Matlab stoqstoolbox for accessing in situ measurements from STOQS

MBARI Summer Internship Project 2012

Francisco Lopez Castejon

(Mentor: Mike McCann)

INDEX

Abstract

This work presents the stoqstoolbox, an extension to Matlab that simplifies the loading of in situ measurement data directly from STOQS databases. STOQS (Spatial Temporal Oceanographic Query System) is a geospatial database tool designed to provide efficient access to data following the CF-NetCDF Discrete Samples Geometries convention. Data are loaded from CF-NetCDF files into a STOQS database where indexes are created on depth, spatial coordinates and other parameters, e.g. platform type. STOQS provides consistent, simple and efficient methods to query for data. For example, we can request all measurements with a standard_name of sea_water_temperature between two times and from between two depths. Data access is simpler because the data are retrieved by parameter irrespective of platform or mission file names. Access is more efficient because data are retrieved via the index on depth and only the requested data are retrieved from the database and transferred into the Matlab workspace.

Applications in the stoqstoolbox query the STOQS database via an HTTP REST application programming interface; they follow the Data Access Object pattern, enabling highly customizable query construction. Data are loaded into Matlab structures that clearly indicate latitude, longitude, depth, measurement data value, and platform name. The stoqstoolbox is designed to be used in concert with other tools, such as nctoolbox, which can load data from any OPeNDAP data source. With these two toolboxes a user can easily work with in situ and other gridded data, such as from numerical models and remote sensing platforms.

Introduction

In situ measurement problem

The in situ measurement collected by platforms as AUV or Glidder that allow to sample during a long time and the development of new sensors, working with a high sampling frequency has allowed oceanographers understand better the ocean dynamics and have new research goals. But usually is very difficult to work with the *in situ* measurement collected for several reason:

How many are?

The number of data collected during one campaign could be huge, for example, during the last CANON (Controlled, Agile, and Novel Observing Network) campaign (http://www.mbari.org/canon/) did by MBARI (Monterey Bay Aquarium Research Institute) in Monterey Bay was collected more than 80,000,000 millions of data. This campaign has been done during the last 4 years, so the total number of measurements did are about 360,000,000 millions.

Where are?

Each dataset did are save by a different research, with different saving protocol. For example, some data will be save in a personal folder in a personal computer and other in a database develop only for a certain type of data. So sometimes is very difficult to find a dataset save by other person some time ago.

What are the format?

There are different formats to save a dataset, ASCII, excel, netcdf, … each research use a different one, so before start to work I must be able to open the file and access to the data.

How could I query them?

Some data formats allow querying them in an efficient way using some software. For example using Matlab a netcdf file can be query and uses only the data needed. But format used commonly as ASCII or Excel, don't allow making any query on them.
So usually to look for a dataset could be a very hard task.

STOQS (Spatial Temporal Oceanographic Query System (STOQS))

In order to resolve all the problems exposed in *in situ* measurement management a project started in MBARI in 2009 finishing in the developing of STOQS. STOQS is a geospatial database web application designed for providing efficient access to *in situ* oceanographic measurement data across any dimension. Where "dimension" is considered in the broadest sense, for example: any spatial dimension, time, the collection of platforms, any other parameter value in the database. STOQS and stoqstoolbox are open source software projects supported by MBARI and the David and Lucile Packard foundation. For more information please see http://code.google.com/p/stoqs.

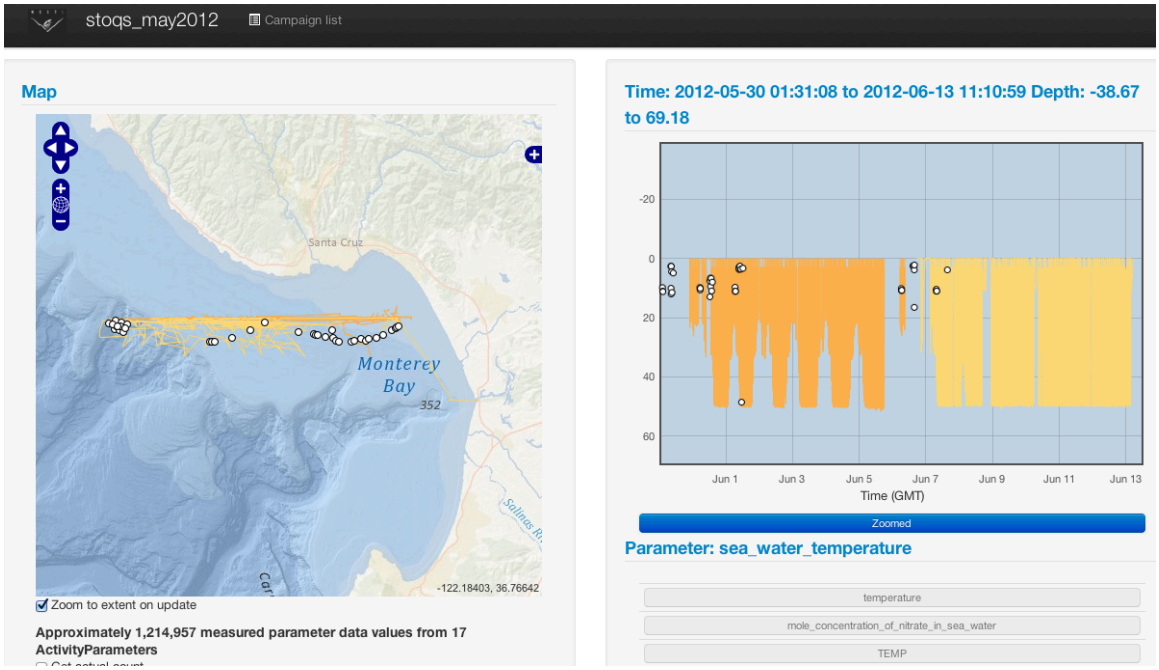


Figure 1 Screencapture of STOQS web application.

STOQS in Matlab

The were two reason to work developing another way to acces STOQS rather than the web application, the first one was to have a less friendly but more efficiency way to do a lot of different kinds of queries and the second was to allow the user to have all the data in one workplace (*in situ* measurement, gridded data,…). In the very beginning of the project we planned to develop it in the programming language Python, but finally we did it in Matlab, cause it's a well-known and widely used software, with a small learning curve.

Stoqs_toolbox

Stoqstoolbox is an opensource Matlab toolbox that provides access to STOQS and query for a dataset base in certain constrains. It can be download as a zip file from http://code.google.com/p/stoqs/downloads/list

Installation

In order to use it you have to unzip the downloaded file and add it to your Matlab path. You will need to install an open-source implementation of a JSON encoder and a decoder/parser for MATLAB/Octave called JSONlab, that can be download from http://www.mathworks.com/matlabcentral/fileexchange/33381. Once added both to you Matlab path, you will be available to use stoqstoolbox.

Stoqstoolbox validation

To ensure the correct working of stoqstoolbox, we made two comparison between the result obtained by the web interface and the toolbox.

Comparing trajectories from STOQS webinterface ouput with stoqstoolbox out for the same query

We query for all the data between 4- 6 m and 2011/06/20 09 - 2011/06/21 09 for the temperature_sea_water variable.

STOQSTOOLBOX

[query,d]=model_vs_stoqs('http://ourocean.jpl.nasa.gov:8080/thredds/dodsC/MBNowcast/mb_das_2011062021.nc',5,1,12,'');

STOQS WEBINTERFACE QUERY

```
http://odss-staging.shore.mbari.org/canon/stoqs_june2011/query/csv?start_time=2011-06-20+08%3A48%3A09&end_time=2011-06-21+09%3A02%3A10&min_depth=4.02&max_depth=6.01&parameters=sea_water_temperature
```

Number of data retrieve

STOQS WEBINTERFACE	STOQSTOOLBOX
7591	7586

Figure 2 The white line is the trajectory get from the STOQS webinterface. The dots are the points get with stoqstoolbox.

Comparing temperature from STOQS webinterface ouput with stoqstoolbox out for the same query

We will do the same query for sea water temperature available in STOQS for the same time and depth:

Date: `'01 Jun 2012 03:00:00' to'03 Jun 2012 15:00:00'`

Depth : 4-6 meters

STOQS CSV QUERY

http://odss-staging.shore.mbari.org/canon/stoqs_may2012/query/csv?start_time=2012-06-01+03%3A00%3A00&end_time=2012-06-03+15%3A00%3A00&min_depth=4&max_depth=6&parameters=sea_water_temperature

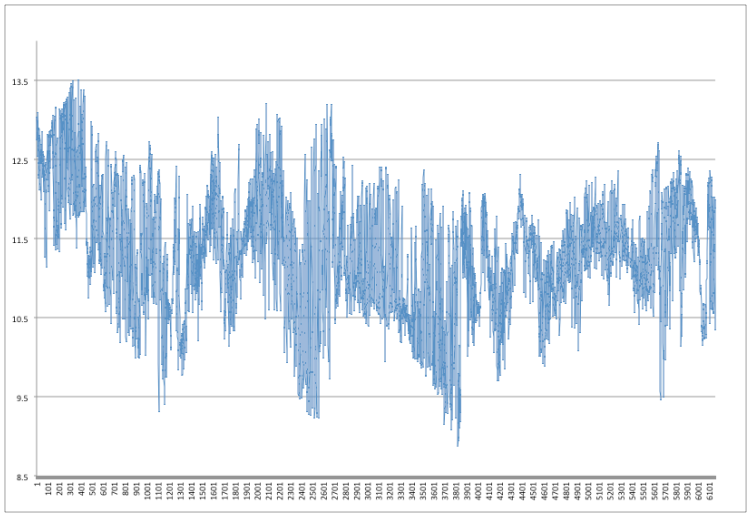


Figure 3 Temperature for the CSV query did to STOQS by the web interface.

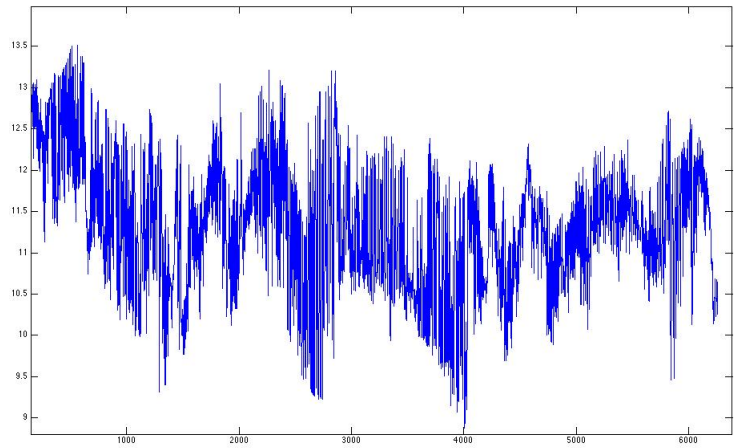


Figure 4 Temperature obtained for the JSON query made with stoqs_toolbox.

Stoqs_toolbox function

This toolbox have small program to do very clear task, for example query for the campaign availables in the server. In this way the users is the one who decides

how to use this toolbox and write their own program. In the program *stoqs_demo* are show some of the capabilities of stoqstoolbox

Stoqs_demo

Running it from Matlab will show how to query to get all the information available on a STOQS server and how to download data from it to Matlab.

stoqs_showall

```
function stoqs_showall(u)
%Show on the screen all the main information in a STOQS server
%
%
%Usage:
%
%    stoqs_showall('http://odss-staging.shore.mbari.org/canon');
%Input :
%
%   u = Url direction of the STOQS data server. Ex:
http://odss.mbari.org/canon
%
%
%
%   Francisco Lopez & Mike McCann & Brian Schlining
%
%   Last modified
%   19/August/2012
```

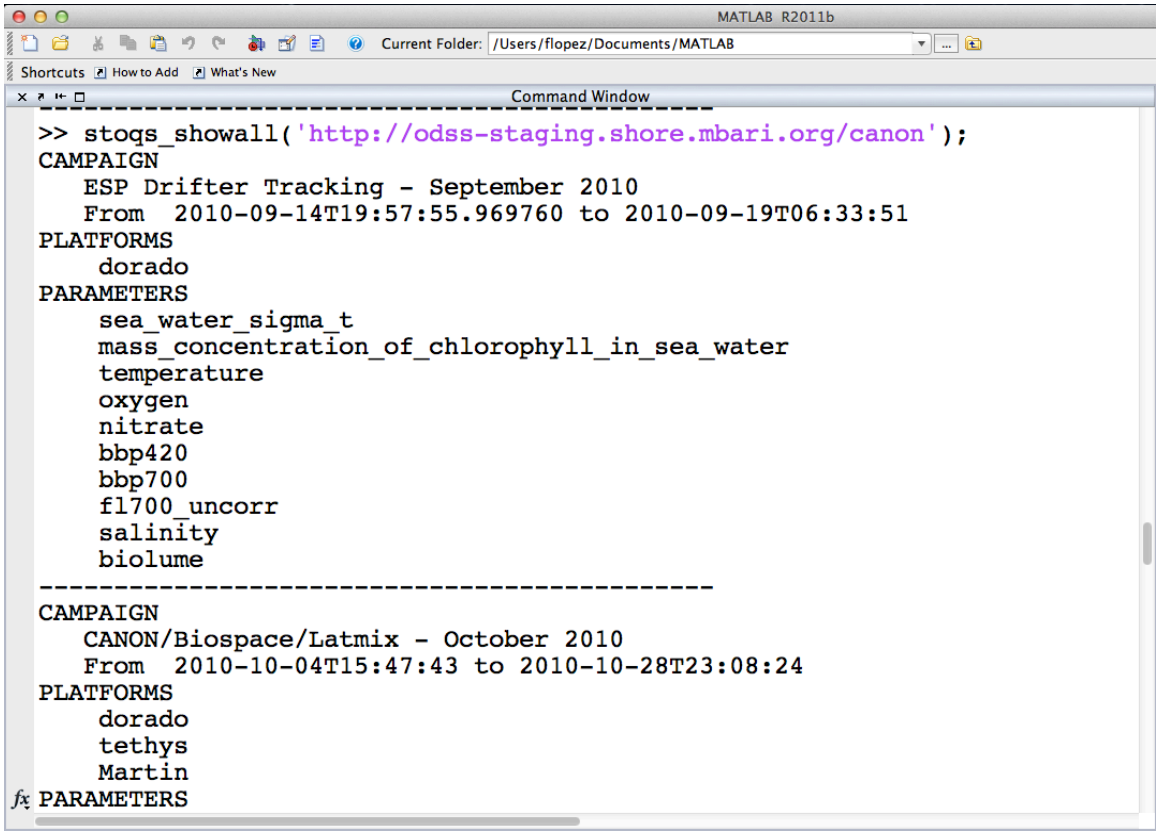


Figure 5 Out from the program stoqs_showall.

stoqs_down

```
function outp=stoqs_down(varargin)
%       Usage
%            Acces to a STOQS server and download a dataset based
in a query.
%
%            Time Queries:
%                The user could write a begging and end date. If,
for example, he doesn't write  a end date, the program will query
to all the data available from the begging date. In the same way,
if the query only have a end date, the program will query to all
the data available until that date.
%     Ex. d=stoqs_down('http://odss-
staging.shore.mbari.org/canon/stoqs_may2012','2012-05-30
00:10:00','2012-05-30
01:10:00','1','2','sea_water_temperature','dorado');
%      Will query to all the data available of sea water
temperature measured by dorado between  2012-05-30 00:10:00 and
2012-05-30 01:10:00 between 1 - 2 meters.
%            Depth Queries:
%                The user could write a top and low leve to
query. If there is not any depth, the program will query for all
the depths. If write the minimum depth, it will search for the
minimum to the maximum depth available,and in the same way  with
```

```
the maximum.
%    Ex. d=stoqs_down('http://odss-
staging.shore.mbari.org/canon/stoqs_may2012','2012-05-30
00:10:00','2012-05-30
01:10:00','','3','sea_water_temperature','dorado');
%      Will query to all the variables measured by  dorado  up 3
m depth between '2010-10-27 21:00:00'and '2010-10-29 23:00:00'
%
%            Variable Query:
%                 MUST use the standard name of he variable
following the CF-Metadata conventions (http://cf-
pcmdi.llnl.gov/documents/cf-standard-names/standard-name-
table/19/cf-standard-name-table.html).
%     In the case you want to get all the variable, leave it
empty
%
%           Platform Query:
%       Name of the platform to query, If is empty will query for
all the platform.
%
%
%
%       Input
%          The input in order are:
%           URL of the STOQS server
%           Start time in the format 'yyyy-mm-dd+HH:MM:SS'
%           End time in the format 'yyyy-mm-dd+HH:MM:SS'%
%           Minimum depth
%           Maximum depth
%           Parameter to get the data. You must use the standard
name of
%                 the variable.
%           Platform name. If want all the platform name don?t
write
%                 any name
%       Ouput
%           outp = Structure with the information,
%           platform,time,longitude,latitude, depth and variable
values of
%           the query point
%
%
%
%   Francisco Lopez & Mike McCann & Brian Schlining
%
%   Last modified
%   19/August/2012
```

Running it , the user will download the data from STOQS to Matlab as a structure variable, containing for each point the latitude, longitude, depth, variable name, platform name who measured it, date and variable value.

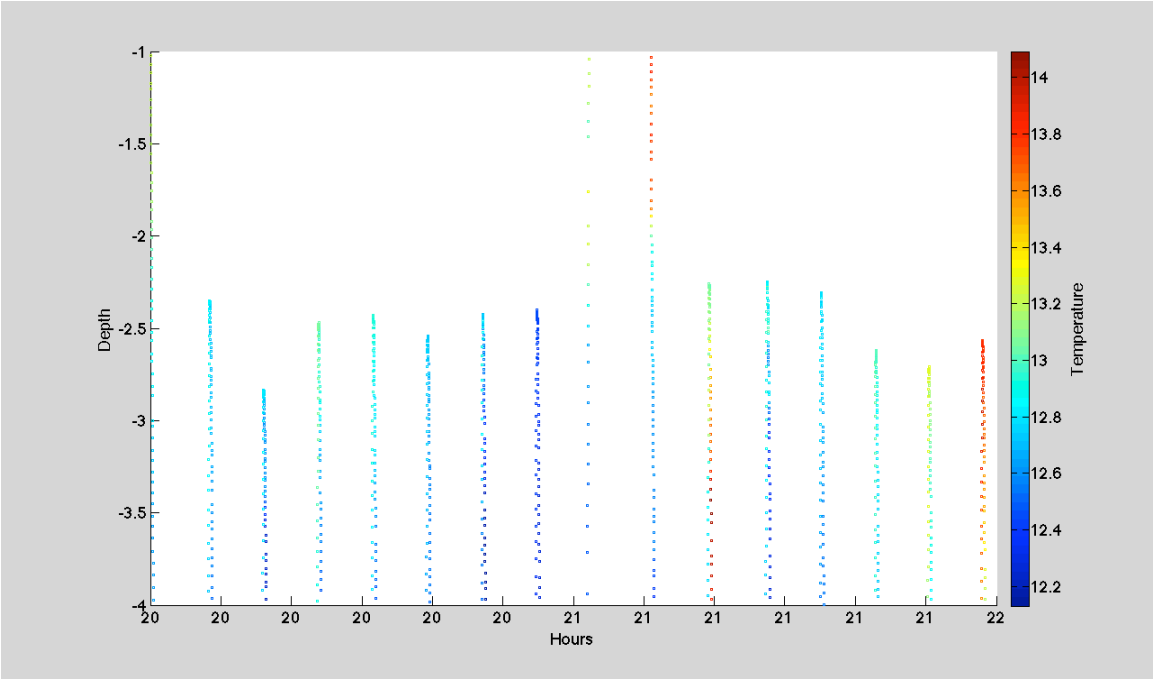


Figure 6 Plot of the output of one query did with stoqs_down.

stoqs_campaignbydate

```
function [inf]=stoqs_campaignbydate(urlst,date)
%Search for the campaign available in STOQS for the date given.
%
%
%Usage:
%       [camp]=stoqs_campaignbydate('http://odss-
staging.shore.mbari.org/canon',datenum(2011,06,21));
%
%Input :
%        urls=Url of the STOQS server
%        date= Date in Matlab format. Use datenum to convert date
to Matlab
%        format
%
%Output
%        inf= All the structure information for the campaign
selected.
%
%   Mike McCann & Brian Schlining & Francisco Lopez
%
%   Last modified
%   19/August/2012
```

stoqs_info

```
function inf = stoqs_info(u,table,var)
%
%Get info from a table in STOQS data base
%       inf=stoqs_info('http://odss-
staging.shore.mbari.org/canon/stoqs_may2012','platform');
%       infc=info_stoqs(u,'activity.json?campaign=','1');
%Usage:
%
%   t=info_stoqs(u)
%Input :
%   Could use 2 o 3 parameter in the input
%   u = Url direction of the STOQS data server. Ex: http://odss-
staging.shore.mbari.org/canon/stoqs_may2012
%   table = table to get the information
%   var= value of the parameter of the query to do to the table
%Output
% Get the struct variable inf with all the information.
%
%   Francisco Lopez & Mike McCann & Brian Schlining
%
%   Last modified
%   19/August/2012
```

stoqs_qcampaigns

```
function infcs=stoqs_qcampaigns(u,show)

%Get the name of all the campaigns available in a STOQS server
%
%Usage:
%
%  inf=stoqs_qcampaigns('http://odss-
staging.shore.mbari.org/canon/',1);
%Input :
%
%   u = Url direction of the STOQS data server. Ex: http://odss-
staging.shore.mbari.org/canon/
%   show = Show the info on the screen or not. If show=1 , show
the info,
%               if show=0 doesn't show it.
%
%Output
% Get the campaigns available on the STOQS data server
%
%
%   Francisco Lopez & Mike McCann & Brian Schlining
```

```
%
%   Last modified
%   19/August/2012
```

stoqs_qcampaign

```
function infc = stoqs_qcampaign(u,show)

%Get info from a campaign in STOQS data base
%
%Usage:
%
%   inf=stoqs_qcampaign('http://odss-
staging.shore.mbari.org/canon/stoqs_may2012',1);
%Input :
%
%   u = Url direction of the STOQS data server. Ex:
http://odss.mbari.org/canon/default
%   show = Show the info on the screen or not. If show=1 , show
the info,
%               if show=0 doesn't show it.
%
%Output
% Get the campaign available on the STOQS data server
%
%
%   Francisco Lopez & Mike McCann & Brian Schlining
%
%   Last modified
%   19/August/2012
```

stoqs_qplatform

```
function [infp,platname] = stoqs_qplatform(u,show)

%Get info from platforms in STOQS campaign data base
%
inf=stoqs_qplatform('http://odss.mbari.org/canon/default',1);
%Usage:
%
%   inf=stoqs_qplatform('http://odss-
staging.shore.mbari.org/canon/stoqs_may2012',1);
%Input :

%   u = Url direction of the STOQS data server. Ex:
http://odss.mbari.org/canon/default
%   show = Show the info on the screen or not. If show=1 , show
the info,
```

```
%                if show=0 doesn't show it.

%Output
%   infc = Get the platform info available on the STOQS data
server
%   platname = Name of the platform to acces it with the platform
ID.
%                Knowing the ID you can get the platname easily.
%
%   Francisco Lopez & Mike McCann & Brian Schlining
%
%   Last modified
%   19/August/2012
```

stoqs_qparameter

```
function [infpa,parname] = stoqs_qparameter(u,show)

%Get info from parameters in STOQS campaign data base
%
%Usage:
%
%   inf=stoqs_qparameter('http://odss-
staging.shore.mbari.org/canon/stoqs_may2012',1);
%Input :

%   u = Url direction of the STOQS data server. Ex:
http://odss.mbari.org/canon/default
%   show = Show the info on the screen or not. If show=1 , show
the info,
%                if show=0 doesn't show it.

%Output
%   infc = Get the parameter info available on the STOQS data
server
%   platname = Name of the parameter to acces it with the
parameter ID.
%                Knowing the ID you can get the platname easily.
%
%   Francisco Lopez & Mike McCann & Brian Schlining
%
%   Last modified
%   19/August/2012
```

stoqs_showcampaign

```
function stoqs_showcampaign(u,camp)
%Show on the screen all the main information about a campaign in
a STOQS
```

```
%database
%
%Usage:
%
%    stoqs_showcampaign('http://odss-
staging.shore.mbari.org/canon/','stoqs_may2012');
%Input :
%
%   u = Url direction of the STOQS data server. Ex:
http://odss.mbari.org/canon
%   camp = name of the campaign you want to do the query.
ex='default'
%
%
%   Francisco Lopez & Mike McCann & Brian Schlining
%
%   Last modified
%   19/August/2012
```

Model validation example

In order to show the capability of stoqstoolbox we will show an example of model validation using data collected during the May-June 2012 field experiment conducted by the Monterey Bay Aquarium Research Institute (MBARI) in Monterey Bay, California. The data are available from the STOQS server at http://odss.mbari.org/canon/stoqs_may2012/query/. Over 14 million data points of 18 parameters from 6 platforms measured over a 3-week period are available on this server. The model used for comparison is the Regional Ocean Modeling System developed by Jet Propulsion Laboratory for the Monterey Bay. The model output are loaded into Matlab using nctoolbox from the JPL server at http://ourocean.jpl.nasa.gov:8080/thredds/dodsC/MBNowcast.

Model validation with in situ measurements can be difficult because of different file formats and because data may be spread across individual data systems for each platform. With stoqstoolbox the researcher must know only the URL of the STOQS server and the OPeNDAP URL of the model output. With selected depth and time constraints a user's Matlab program searches for all in situ measurements available for the same time, depth and variable of the model.

STOQS and stoqstoolbox are open source software projects supported by MBARI and the David and Lucile Packard foundation. For more information please see http://code.google.com/p/stoqs.

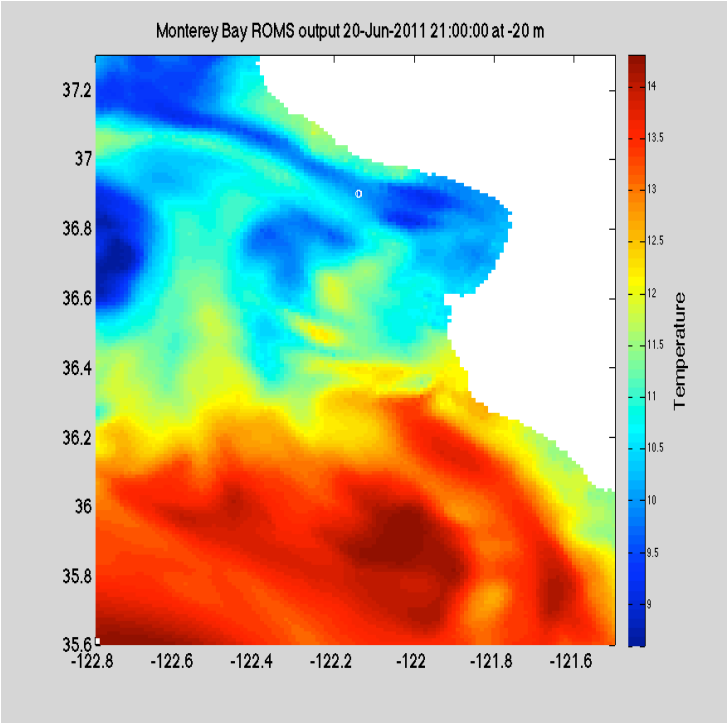


Figure 7 JPL model output for Monterey Bay.

Model_vs_stoqs demo

The goal of this demo is to show how to get the *in situ* measurement available in STOQS server for the time, depth and variable of the model output in an easy way. For this we will use nctoolbox to connect to the OPeNDAP server and get the model output data in Matlab

load_mb_1km

```
%Get the information of the OPeNDAP data set selected(url).
Information:Date,
%depth, latitude,longitude, and parameter value
%
%Usage:
%
%
[date]=load_mb_1km('http://ourocean.jpl.nasa.gov:8080/thredds/dod
```

```
sC/MBNowcast/mb_das_2012052515.nc',5)
%
%
%Input :
%
%   url = Url direction of the OpenDap data server. Ex:
http://ourocean.jpl.nasa.gov:8080/thredds/dodsC/MBNowcast/mb_das_
2012052515.nc
%
%
%
%Output
%   query = Structure with all the information.
```

extract_points

```
%  Usage
%
%   Get the value for the nearest node of the model to the insitu
%   measurement. It's use in the function model_vs_stoqs. You
will get one model value for each in situ measurement.
%
% Input
%
%   Model= Model information getting from model_vs_stoqs
%   insit= Insitu measuremente getting from model_vs_stoqs
%
% Ouput
%   extra = Structure with the x,y index(indx,indy), the data
model value(pointdata).
%    Get one model output for every insitudata
%       .pointdata = Value of the nearest node to the insitu data
%       .indx,indy = Index of the node nearest to the insitu data
%       .modeltime = Model time.
%       .time = Time of each of the in situ measurement
%      .longitude,latitude = Longitude and Latitude of the node
that have in situ measurement.
%
%
%   Francisco Lopez-Castejon
%   19/August/2012
```

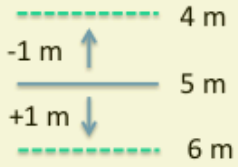
model_vs_stoqs

We must to know

Url of the OPeNDAP model output:

http://ourocean.jpl.nasa.gov:8080/thredds/dodsC/MBNowcast/mb_das_2011062021.nc

Depth Range: *+- X m of the selected depth. Ex: +-1m of 5 m*



Date Range: *+- X hours of model date. Ex: +-2 hours of 2011/06/20 21:00*

2011/06/20 19:00 <-> 2011/06/20 23:00

So we can query :

[query,d]=model_vs_stoqs('http://ourocean.jpl.nasa.gov:8080/thredds/dodsC/MBNowcast/mb_das_2011062021.nc',5,1,2,1);

And we will get all the data from STOQS and OPeNDAP in Matlab.

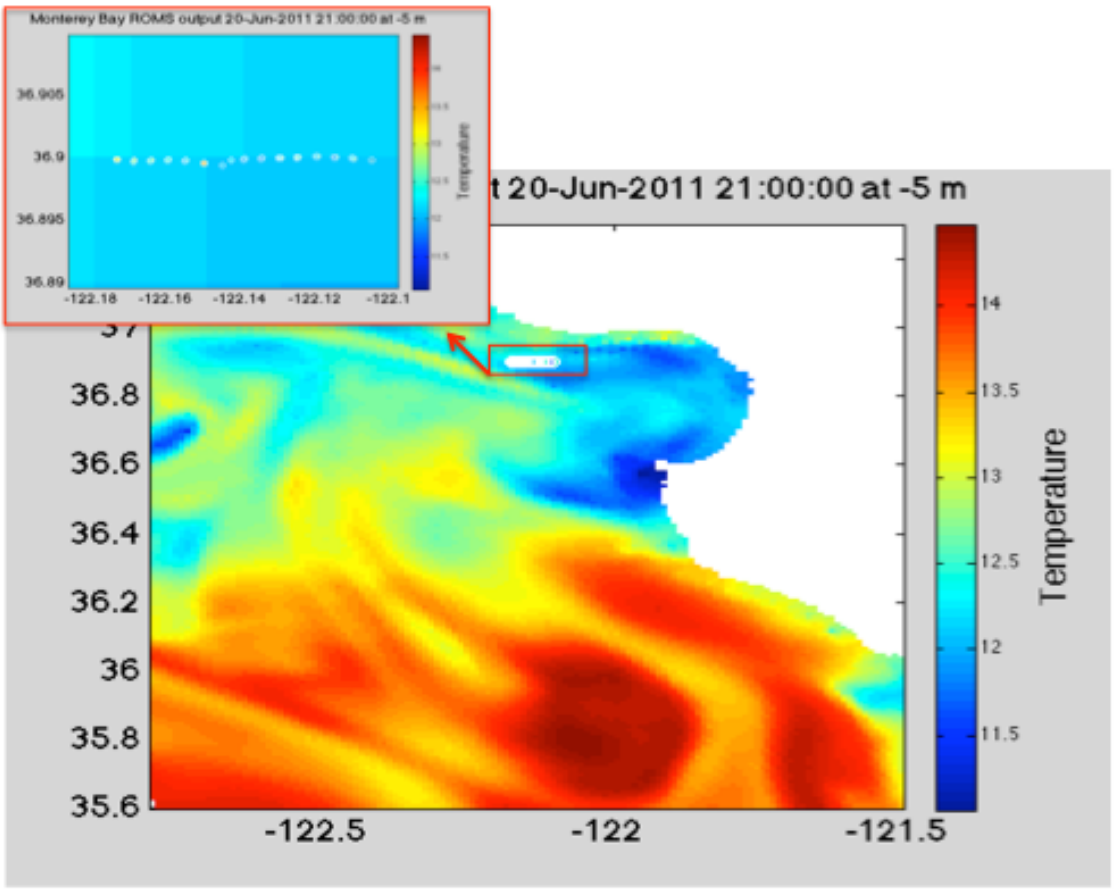


Figure 8 Plot of the output from model_vs_stoqs. The white spot are the in situ measurements download for our query.

```
function
[model,insit]=model_vs_stoqs(urlo,depth,range,drange,show)
%Compare in-situ data with model output
%
%
%    Input
%        urlo= OPeNDAP ROMS output to use
%        depth= Depth at we want to compare the data (m)
%        range = Range of +/- depth where to look for in-situ
measurement.
%        drange = Date regane where to look for in-situ data
(hours)
%        show = If is 1 make some plots.
%
%    Usage
%
[query,d]=model_vs_stoqs('http://ourocean.jpl.nasa.gov:8080/thred
ds/dodsC/MBNowcast/mb_das_2011062021.nc',5,0.1,2,1);
%
%    Output
%
```

```
%
%    Francisco Lopez-Castejon
%    19/August/2012
```

canon_may2012_validation

We will show how to query for different model output at the same time.

```
%Program to compare all the model output between two dates
%
% The Monterey Bay ROMS model, give an ouput every 6 hours
(3,9,15,21)
% We will try to compare the model ouput for all the time of the
CANON 2012
% May campaign (15 May 2012 01:23:30  to 13 Jun 2012 18:24:52 )
%
%    Francisco Lopez-Castejon
%    19/August/2012


sdate=datenum('01 Jun 2012 03:00:00');     %Start date to compare
edate=datenum('01 Jun 2012 21:00:00');     %End date to compare
intm=6; %Output model interval
depth=5;
depth_range=0.1;
time_range=3;
```

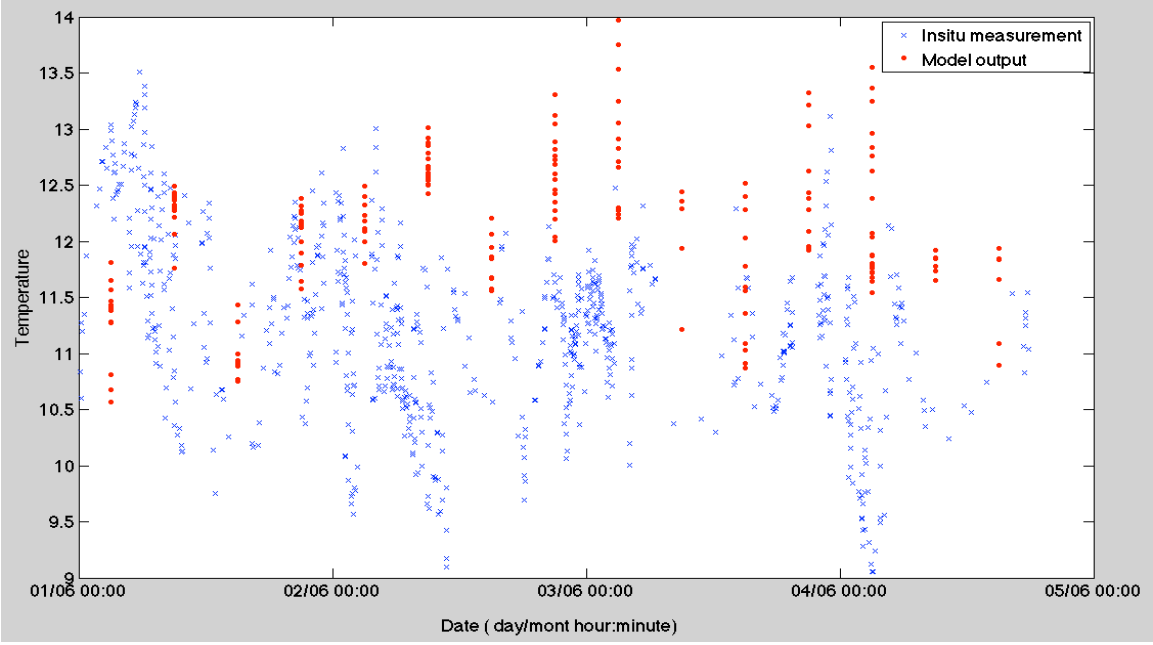


Figure 9 Comparasion of the model ouput (red spot) with the in situ measurement for our query (blue spots).

Acknowledge:

- **George Matsumoto** and **Linda Kuhnz** for help us a lot, take care of us and offer us so many fun things to do.
- All **the intern**, you all are very very nice and have been my little family during this time.
- All the **MBARI staff** to be so friendly and help me in every problem that I have had, specially **Erich Rienecker**.
- My mentor **Mike McCann** to make all the thing very easy to me, be so nice, and help me so much to have a very good experience.
- To Monique Messié, Fred Bahr, Brian Schlining ,John Ryan for the feedback received.
- **My family** for support me in this adventure.