US IOOS/SURA Testbed Data Compilation

1. Overview

We compiled biogeochemical observations from five sources:

- LUMCON (multi-year hypoxia monitoring program led by Nancy Rabalais of LUMCON)
- MCH (a series of NOAA NGOMEX programs led by Steven DiMarco of Texas A&M)
- EPA (observational program by the EPA Lab in Gulfbreeze)
- SEAMAP (NOAA fisheries surveys)
- MAGMIX (data from Alan Schiller of USM)

1.1 LUMCON

We received data from 2 sources:
(1) From Cathy Feng (cathyyangfeng@neo.tamu.edu) who obtained it from NODC's World Ocean Database
Date: Jan 2010
Original filename: LUMCON_ALLDATA_04to07.mat (located in Catz: /share/apps/user_apps/jiatang/project/SURA/DATA/LUMCON_data/)
Data period: 2004-2007
Project: LUMCON's hypoxia monitoring cruises
Comments: There are some stations with wrong year of 200 (15 stations) and some stations with temperature values of zero (121 stations, Mar-Jun and Aug-Oct of 2005)
Matlab file created by Jiatang Hu: TAMU_LUMCON_04to07_V4.mat (this matlab file was created through reorganizing the data from the original file, i.e. LUMCON_ALLDATA_04to07.mat into a consistent data format, using a matlab structure array "sta"; see "Data structure" below for detailed descriptions on this structure array; see the script 'TAMU_LUMCON_04to07.m' used to create this matlab file in: /share/apps/user_apps/jiatang/project/SURA/DATA/LUMCON_data/, and see the matlab file in the 'matfile' subdirectory)

(2) From John Lehrter (lehrter.john@epa.gov) from his LA shelf hydrographic and nutrient database
Date: Oct 2010
Original filename: Hypoxia_Public_v3.mdb (Access database format, located in: /share/apps/user_apps/jiatang/project/SURA/DATA/)
Data period: 2004-2006
Project: LUMCON's hypoxia monitoring cruises
Comments: There are many duplicate profiles (measured at the same time and

location yet measured as distinct profiles with different depths). Some duplicates go deeper than the bottom depth, or have different bottom depths.

Matlab file created by Jiatang Hu: DB_LUMCON_04to06_V4.mat (see the script 'DB_LUMCON_04to06.m' used to create this matlab file in: /share/apps/user_apps/jiatang/project/SURA/DATA/LUMCON_data/, and see the matlab file in the 'matfile' subdirectory).

1.2 MCH

We received data from 2 sources:

(1) From Cathy Feng (TAMU)

Date: Jan 2010

Original filename: MCH_ALLOXY.mat (located in: /share/apps/user_apps/jiatang/project/SURA/DATA/MCH_data/)

Data period: 2003-2005

Project: TAMU's Mechanisms Controlling Hypoxia program

Comments: File only contains oxygen data. There are over 490 stations with empty entries.

Matlab file created by Jiatang Hu: TAMU_MCH_03to05_V4.mat (see the script 'TAMU_MCH_03to05.m' used to create this matlab file in: /share/apps/user_apps/jiatang/project/SURA/DATA/MCH_data/, and see the matlab file in the 'matfile' subdirectory).

(2) From Steven DiMarco (dimarco@tamu.edu) from his MCH Data Archive

Date: Jan 2011

Original filename: MCH_data_archive (located in: /share/apps/user_apps/jiatang/project/SURA/DATA/MCH_data/);

Data period: 2004-2005 and 2007-2008

Project: TAMU's Mechanisms Controlling Hypoxia program

Comments: includes data from MCH1, MCH2 subdirectories (MCH3 only has nutrients; information regarding the time and positions of observations is missing); includes the bottle DO data (for MCH1); no data for 2006; There are some nutrient data with wrong year of 1905;

Matlab file created by Jiatang Hu: TAMU_MCH_ 04to08_V4.mat (this matlab file contains the combined data for MCH1 and MCH2 and can be found in: /share/apps/user_apps/jiatang/project/SURA/DATA/MCH_data/matfile/; it was created as follows:
First, the scripts 'MCH_CTD_V1_M0_M7.m' and 'MCH_NUT_M0_M7.m' in /share/apps/user_apps/jiatang/project/SURA/DATA/MCH_data/MCH_data_archive/MCH1/ were used to get the MCH1 data, then
the scripts 'MCH_CTD_M8_M16.m' and 'MCH_NUT_M8_M16.m' in /share/apps/user_apps/jiatang/project/SURA/DATA/MCH_data/MCH_data_archive/MCH2/ were used to get the MCH2 data,

and then the script 'MCH1_MCH2_combine.m' in /share/apps/user_apps/jiatang/project/SURA/DATA/MCH_data/ was used to combine the data for MCH1 and MCH2).

1.3 EPA

We received data from John Lehrter (EPA)
Date: Oct 2010
Original filenames:
1) Hypoxia_Public_v3.mdb
2) Prim_Prod_2005-2007_for_Katja_Fennel_vertically_binned.xls (includes primary production and chl)
Both located in:
/share/apps/user_apps/jiatang/project/SURA/DATA/;
Data period: 2004-2008
Project: EPA program
Comments: includes EPA inshore data for 2007-2008; includes chlorophyll and primary production data for 2005-2007; some profiles go deeper than the bottom depth;
Matlab file created by Jiatang Hu: DB_EPA_with_PP_04to08_V4.mat (used the scripts 'DB_EPA_04to07.m', 'DB_EPA_inshore_07to08.m', 'Lehrter_PP_data_05to07.m' and 'EPA_data_combine.m' in /share/apps/user_apps/jiatang/project/SURA/DATA/EPA_data/ to create this matlab file; see the matlab file in the 'matfile' subdirectory).

1.4 SEAMAP

We received data from the following sources:
(1) John Lehrter (EPA)
Date: Oct 2010
Original filename: Hypoxia_Public_v3.mdb (located in: /share/apps/user_apps/jiatang/project/SURA/DATA/)
Data period: 2004-2006
Project: SEAMAP program
Comments: no data for chl and nutrients
Matlab file created by Jiatang Hu: DB_SEAMAP_04to06_V4.mat (see the script 'DB_SEAMAP_04to06.m' used to create this matlab file in: /share/apps/user_apps/jiatang/project/SURA/DATA/SEAMAP_data/, and see the matlab file in the 'matfile' subdirectory).

(2) From David Sallis (David.Sallis@noaa.gov)
Date: Dec 2010
Original filename:
seamap_nerrs.mat; seamap_nmfs.mat; seamap_wod_CTD.mat;

seamap_wod_DRB.mat; seamap_wod_GLD.mat; seamap_wod_PFL.mat; seamap_wod_XBT.mat; (these matlab files contain the GeoJSON data from the NMFS, WOD, and NERRS database, and they were created using the scripts 'seamap_nerrs.m', 'seamap_nmfs.m', 'seamap_wod_new.m', and 'parse_json.m' in /share/apps/user_apps/jiatang/project/SURA/DATA/SEAMAP_data/; see these matlab files in: /share/apps/user_apps/jiatang/project/SURA/DATA/SEAMAP_data/);

Instructions from David Sallis: "get the GeoJSON data from the NMFS, WOD, and NERRS database (for example, use this URL to request the data from NMFS for 2006: http://service.ncddc.noaa.gov/moss?source=nmfs&units=m&cf=0&output=json&indent=1&timeframe=2006:2007&p=depth,temperature,oxygen,salinity,chlorophyll)"

Data period: 2004-2008

Project: SEAMAP program

Comments: used matlab's JSON parser to get the GeoJSON data into matlab format

Matlab file created by Jiatang Hu: NCDDC_SEAMAP_04to08_V4.mat (see the script 'NCDDC_seamap_04to08.m' used to create this matlab file in: /share/apps/user_apps/jiatang/project/SURA/DATA/SEAMAP_data/, and see the matlab file in the 'matfile' subdirectory);

(3) From Steven DiMarco (TAMU)

Date: Jan 2011;

Original filename: 2007-Cruise276; 2008-Cruise282 (located in: /share/apps/user_apps/jiatang/project/SURA/DATA/MCH_data/MCH_data_archive/SEAMAP/SEAMAP_summer_environmental_profiles1992-2008/)

Data period: 2007-2008;

Project: SEAMAP program;

Comments:

Matlab file created by Jiatang Hu: TAMU_SEAMAP_07to08_V4.mat (see the script 'SEAMAP_07to08.m' used to create this matlab file in: /share/apps/user_apps/jiatang/project/SURA/DATA/MCH_data/MCH_data_archive/SEAMAP/SEAMAP_summer_environmental_profiles1992-2008/, and see the matlab file in the 'matfile' subdirectory).

1.5 MAGMIX

We received data from Jerry Wiggert (Jerry.Wiggert@usm.edu)

Date: Nov 2010;

Original filename:

1) mm1-rep.txt; mm2-rep.txt; mm3-rep.txt; (CTD data)

2) magmixWinkler.tar (bottle DO data);

3) mm-dethps.txt (bottom depths);
All located in:
/share/apps/user_apps/jiatang/project/SURA/DATA/MAGMIX_data/;
Data period: 2008-2009
Project: MAGMIX cruises
Comments: includes CTD and bottle DO data (winkler titrations)
Matlab file created by Jiatang Hu: MAGMIX_08to09_V4.mat (used the scripts 'magmix_CTD.m', 'magmix_Bottle_DO.m' and 'magmix_CTD_Bottle_combine.m' in /share/apps/user_apps/jiatang/project/SURA/DATA/MAGMIX_data/ to create this matlab file; see the matlab file in the 'matfile' subdirectory).

2. Data processing

The SURA data was processed using Matlab. First, data from different sources was combined (for example, LUMCON data in TAMU_LUMCON_04to07_V4.mat and DB_LUMCON_04to06_V4.mat), then duplicates with identical profiles were removed, and then duplicates with distinct profiles were flagged (see "Data issues and treatment" below). The final version of the data compilation is listed as

For LUMCON: LUMCON_combined_04to07_V4_cleaned.mat (used the scripts 'clean_and_combine_LUMCON_SURA_data_V4.m', 'remove_combine_duplicate.m' and 'set_SURA_data_flag.m' in /share/apps/user_apps/jiatang/project/SURA/DATA/LUMCON_data/ to combine the LUMCON data and to remove/flag the duplicates; see the matlab file in the 'matfile' subdirectory);

For MCH: MCH_combined_04to08_V4_cleaned.mat (used the scripts 'clean_and_combine_MCH_SURA_data_V4.m', 'remove_combine_duplicate.m' and 'set_SURA_data_flag.m' in /share/apps/user_apps/jiatang/project/SURA/DATA/MCH_data/ to combine the MCH data and to remove/flag the duplicates; see this matlab file in the 'matfile' subdirectory);

For EPA: EPA_combined_04to08_V4_cleaned.mat (used the scripts 'clean_and_combine_EPA_SURA_data_V4.m', 'remove_combine_duplicate.m' and 'set_SURA_data_flag.m' in /share/apps/user_apps/jiatang/project/SURA/DATA/EPA_data/ to combine the EPA data and to remove/flag the duplicates; see this matlab file in the 'matfile' subdirectory);

For SEAMAP: SEAMAP_combined_04to08_V4_cleaned.mat (used the scripts 'clean_and_combine_SEAMAP_SURA_data_V4.m', 'remove_combine_duplicate.m' and 'set_SURA_data_flag.m' in /share/apps/user_apps/jiatang/project/SURA/DATA/SEAMAP_data/

to combine the SEAMAP data and to remove/flag the duplicates; see this matlab file in the 'matfile' subdirectory);

For MAGMIX: MAGMIX_combined_08to09_V4_cleaned.mat (used the scripts 'clean_and_combine_MAGMIX_SURA_data_V4.m', 'remove_combine_duplicate.m' and 'set_SURA_data_flag.m' in /share/apps/user_apps/jiatang/project/SURA/DATA/MAGMIX_data/ to combine the MAGMIX data and to remove/flag the duplicates; see this matlab file in the 'matfile' subdirectory);

For All Five Datasets: SURA_DATA_combined_04to09_V4_cleaned.mat (used the script 'combined_SURA_alldata_V4.m' in /share/apps/user_apps/jiatang/project/SURA/DATA/combine_data/ to combine the data from the five datasets above; see this matlab file in the same directory).

3. Data issues and treatment

(1) Some stations contain temp values of zero

-> Treatment: This applies to the LUMCON data from Cathy Feng (TAMU). Removed these temp values (121 stations, Mar-Jun and Aug-Oct of 2005).

(2) Some stations have wrong years

-> Treatment: This applies to the LUMCON data from Cathy Feng (TAMU) with the wrong year of 200 (15 stations) which was corrected to 2005, and the MCH1 nutrient data from Steven DiMarco (TAMU) where the wrong year of 1905 was corrected to 2005.

(3) There are many duplicate CTD profiles (measured at the same time and location yet measured at different depths, probably upcast and downcast; see Table 1 for an overview of duplicate profiles; see Figure 1 for some examples)

Table 1 Number of duplicate profiles

Variables	LUMCON		MCH		EPA		SEAMAP	
	Duplicate	Total	Duplicate	Total	Duplicate	Total	Duplicate	Total
Oxy	492	1773	579	1566	0	918	342	2391
Salt	699	2005	1	1351	328	1326	450	2576
Temp	699	2005	1	1351	328	1280	456	2576
NH4	179	1421	0	868	0	709	0	0
PO4	179	1422	0	868	0	785	0	0
chl	0	0	0	950	265	844	0	183

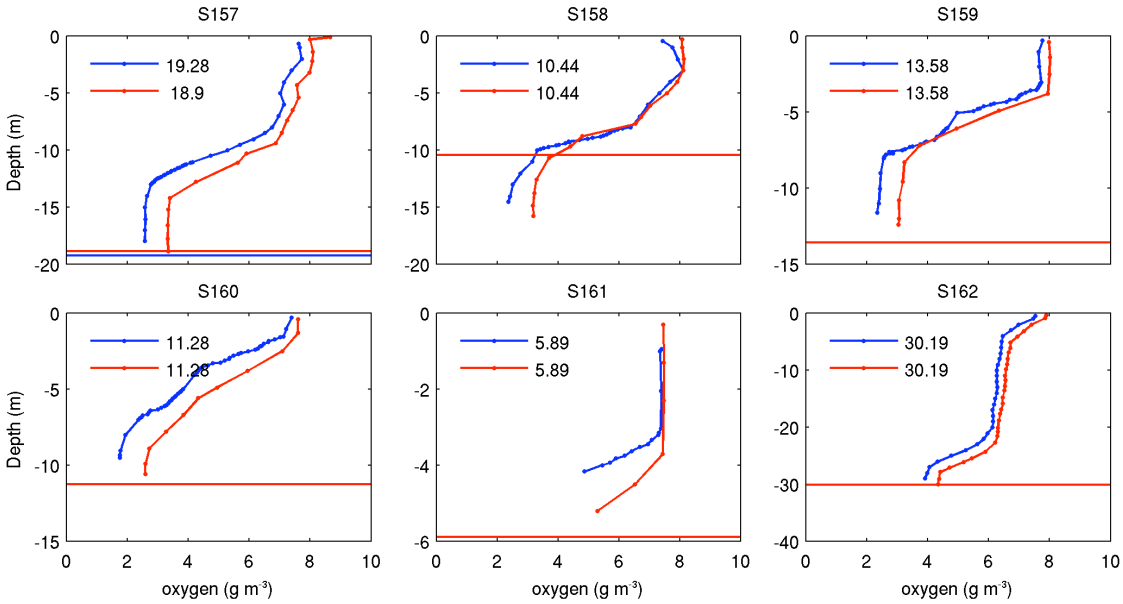

Figure 1 Duplicate oxygen profiles from LUMCON. The legend gives the bottom depths, which are also shown as the horizontal lines.

-> Treatment: set up a flag 'Dup_CTD_flag' for the duplicate CTD profiles (see "Data structure" below for details).

(4) Some profiles go deeper than the bottom depth (see S158 in Figure 1), and some duplicates have different bottom depths (see S157 in Figure 1);
-> Treatment: set up a data quality flag 'Q_*_flag' for the profiles that don't go all the way to the bottom or go deeper than the bottom depth (Q_*_flag = 3, if the bottom most data point is not within 2m of bottom depth; see "Data structure" below for details).

4. Data structure

The data is organized in the matlab structure array "sta", where every entry corresponds to one station. Each entry (each station) contains the 34 fields listed below:

sn = station name;
source = data source;
cruise = cruise ID;
year = year of observations;
month =month of observations;
day = day of observations;
datestr = a string showing the time of observations;
time = time in matlab serial data number format;
lat = latitude of station (^{o}N);

lon = longitude of station (°E);
Bdepth = bottom depth of station (m);
CTD_h = depth for the CTD profile ('CTD' as data type);
oxy = oxygen from the CTD profile;
salt = salinity from the CTD profile
temp = temperature from the CTD profile;
Bottle_h = depth for the bottle data ('Bottle' as data type);
Bottle_oxy = oxygen from the bottle;
Bottle_salt = salinity from the bottle;
Bottle_temp = temperature from the bottle;
Nutr_h = depth for nutrients ('Nutr' as data type);
NH4 = ammonium;
NO23 = nitrate plus nitrite;
PO4 = phosphate;
chl_h = depth for chlorophyll ('chl' as data type);
chl = chlorophyll;
PP_h = depth for primary production ('PP' as data type);
PP = primary production;
unit = units for the observed variables;
Q_CTD_flag = quality flags for oxygen, salinity, and temperature from the CTD profile (with values of 0, 1, 2, 3, 4, 5);
Q_Bottle_flag = quality flags for oxygen, salinity, and temperature from the Bottle data;
Q_Nutr_flag = quality flags for ammonium, nitrate plus nitrite, and phosphate;
Q_chl_flag = quality flag for chlorophyll;
Q_PP_flag = quality flag for primary production;
Dup_CTD_flag = flags for the duplicate CTD profiles (with values of 0, 1, 2).

For example, loading 'LUMCON_combined_04to07_V4_cleaned.mat' into Matlab, and typing "sta(20)" will give

```
>> sta(20)
ans =
    sn: 'C7'
    source: 'LUMCON'
    cruise: 'LC0401'
    year: 2004
    month: 1
    day: 13
    datestr: 'Jan-13-2004 09:40:00'
    time: 7.3196e+005
    lat: 28.8320
    lon: -90.3930
    Bdepth: 21.0400
```

CTD_h: [1x20 double]
oxy: [1x20 double]
salt: [1x20 double]
temp: [1x20 double]
Bottle_h: NaN
Bottle_oxy: NaN
Bottle_salt: NaN
Bottle_temp: NaN
Nutr_h: [0.8000 20]
NH4: [0.5000 0.4000]
NO23: [NaN NaN]
PO4: [0.7000 0.7000]
chl_h: [1x20 double]
chl: [1x20 double]
PP_h: NaN
PP: NaN
unit: {'oxy (mg/l)' 'salt (PSU)' 'temp (deg C)' 'NH4 (mmol/m^3)' 'NO23 (mmol/m^3)' 'PO4 (mmol/m^3)' 'chl (ug/l)' 'PP (mg C/m^3/d)'}
Q_CTD_flag: [0 0 0]
Q_Bottle_flag: [1 1 1]
Q_Nutr_flag: [4 1 4]
Q_chl_flag: 1
Q_PP_flag: 1
Dup_CTD_flag: 1

Explanation of the quality flags for the different data types

For each flag,

Q_*_flag = 0, if data is good; asterisk refers to the data types;
= 1, if no data available;
= 2, if bottom depth < 5m;
= 3, if bottom depth < or > 2m from the bottom most data point (see S158 in Figure 1);
= 4, if data gap > 5m (see Figure 2);
= 5, if there might be problems (see the red lines in Figure 3).

Explanation of the flags for the duplicate CTD profiles

Dup_CTD_flag = 0, if there are no duplicate stations;

Dup_CTD_flag > 0, if there are duplicate stations (measured at the same time and location yet measured as distinct profiles with different depths):
if the same observation types are included, Dup_CTD_flag = 1 for the profile with the bottom most data point, and Dup_CTD_flag = 2 for the other one; otherwise, set Dup_CTD_flag = 1 for the profile with more data types available, and set Dup_CTD_flag = 3 for the other one (with less data).

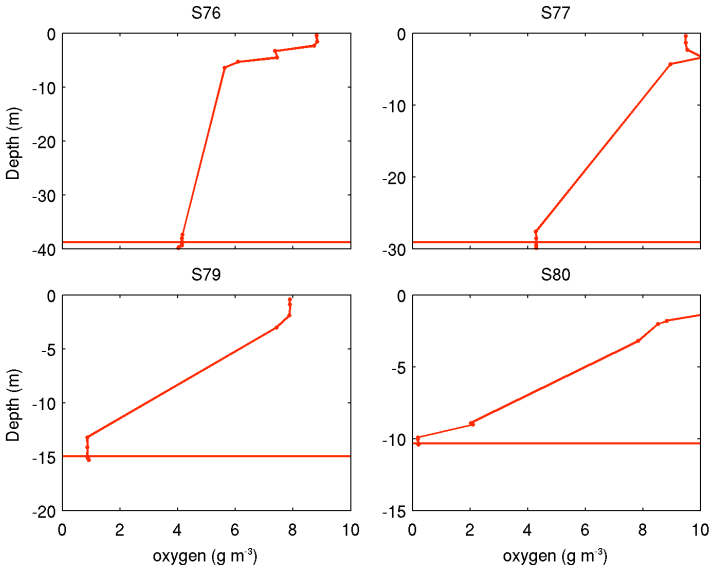

Figure 2 oxygen profiles with data gap > 5m (with a data quality flag of 3). Bottom depths are shown as the horizontal lines.

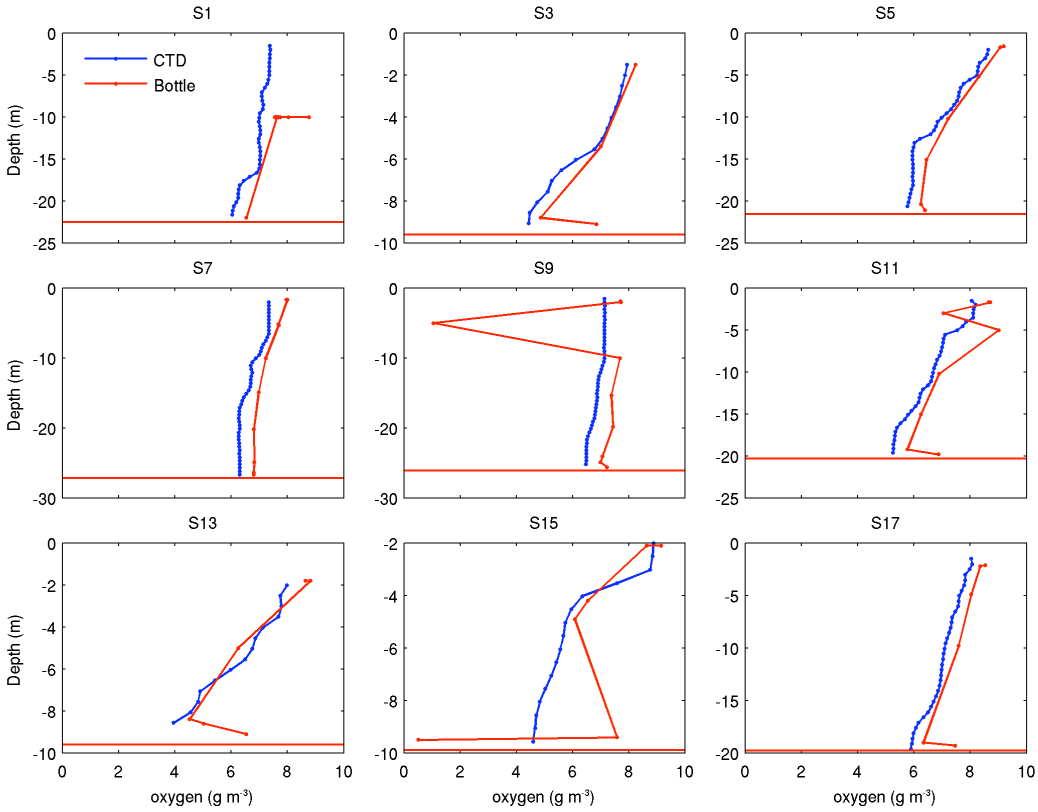

Figure 3 comparison of oxygen for the CTD (blue) and the Bottle (red) data from MCH. There might be problems with the bottle DO data (with a data quality flag of 5). Bottom depths are shown as the horizontal lines.

If you have any questions on above or need clarification please feel free to contact us via Jiatang.hu@dal.ca and katja.fennel@dal.ca.