Abstract—OurOcean (http://ourocean.jpl.nasa.gov) is a web portal developed in JPL to enable users to easily access ocean science data, run data assimilation models, and visualize both data and models. Currently, OurOcean provides both real-time and retrospective analysis of observation data and ocean model simulations in the Pacific Ocean. We are particularly interested in the U.S. West Coastal Ocean with focused areas around Southern California, Central and Northern California, and Prince William Sound in Alaska. OurOcean serves four primary functions: 1) Retrieve, process, and archive real-time in-situ and satellite observation data and model outputs from distributed sources. 2) Produce value-added products by combining multiple data sources. 3) Supply various methods to access and visualize the data and model output. 4) Provide on-demand modeling capability for user to configure his/her own ocean model, submit and monitor the job status, and finally retrieve and visualize the model output.

In this paper, we will present the infrastructure of the OurOcean portal, the architecture of the four servers and its underlined technologies. We will also discuss some state-of-the-art web technologies and computing technologies to be incorporated in OurOcean for interactive visualization and on-demand modeling.

1. INTRODUCTION

A comprehensive understanding of ocean process requires sustained and long-term measurement via in-situ instruments at the sea surface and in the water column; through remote-sensing observations for global measurements such as SST, SSH, ocean vector winds and ocean color; and 3-dimensional computational modeling that models ocean’s biological, chemical and physical phenomena in continuous spatial and temporal dimensions.

The emerging ocean observing systems and the advanced computer models powered by rapidly changing supercomputing technology produce terabytes of data and images. These data sets are collected by different agencies, stored in different formats, and served using different protocols. They cover the entire world ocean with multiple scales ranging from global at hundreds of kilometers to regional/coastal at a few kilometers. How to manage this multi-phenomenon, multi-resolution, constantly growing volume of ocean observation data and model output presents a great challenge to the state of the art information technology. More importantly, how to make these data accessible to scientists and general public in a friendly user interface becomes an imminent task for the Ocean Science community.

OurOcean is a web portal developed by JPL to enable users to easily access ocean science data, run data assimilation models, and visualize both data and models. Currently, OurOcean provides both real-time and retrospective analysis of observation data and ocean model simulations in the Pacific Ocean. We are particularly interested in the U.S. West Coastal Ocean with focused areas around Southern California, Central and Northern California, and Prince William Sound in Alaska. OurOcean serves four primary functions: 1) Retrieve, process, and archive real-time observation data and model output from distributed sources, including global QuikSCAT wind, ocean current from HF Radar, Mooring and glider data for Monterey Bay, wind products from COAMPS model at different resolutions, from MM5 for Southern California, and from RAMS for Prince William Sound. 2) Produce value-added products by combining multiple data sources. For example, we provides real-time blended wind product for the Northeast Pacific by combining the QuikSCAT wind and the COAMPS simulated wind. We also provide real-time wind plots and animation in the Monterey Bay area using analysis and forecast wind product from COAMPS and measurements from moorings and coastal stations. 3) Supply various methods to access and visualize the data and model output. For 3D model output, we use Live Access Server (LAS) to subset and visualize the data on-the-fly. For the observation data, we serve pre-generated plots together with the raw data. In some cases, we provide custom plotting capability for user to select a subset of the data or to choose their own plotting parameters. 4) Provide on-demand modeling capability for user to configure his/her own ocean model, submit and monitor the job status, and finally retrieve and visualize the model output.

2. OUROCEAN ARCHITECTURE

OurOcean consists of a data server, a web server, a visualization server, and an on-demand server. The data server is in charge of real-time data retrieval and processing. Currently, our data server manages a MySQL database and a 5 terabyte RAID disk. The web server is an apache2 server with Tomcat running on a Linux workstation. The web
server dispatches user requests to the visualization server and on-demand server to generate custom plots or invoke on-demand modeling. The visualization server consists of a set of plotting programs written in GMT and Matlab. Finally, the on-demand server manages the custom model runs and the computing resources. We have a 12 processor SGI Origin 350 and a 16 processor SGI Altix cluster as our modeling engines. Both machines are running the PBS batch system. Figure 1 depicts the OurOcean hardware configuration which consists of three Linux workstations and two SGI parallel computers. The three workstations are served as the web server, the data server, and the visualization server, respectively. The data server machine is also configured as a backup web server. When the primary web server fails, we can switch to the backup server without interrupting the service. In addition to the 5 Terabyte RAID disk, we have an identical RAID disk that mirrors the primary disk contents on the daily basis. The two SGI systems are mainly used for model execution. The SGI Origin 350 is dedicated to on-demand modeling and the SGI Altix system is used for other real-time and retrospective modeling tasks. The Altix machine will be the main modeling engine for the MB06 experiment during the summer of 2006.

![Figure 1. OurOcean Hardware Configuration](image)

### A. OurOcean Data Server

The data server is responsible for the retrieval, processing, and archival of real-time observation data and model output from distributed sources. The real-time observation data include global QuikSCAT wind, ocean current from HF Radar, Mooring and glider data for Monterey Bay, Sea Surface Temperature (SST) data from various satellites, such as AMSRE, GOES, MCSST, and TMI, temperature profile from the ship. The real-time modeling output include wind products from COAMPS model at different resolutions, from MM5 for Southern California, and from RAMS for Prince William Sound.

We are pulling data on hourly basis from more than ten difference sources, including COCMP (Coastal Ocean Currents Monitoring Program) website, NOAA’s Center for Operational Oceanographic Products and Services (CO-OPS), JPL’s Physical and Oceanography Data Center (PO.DAAC), Remote Sensing Systems (RSS), Monterey Marine Meteorology Division from NRL, US Navy’s Global Ocean Data Assimilation Experiment (GODAE), Monterey Bay Aquarium Research Institute (MBARI), Naval Postgraduate School, and Aerospace Corporation.

The downloaded data are mainly used for 1) data assimilation into the ocean model, 2) the forcing to drive the model, or 3) comparison with the model output. Some of the data are used to produce value-added products, such as blended wind product and real-time wind product for Monterey Bay. Most of the downloaded data are also served on OurOcean portal with pre-generated images.

Currently, our data server manages a MySQL database and a 5 terabyte RAID disk. The metadata of each data set is stored in the database while the raw data and images are stored as plain files. Some sparsely distributed point datasets, such as tagged animals or mooring data, are also stored in the database for more flexible query.

### B. OurOcean Web Server

OurOcean web server provides three major functions: 1. serve observation data products out of our data server, 2. serve the model output using Live Access Server (LAS), and 3. provide interface for user to run and monitor customized ocean models (On-Demand Modeling).

The data products pages are dynamically constructed using CGI scripts and Java Servelets based on the metadata stored in the database. The data are organized into five categories based on their locations. Under each category, there are multiple data products based on the instrument and the data source. Each data product is further divided into sub-products according to the product resolution, product type, etc. For most of the products, we serve both raw data and pre-generated images. For some of them, we also provide custom plotting tools, e.g., California Sea Lion under CIMT category.

All the 3D time-series model output are served via Live Access Server (LAS) [1]. LAS, designed by NOAA’s Pacific Marine Environmental Laboratory (PMEL), is a highly-configurable web server to provide flexible access to georeferenced scientific datasets. It allows user to subset variables and produce output in a choice of file formats, to visualize data with on-the-fly graphics, and to compare variables from different datasets. Currently, we have real-time ROMS (Regional Ocean Modeling System) model output for the Monterey Bay and Southern California Bight. We also have Pacific Model output at 50-km resolution from 1948 to 2003 and a 10-year Pacific Model surface output at 12.5km resolution.

The on-demand modeling page contains a login page, a user registration page, a model configuration page, a model status page, and administration pages to allow administrators to manage the user accounts and the user jobs. The custom model output, once generated, will be served under the “on-demand models” category in LAS.

The web server is an apache2 server with Tomcat running on a Linux workstation. The web server
communicates with the data server on another workstation to get the data information. It also dispatches user requests to the visualization server and on-demand server to generate custom plots or invoke on-demand modeling. A backup web server is running side-by-side with the primary web server and ready to take control whenever necessary.

C. OurOcean Visualization Server

We supply various methods to access and visualize the data and model output. For 3D model output, we use Live Access Server (LAS) to subset and visualize the data on-the-fly. For the observation data, we serve pre-generated plots together with the raw data. In some cases, we provide custom plotting capability for user to select a subset of the data or to choose their own plotting parameters.

The visualization server consists of a set of plotting programs written in GMT and Matlab. Currently, the plotting programs are custom built for specific datasets and specific type of plots. We are in the process of designing a more general tool with a web interface to select, combine or compare multiple data sets and/or multiple variables in one plot. We are also exploring more interactive visualization tools such as Google map and Google Earth for overlapping gridded data sets and point data sets.

D. OurOcean On-Demand Server

The on-demand server manages the custom model runs and the computing resources. The custom model includes a parallel data assimilation module and a parallel ROMS ocean model. We have a 12 processor SGI Origin 350 and a 16 processor SGI Altix cluster as our modeling engines. Both machines are running the PBS batch system. The on-demand server consists of two service modules. The account manager is responsible for new user registration, user authentication and user status update. The job controller is responsible for configuring the model run, collecting all the input data, submitting a batch job to a computing engine, monitoring the job execution, post-processing the model output, and notifying the user. The job controller manages its own job queue in order to keep track of the job status and report it to the user. Figure 2 depicts the major functions and its control flow of the On-Demand Server.

Currently, we only allow on-demand modeling on one specific domain, i.e., Monterey Bay. Eventually, we will expand this capability to cover multiple domains in the Pacific Ocean and we are exploring alternative computing resources to accommodate the increasing computing demands. We will discuss the alternative computing resources in Section 4.

3. OUROCEN APPLICATIONS

OurOcean web portal has been used for several specific applications. We will present two real-time applications here.

A. CIMT Real-time Wind

As part of the Center for Integrated Marine Technologies pilot ocean observing system project (CIMT) [3], NPS and JPL have developed an online near real-time surface wind demonstration product for Monterey Bay (http://cimt.jpl.nasa.gov). This product, aimed at the marine recreational user, shows surface winds from a mesoscale atmosphere model (COAMPS)[4] together with winds measured at four ocean buoys and seven coastal stations from various CeNCOOS partner institutions.

Figure 2. On-Demand Modeling Server

Figure 3. COAMPS Wind overlaid with Realtime Observations
Twice a day, we download the 3km resolution COAMPS 48 hour forecast output from Naval Research Laboratory in Monterey. At every hour, we download the wind data from four ocean buoys and seven land stations in the Monterey Bay area and produce the wind plot shown in Figure 3. In addition, we produce time series plots for a 96 hour time period at every observation location (Figure 4). We also animate the wind plots and display the station location and the current wind direction and speed at the station if you click on the red dot in the plot.

This demonstration shows how we combine distributed real-time observation with model output to generate value-added products.

B. MB06 Ocean Forecasting System

The Monterey Bay (MB) ocean forecasting system is based on the Regional Ocean Modeling System (ROMS) [5,6,7] using a 3-dimensional variational (3DVAR) data assimilation algorithm [8]. The ROMS configuration consists of three-level nested domains covering the U.S. West coast, central California coast, and Monterey Bay at 15-km, 5-km and 1.6-km, respectively. This MB ROMS forecasting system is developed during the 2003 AOSN (Autonomous Sampling Ocean Network) field program [9] and further improved in preparing the MB06 ASAP (Adaptive Sampling and Prediction) program [2]. Figure 6 shows a snapshot of the Sea Surface Temperature shaded relieved with Sea Surface Height from the three domains. The black box indicates the model domain for the next level and its relation location within the bigger domain.

This system is by far the most complicated operational system we have built. We run four ROMS models -- a 3-level nested ROMS with Data Assimilation that produces 24 hour nowcast output at every 6-hour interval, a ROMS forecast model that produces forecast output for the next 48 hours, two ROMS models with Tide [10], one at 1.5Km resolution and the other at 600 meter resolutions. All four models run on our SGI Altix system. The model outputs are interpolated onto a regular lat-lon grid and served on the web using LAS. In addition, we generate a set of plots on pre-defined fields including the temperature, salinity and current at sea surface, from 0 to 400 meter average, at isopynal (constant-density) surface, and an offshore cross section. Finally, we use Mangen [9] system to analyze the Lagrangian Coherent Structure (LCS) in the surface current velocity field generated by the ROMS forecast model and the ROMS 1.6Km Tide model.

All the assimilated observation data, model outputs, image products, Mangen outputs and plots are stored in the OurOcean database and then served on the OurOcean MB06 website (http://ourocean.jpl.nasa.gov/MB06). Figure 6 describes the work flow of the system with each color representing a different computer. We use five different computers, each performing a different type of operation. The independent operations are executed concurrently on different computers in order to complete the end-to-end processing within minimal time. The whole process starts at 2AM every night and the main operations finish before 6AM. The model output is available for next day’s field experiment at Monterey Bay via the OurOcean portal.

This system demonstrates how the data server, web server and the visualization server in the OurOcean portal interact with each other in a real-time operational system.
4. FUTURE WORK

We are enhancing the OurOcean capabilities to better handle the increasing size and diversity of the observation data and model output and the increasing workload on on-demand modeling. We focus our effort in two areas: more robust visualization tools and alternative computing resources.

A. Interactive Exploration and Animation

We are investigating new ways to do 3D interactive visualization of time-sequence data over the internet. This tool will allow users to browse the dataset interactively in both the spatial and temporal domains, and to animate different features of the data in a variety of ways to reveal the dynamic structure of the data. This tool should also allow users to visualize the observation data in the context of model simulations. We consider using Google Earth as the base front end visualization engine because it can smoothly pan and zoom large remote earth image and digital elevation map in 3D perspective. It can also overlay geo-referenced imagery or vector dataset using a standard language (KML) definition. Moreover, it can serve data from geographically distributed locations seamlessly. However, Google Earth is lack of capability to visualize 3D volume data and interface to animate time-sequence data. We are currently investigating various methods to represent 3D time-sequence model outputs in Google Earth.

B. High Throughput Computing

Our SGI Origin 350 is only capable of handling 5 to 10 on-demand modeling requests per day. We will need to pursue alternative computing resources when the modeling demand increases. A potential candidate for such an inexpensive but powerful system is Condor.

The Condor High Throughput Computing Software (Condor) [11] is an established open-source system that enables scientists to perform large-scale computations by harnessing the power of available computers, including both dedicated rack-mounted compute clusters as well as non-dedicated systems distributed across a classroom, a campus, or across the world in a grid environment. The Condor software includes tools for the scheduling and management of serial jobs, tightly-coupled parallel jobs using libraries such as MPI and PVM, and entire workflows containing complex interdependencies encompassing both computation and data placement.

There are over a thousand Condor pools over the world with more than 50,000 computers contributing their resources. If we can use such abundant computing power for the on-demand modeling, we will be able to handle hundreds of requests per day. However, it requires extensive work in configuring the model jobs, scheduling and managing distributed tasks and reliably transferring data back and forth.

5. CONCLUSION

In the last three years, OurOcean has evolved from a simple portal serving a limited number of ocean data products to today’s multi-functional, distributed computing and data management server. Our goal is to provide an integrated solution for ocean monitoring and forecasting. We will continue develop new products and incorporate new technologies into OurOcean to make it more versatile and user-friendly for scientists, educators, students, and general public.

ACKNOWLEDGMENT

We would like to thank all our collaborators who supply us with real-time observation data, including Dick Lind, Jim Doyle, Francisco Chavez, Arlene Kishi and Jeff Paduan. We also like to thank Leslie Rosefeld, Todd Anderson, and Rondi Robinson for their help on the construction of the CIMT real-time Wind products; Shawn Shadden and Francois Lekien for setting up Mangen software for the MB06 experiment.

The research described in this publication was carried out, in part, at the Jet Propulsion Laboratory (JPL), California Institute Technology, under a contract with the National Aeronautics and Space Administration (NASA). Funding by the Office of Naval Research (ONR) through a subcontract from the Monterey Bay Aquarium Research Institute (MBARI) is also acknowledged.

REFERENCES


