

# Data Management and Real-time Distribution in the HF-Radar National Network

E. Terrill<sup>1</sup>, M. Otero<sup>1</sup>, L. Hazard<sup>1</sup>, D. Conlee<sup>2</sup>, J. Harlan<sup>3</sup>, J. Kohut<sup>4</sup>, P. Reuter<sup>1</sup>, T. Cook<sup>1</sup>, T. Harris<sup>1</sup>, K. Lindquist<sup>5</sup>

<sup>1</sup>Coastal Observing R&D Center, Marine Physical Laboratory  
Scripps Institution of Oceanography  
9500 Gilman Drive #0213  
La Jolla, CA 92093-0213

<sup>2</sup>DOC/NOAA/National Weather Service  
National Data Buoy Center  
1100 Balch Boulevard  
Stennis Space Center, MS 39539

<sup>3</sup>National Oceanic and Atmospheric Administration  
HF Radar Ocean Remote Sensing  
NOS/Center for Operational Oceanographic Products and Services  
Boulder, CO 80303

<sup>4</sup>Rutgers University  
Coastal Ocean Observation Lab  
Institute of Marine and Coastal Sciences  
New Brunswick, NJ 08901

<sup>5</sup>Lindquist Consulting Incorporated  
59 College Rd. Suite #7  
Fairbanks, AK 99701

*Abstract*—The architecture, status and applications of a real-time data access, distribution, processing and storage system designed for networking radial data from surface current mapping HF-Radar instruments across the United States is presented. By leveraging the system design of HF-Radar sites, data access is generalized to nearly all sites while still providing alternate access options where needed. Data format convergence, while not required, is achieved for data from all systems through careful metadata mapping and code development. Object ring buffers (ORBs) and ORB communication protocol provide robust and flexible data transport while a relational database facilitates data storage.

The HF-Radar Network has evolved from a prototype project to an operational status over the last 2.5 years with 4 data access sites (portals) and 1 data aggregation site (node) deployed. By early 2007, an additional portal and 2 additional nodes will be added to create a distributed network. To date, the repository contains over 356,000 radial files produced by 45 sites from 10 participating institutions.

Recent development has focused on real-time total vector processing on a national scale. Base grids for the U.S. West and East/Gulf Coast of 1 km nominal resolution extending 300km offshore are created using an equidistant cylindrical projection. A community standard MATLAB toolbox for total vector processing is optimized for production on large grids and integrated into the real-time system to produce hourly surface current maps on a national scale at 1 km, 2 km and 6 km resolutions.

Current applications of the HF-Radar network include an interactive radial diagnostic site for use by site operators and a prototype interactive web site providing the first images of real-time surface currents integrated across a national scale.

## I. INTRODUCTION

Local, state, regional, and federal discussions directed towards the establishment of an Integrated Ocean Observing System continue to emphasize a desire for the installation, development, and operation of a network of surface current mapping systems for use by a broad range of end users. Central to the operational success of a large scale network will be a scalable data management, storage, access, and delivery system. The system under development with collaborators from the National Oceanic and Atmospheric Administration (NOAA), National Data Buoy Center (NDBC) and National Ocean Service (NOS) builds upon a prototype software architecture developed with funding from the National Science Foundation (Real-time Observatories, Applications, and Data Management Network, ROADNet). The architecture of the HF-Radar Network lends itself well to a distributed real-time network and may serve as a model for networking sensors on a national level. This joint university-NOAA partnership is focused on defining and meeting the

expressed needs for an IT infrastructure supporting a national network of surface current mapping systems.

Surface current velocity measurements collected by HF-Radar benefit from aggregation and integrated processing of data from geographically distributed sites. The network of surface current mapping systems is characterized by a tiered structure that extends from the individual field installations to regional operations maintaining multiple sites and on to aggregation sites obtaining data from all regions. The data system development effort focuses on building robust data communications from remote field locations (sites) for ingestion into the data system via data portals. Portals are computer systems enabled with the Antelope Real-Time System (ARTS), allowing the acquisition, transfer, buffering and delivery of data. Once surface current data is within the ARTS framework, it is buffered and transported through Object Ring Buffers (ORBs), a set of code specific to real-time data delivery. Each portal is designed to interact with any number of data repositories (nodes) which collect data from any number of regional portals. A data system built around the concept of a distributed network provides redundancy by allowing multiple locations to house data while addressing throttling issues during high usage periods. Aggregation of surface current data across regional associations enables integrated total vector processing on large scale national grids resulting in products for ingestion by tools such as U.S. Coast Guard search and rescue models, ocean circulation models, ecosystem management, public health and water quality warning systems.

A pilot program, involving both the Southern California Coastal Ocean Observing System (SCCOOS) and the Central and Northern California Ocean Observing System (CeNCOOS), through the California wide Coastal Ocean Currents Monitoring Program (COCMP) will be implementing this data management system throughout California. The scope of the pilot program is greatly expanded by partnering with Rutgers University and the NDBC. The growth of the program has resulted in data contributions from 45 sites across 10 institutions around the country.

In this paper, details on the HF-Radar Network architecture and total vector processing are presented along with operational and prototype applications. Active and planned development efforts for the network are also discussed.

## II. HF-RADAR NETWORK ARCHITECTURE

### A. *Antelope Real-time System*

Antelope, a product of Boulder Real Time Technologies (BRTT, [www.brтт.com](http://www.brтт.com)), is an integrated collection of programs for real-time data collection of environmental monitoring information. Through its open system design principles and modular approach, a system can be developed to provide specific solutions to common real-time data problems related to reliable acquisition, distribution, processing and storage.

Antelope has been developed and refined for over a decade through BRTT's mission to provide software support to operational seismic networks worldwide. The real-time system is build around large, flexible, non-volatile object ring buffers (ORBs). Datascope, the relational database, provides a bridge between real-time processing at the ORB level and long-term storage to a local database. Because Antelope provides a flexible, robust and scalable solution to real-time processing, it is used at the core of the HF-Radar Network.

### B. *Portals and Nodes*

From a broad perspective, the HF-Radar Network architecture is comprised of two building blocks, portals and nodes, each representing Antelope enabled computers with distinct roles. Portals serve as 'point of entry' machines by acquiring and serving radial data from any number of HF-Radar sites. Nodes serve as data concentrators by collecting radial data from any number of portals (or nodes). This design minimizes data requests through sometimes unstable network connections to individual sites by serving data through portals while maintaining a high degree of network flexibility through selective data collection at nodes.

### C. *Portal Design Elements*

Portals utilize two executables within the Antelope framework, named `hfradar2orb` and `orbserver`, for acquiring and serving radial data, respectively (Fig. 1). The `orbserver` executable manages an object ring buffer (ORB) and honors requests to accept packetized data and/or deliver copies of previously received packetized data. Each packet may consist of any desired binary information (therefore including ASCII). Packets taken into the `orbserver` are preserved and reproduced without modification. The `hfradar2orb` executable was written specifically for the HF-Radar Network. It acquires radial files, either from a local intake directory or from a remote site over SSH protocol, encapsulates each file as an ORB packet, and puts the packet on a specified `orbserver`. Two of the most notable features contributing to network scalability and high data integrity are data access over SSH protocol and on-the-fly file content checking and conversion.

Data access over SSH protocol evolved primarily from the need to relieve host sites of any executable code (and associated maintenance) needed to make data available to the HF-Radar Network. By leveraging the system design of HF-Radar sites, data access was generalized to all standard sites through secure requests. For non-standard sites, the only requirements for remote data access using `hfradar2orb` are access over SSH and a real-time-directory (i.e. a single static path where recent radial files are available). For any site that does not satisfy remote data acquisition requirements, alternate custom arrangements may still be made with ingestion by `hfradar2orb` ultimately taking place from a local directory. Eliminating code external to the portal and implementing remote data acquisition through `hfradar2orb` resulted in increased data continuity and scalability as

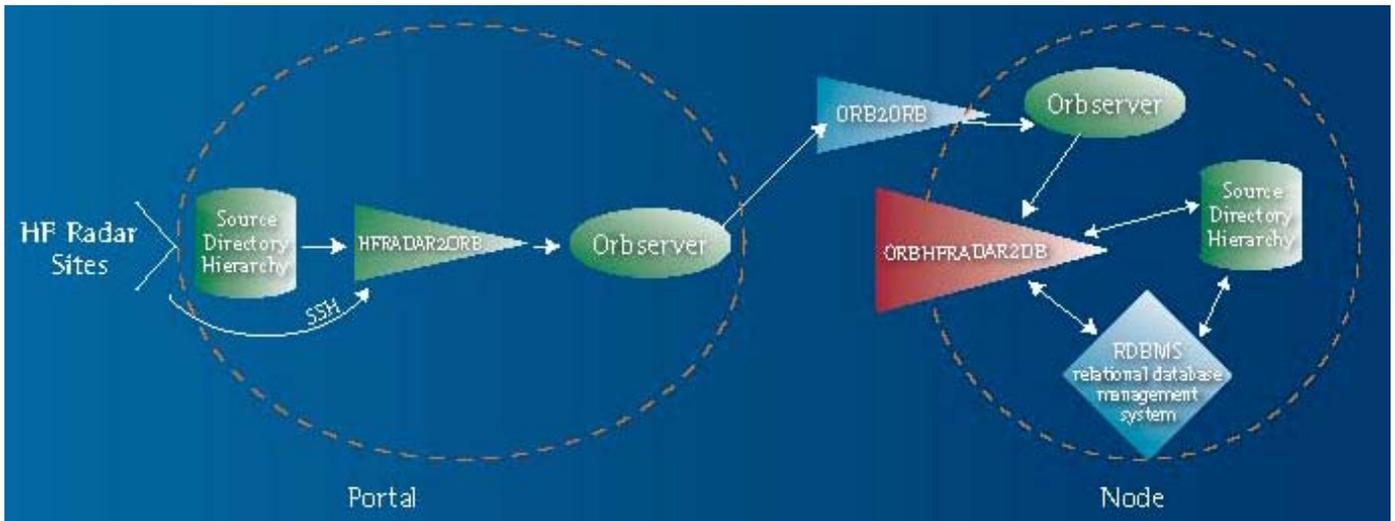


Figure 1. Schematic depicting data flow through the HF-Radar Network and the relationship between a portal and node. Data is acquired by hfradar2orb, either over SSH protocol from a remote location or from a local source directory, and then placed on the orbserver. The node transfers packets served out by the portal orbserver into its own orbserver via orb2orb. Then orbhfradar2db unwraps packets, stores file metadata (path, filename, arrival time, etc.) to the relational database management system and stores files to a directory hierarchy.

observed by reduced gaps in site records and increased speed of network growth.

Throughout the growth of the radial data repository in the HF-Radar Network, careful note has been taken of variations in radial file formats so that information can be extracted in a consistent and controlled manner by downstream processing. File formats fall into two broad categories known as range-bin and LatLonUV (LLUV). CODAR Ocean Sensors SeaSonde systems originally produced files in range-bin format which is currently being phased out in favor of the newer LLUV format. Minimal documentation is available for the range-bin format and considerable variation is observed in essential metadata fields such as the timestamp and site location. Conversely, the newer LLUV format has clear documentation and consistent formatting.

As the HF-Radar Network evolved, it became apparent that conversion of range-bin format files to LLUV format would streamline all downstream processing. Through careful metadata mapping and code development against 269,000 radial files produced by 35 CODAR SeaSonde systems from 8 institutions located throughout the United States, a Perl module (codartools) was written to convert range-bin format files to LLUV and verify LLUV format file contents. The codartools module was then integrated into the Antelope environment through hfradar2orb for on-the-fly conversion of any range-bin files encountered and verification of file contents prior to placement in the ORB. These upgrades resulted in a cleaner development environment for downstream processing by providing unambiguous file formats and eliminating corrupt files from entering the system.

#### D. Node Design Elements

Once files are ingested by hfradar2orb and made available through the orbserver on a portal, it is up to the node(s) to retrieve packets and unpack them into the local Datascope

database. Nodes accomplish this through three executables, orb2orb, orbserver and orbhfradar2db (Fig. 1).

Nodes run orbserver, as portals do, and the orb2orb executable connects the two orbserver by transferring specified packets of interest from the source orbserver (the portal orbserver) to the destination orbserver (the node orbserver). When packets of interest arrive on the source orbserver, they are copied reliably, without modification, and immediately to the destination orbserver. Since data transfer within the Antelope framework is accomplished via orb2orb and orbserver, a node can be used as a source orbserver to another node, and vice versa. This may be done to reduce load on portals or otherwise distribute network traffic.

Packets from the node orbserver are unwrapped and put into a relational (Datascope) database by orbhfradar2db, an executable written specifically for the HF-Radar Network. Datascope relational database tables are stored as ASCII text files, making them readable by standard UNIX tools. Datascope is able to store files external to its tables which allows files to be accessed with or without the Datascope database layer. The current usage of Datascope within the HF-Radar Network stores radial files externally so that the native data format is preserved and files remain accessible outside of Datascope. Paths to the radial files, record modification times, data timestamps, site information and other metadata are all stored within Datascope tables by orbhfradar2db as packets arrive in the local ORB.

#### E. Network Status

The HF-Radar Network started as a prototype at Scripps Institution of Oceanography with a single portal and node and 4 sites in December 2003 and has since grown to an operational status with over 356,000 radial files produced by 45 sites from 10 participating institutions (as of July 2006). Three additional portals have been deployed at Rutgers

University, University of California at Santa Barbara and San Francisco State University to serve radial files from their local regions. The node at Scripps is now used for operational production of radial diagnostics used by HF-Radar site operators. The network has proven to be stable and robust through network outages, power failures and system upgrades. Under normal circumstances, radial files are often available at the node within minutes of being produced at remote sites or being made available to the portal.

### III. REAL-TIME TOTAL VECTOR (RTV) PRODUCTION

#### A. Wide Area Grid Development

Coastal grids have been developed around the United States in order to seamlessly integrate radial data for total vector production on a nationwide scale. Due to the large spatial extent of these regions (along coast distances up to 3000 km) and relatively high resolution of HF-Radar data (1 – 6 km), careful consideration needed be given to the advantages and disadvantages of grid generation methodologies. The adopted grid aims to preserve data integrity as well as provide a practical format for data dissemination.

Initially, an equal area grid based on the WGS84 ellipsoid was generated for the U.S. West Coast (Fig. 2). However, the resulting grid was not orthogonal throughout. It forms 1 km

squares around the central longitude but forms parallelograms with increasing tilt as distance from the central longitude increases. While properties of equal area and constant resolution retain the native resolution of data throughout the grid, it complicates data dissemination. Dissemination is mainly complicated by the fact that the grid's latitude and longitude pairs are determined by converting distances to latitude and longitude coordinates based on a stepwise approach originating from a center location on the WGS84 ellipsoid making it difficult for users to modify, re-create or interpolate the grid. Ease of interpolation is an important factor for dissemination of spatial velocity data since products will likely be based upon interpolations of the data. For example, since the eastward ( $u$ ) and northward ( $v$ ) components of the total vector will be co-located on the grid ("A grid") interpolation onto a staggered grid ("C grid") may be done for calculating geophysical parameters such as vorticity and divergence. Other examples include interpolation onto various grid types for model ingestion.

Problems associated with equal area grid cells have been encountered within the ocean modeling community which has adopted approaches that can be extended to developing wide area HF-Radar grids. The most common approach is to develop a grid based upon constant latitude and longitude spacing using an equidistant cylindrical projection. The resulting grid will have variable resolution in longitude that increases poleward, but constant resolution in latitude. In return for sacrificing constant resolution and equal area, orthogonality is gained in addition to a grid with very simple construction and interpolation (Fig. 2).

Nominal 1 km grids based on equidistant cylindrical projections were developed for the U.S. West Coast and East/Gulf Coast extending 300km offshore. Grid points falling over land or within 0.5 km were removed using polygons produced from the World Vector Shoreline database available through the National Geophysical Data Center (<http://rimmer.ngdc.noaa.gov/mgg/coast/getcoast.html>).

#### B. RTV Processing

HFRadarmap, a MATLAB toolbox for processing HF-Radar data, is a commonly used program within the community for a wide variety of applications ranging from research to real-time processing. HF-Radar processing continues to be an active area of research and MATLAB is used because it provides a robust platform for operational processing as well as the flexibility needed to develop new products and processing techniques. A new version of the toolbox, currently under development at the Naval Postgraduate School, is being incorporated into a broader suite of programs geared towards real-time processing collectively called HFR\_Progs.

Using a beta-version of the MATLAB toolbox in HFR\_Progs, modifications have been made to optimize performance for large national grids and for integration with the Antelope Toolbox for MATLAB (ATM). The ATM is a set of MATLAB tools, delivered with Antelope, that allows

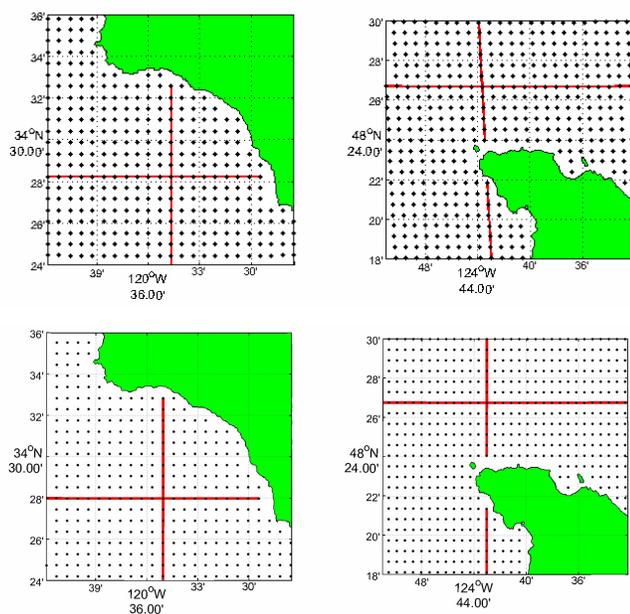


Figure 2. Images from regions near the central longitude (top left) and far from the central longitude (top right) of an equal area grid show how grid cells change shape from squares to parallelograms. Along lines of latitude, grid points have constant values. However, along lines of longitude, grid points only have constant values along the central line of longitude. The grid points along the central longitude and an arbitrary latitude are highlighted on the top left. The image on the top right is from a region approximately 4 degrees away from the central longitude where a pair of arbitrary grid lines along latitude and longitude are highlighted to show their non-orthogonal nature. Images from the same region shown in the upper panels are shown in the lower panels for a grid based on an equidistant cylindrical projection. Arbitrary grid lines highlighted in red show orthogonality is preserved throughout the grid.

access to Datascope relational database tables. The resulting modified toolbox brings real-time total vector processing under the management Antelope real-time utilities.

Real-time total vector processing is currently done on an hourly basis for both the U.S. West Coast and U.S East/Gulf Coast. Since sites don't always report at the same time, re-processing of total vectors is often required as lagging sites report older data. However, in order to protect against large back-processing jobs, throttling is built in to total vector processing limiting the oldest dates allowed to be processed. Based on network downtime latencies, processing constraints and real-time data relevancy, 26 hour back-processing limits are currently used. An important note is that the main objective of RTV processing is to produce the best real-time total vector data available at the time of processing. These real-time products are not necessarily research level products. However, research level refined total vector products can be produced outside of real-time processing within the system.

Total vectors are processed on three grid resolutions for both coasts to accommodate for the range of resolutions in radial data. Systems greater than approximately 24MHz, 13MHz and 5MHz contribute to totals on 1 km, 2 km and 6 km grids, respectively. These hourly total vectors form the

basis for further product development. Hourly 25-hour averages, computed from the hourly total vector products, are also in real-time production for both grids at all three grid resolutions. Merging all grid resolutions in to a single product for both hourly and averaged totals is currently in development.

#### IV. CURRENT HF-RADAR NETWORK APPLICATIONS

##### A. Web Accessible Radial Diagnostics

The first application of the HF-Radar Network was to develop an on-line radial diagnostic utility for use by HF-Radar operators (<http://cordc.ucsd.edu/projects/mapping/>). Through the diagnostic utility, site operators are able to check the status and quality of radial data from each site contributing to the HF-Radar Network. The most recent radial file in the system is displayed along with metadata pertaining to the site and file format (Fig. 3). User selectable time histories and statistical summaries are available for data latencies, maximum range of radial data and the number of radial solutions. In addition to providing utilities for monitoring data integrity, active development will lead to interactive tools for system diagnostics such as module temperatures and forward and reflected power. The entire diagnostic utility

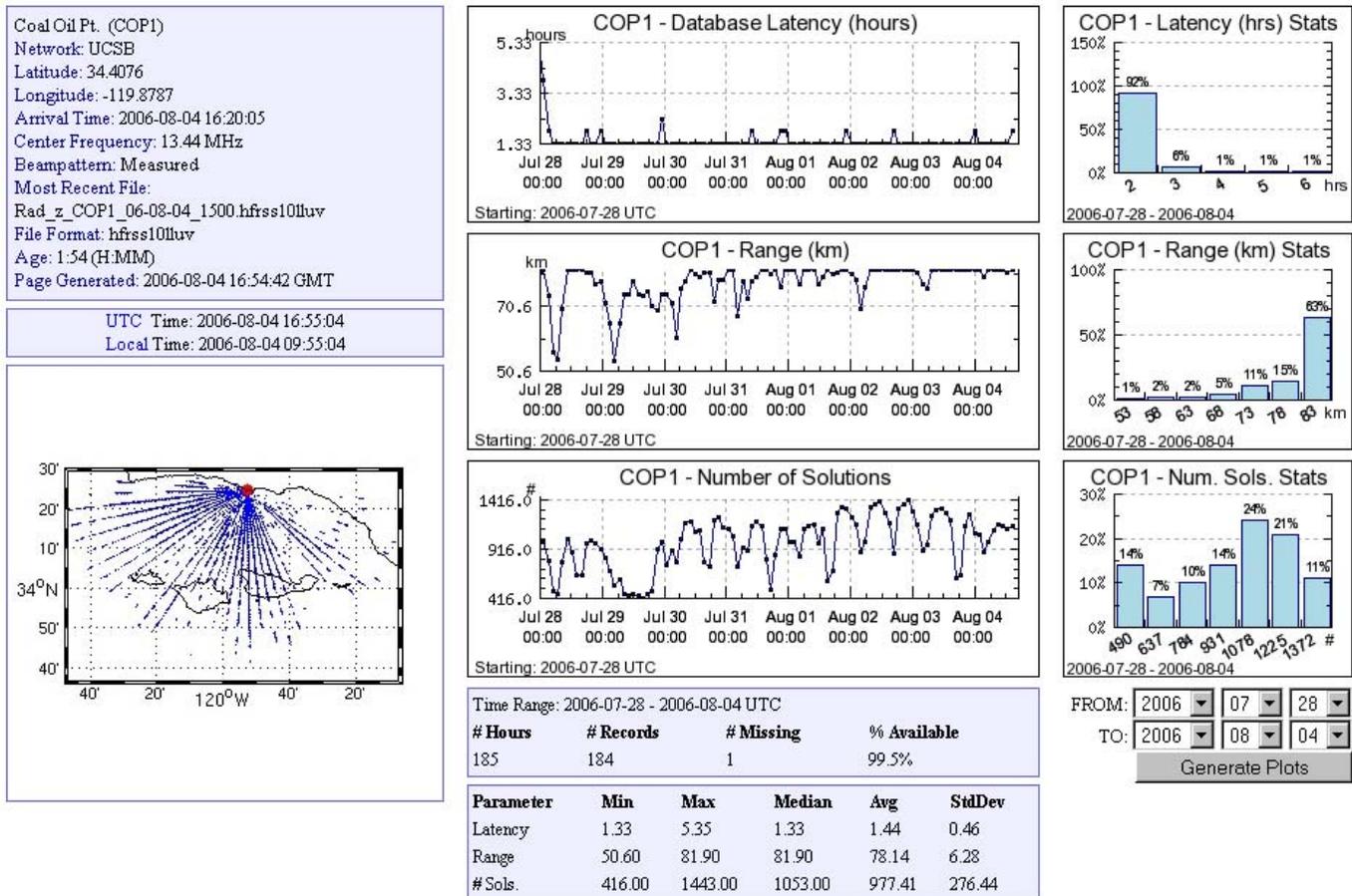


Figure 3. Screenshot from the radial diagnostic web site (<http://cordc.ucsd.edu/projects/mapping/>) of diagnostic information for a U.C. Santa Barbara radar installed at Coal Oil Point, COP1.

directly accesses the Datascope database creating an on-line extension of the HF-Radar Network.

### B. RTV Interactive Web Interface

A prototype interactive web site for display of hourly and 25 hour averaged RTV products produced from the HF-Radar Network is providing the first images of real-time surface currents integrated across a national scale (<http://cordc.ucsd.edu/projects/mapping/rtv/>, Fig. 4). The last four days of RTV products are currently available for both U.S. West and East/Gulf Coast grids at 1 km, 2 km and 6 km resolutions. As with the radial diagnostic utility, on-line RTV product utilities will become closely integrated with the real-time nature of the Datascope database as the network develops.

## V. SUMMARY

The HF-Radar Network has developed into a data management utility that will serve as a working model to other

real-time data streams. Although still in development, its present working state provides an end-to-end solution to real-time data access, distribution, processing and storage. New developments will integrate Quality Assurance and Quality Control (QA/QC) into real-time processing and expand the database schema to include metadata and system diagnostics available within radial files. An additional portal will be deployed through NDBC to the Southeast U.S. by 2007 and nodes will be deployed to both NDBC and Rutgers University creating a distributed network.

Products derived from the HF-Radar Network already provide important information to HF-Radar site operators and will eventually provide critical information up through end user applications. At the end user level, integrated surface current information will become a valuable resource for combating the spread of pollution, improving navigation safety, search and rescue operations, fisheries management and oil spill cleanup efforts.

## ACKNOWLEDGMENT

The Coastal Observing Research and Development Center acknowledges the earlier support of the National Science Foundation (NSF/ATM 03-25963) and the present funding from NOAA/NDBC (NA17RJ1231).

The SCCOOS deployment of the HF-Radar data management system in Southern California is supported by both the State of California via the COCMP program and through the NOAA Coastal Ocean Technology System (COTS) program.

The authors would like to acknowledge contributions of radial data made throughout the development of the HF-Radar Network from the following participants: Rutgers University, U.C. Santa Barbara, Naval Postgraduate School, U.C. Santa Cruz, Universidad Autonoma de Baja California, Texas A&M, University of Maine, San Francisco State University, University of Southern California and the NOAA/NOS/Integrated Observations Team.

Thanks to Mike Cook at the Naval Postgraduate School for providing a beta-version of HFR\_Progs and Bill Rector at CODAR Ocean Sensors for clarification on radial data formats.

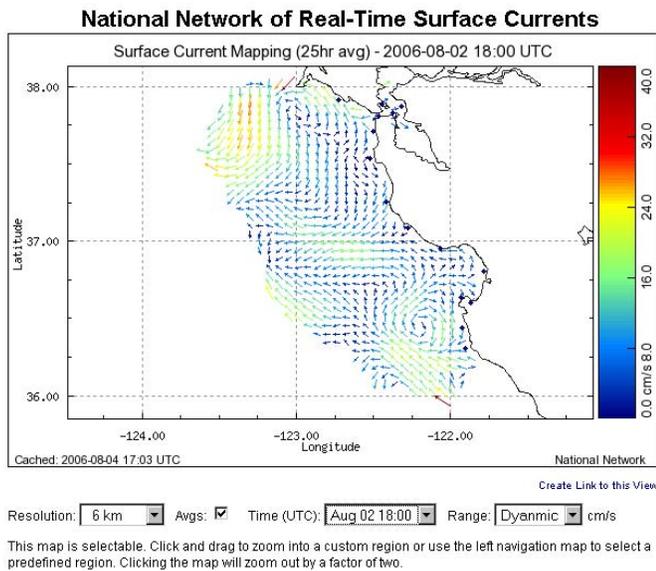


Figure 4. Screenshot from the prototype RTV interactive web site (<http://cordc.ucsd.edu/projects/mapping/rtv/>) of 25 hour averaged flow at 6 km resolution between Monterey Bay and Pt. Reyes.