CPU resources on computer hardware. To the model runner, a computer with only one core can only run the transport calculations for one particle at a time. If a computer has four cores, the model runner can perform four transport calculations at the same time, reducing processing time by 75 percent. Amazon offers a wide variety of instance types, each with dedicated amounts of RAM and CPU resources [aws.amazon.com/ec2/instance-types]. Because the model runner was highly dependent on CPU, and not RAM, a "high-CPU" instance was chosen with eight CPU cores and 7GB memory. This allowed multiple model runners to run on the

same instance (reducing costs) and gave the runners enough processing power to complete large-scale model runs relatively quickly.

3 CONCLUSIONS AND RECOMMENDATIONS

Running particle tracking models and integrating the results with environmental or *in situ* data can be a powerful tool to understand the early life history and success of marine fish species. But, running these models is frequently a complex and time-consuming process. An intuitive interface for parameter setting combined with harnessing the power of cloud computing resources can make running these models intuitive and rapid. This kind of model deployment can make it possible for scientists and resource managers to use these tools for “what if” scenario testing and the exploration of thought models.

In the future, we plan to expand LarvaMap by including biological behaviors and responses for the particles, currently a notable gap in our tool. We are working on a joint project to create a framework where users can specify a set of biological behaviors, such as time spent as passive particles or swimming strength or timing of the development of eyes capable of seeing well enough for predator avoidance, for a species of interest. We will also implement links to THREDDS or other data servers to allow a variety of circulation models to be implemented. These improvements will build upon a robust computational structure to create a fully functional biological model.

ACKNOWLEDGMENTS

The US National Oceanic and Atmospheric Administration (NOAA) funded the development of LarvaMap and associated work through generous support from its High Performance Computing and Communications Office and the IT Innovation Program [<http://www.cio.noaa.gov/HPCC/innovation/index.html>]. The NOAA/NMFS AFSC Recruitment Processes Program also supported the project. Deployment of LarvaMap in the cloud was supported by the US Federal Geospatial Data Committee (FGDC) via their GeoCloud Sandbox initiative [cisc.gmu.edu/scc/presentation/GeoCloud-doug-panel.pdf]. The authors thank the three anonymous reviewers for their helpful comments.

REFERENCES

- Brickman, D., G. Marteinsdottir, K. Logemann, and I. H. Harms, Drift probabilities for Icelandic cod larvae. *ICES Journal of Marine Science: Journal du Conseil Advance Access* Jan 1, 2007, 49 – 59, 2007.
- Chesapeake Community Modeling Program, <http://ches.communitymodeling.org/models.php>, viewed 2/24/2102.
- Grzechnik, M. P., and B. J. Noye, A Lagrangian-Stochastic particle tracking procedure for coastal seas. In Noye, B. J., M. D. Teubner, and A. W. Gill, editors, *Computational Techniques and Applications: CTAC-97*, 257- 264. The University of Adelaide, South Australia, World Scientific, Singapore, 1998.
- Hermann, A.J., W.C. Ruge, P.J. Stabeno, and N. A. Bond, Physical transport of young pollock larvae (*Theragra chalcogramma*) near Shelikof Strait as inferred from a hydrodynamic model, *Fish. Oceanogr.*, 5(Suppl. 1), 58-70, 1996.

- Lanksbury, J.A., J.T. Duffy-Anderson, K.L. Mier, M.S. Busby, and P.J. Stabeno, Distribution and transport patterns of northern rock sole, *Lepidopsetta polyxstra*, larvae in the southeastern Bering Sea. *Prog. Oceanogr.*, **72**, 39–62, 2007.
- North, E. W., Z. Schlag, R. R. Hood, M. Li, L. Zhong, T. Gross, and V. S. Kennedy, Vertical swimming behavior influences the dispersal of simulated oyster larvae in a coupled particle-tracking and hydrodynamic model of Chesapeake Bay, *Marine Ecology Progress Series* 359: 99-115, 2008.
- Ribergaard, M.H., S.A. Pedersen, B. Aadlandsvik, and N. Kliem, Modeling the ocean currents on the West Greenland shelf with special emphasis on northern shrimp recruitment, *Continental Shelf Research*, **24** (13-14), 2004.
- Vance, T.C., S.L. Cross, B.A. Megrey, S.M. Mesick, and C.W. Moore, GeoFish – visualization and analysis of particle tracking model output for fish and shellfish larvae, ICES 2008 Annual Science Conference - theme session R (Data Management) Halifax, Nova Scotia. 22-26 September 2008.
- Wilcox, K. and T.C. Vance, Ride the wave: ocean scientists use cloud computing to model dispersion, *GeoWorld*, September 2011, 18-21, 2011.