Physical Oceanographic Mooring Time Series Protocol
Data Processing Procedures
Moorea Coral Reef LTER

By C. Gotschalk, 7 January 2010

Preliminary note: All these processing steps are being preformed off campus without the ability to map network drives. From on campus these could be set up very differently. I suggest working within a VNC window and pretending you cannot map drives.

MCR LTER Site Codes = CBYTS, FOR01, FOR04, FOR05.

This document explains the steps required to take the raw physical oceanographic data collected at the 4 Moorea LTER moorings through all processing required to generate 'Monster' files. Monster files are annual flat files with sensor data from multiple sites and various sample intervals all interpolated onto a common 20 minute time base. The list of sensors includes:

SBE37 Moored CTD
SBE39 Thermistors (some record pressure) (also bottom mounted BTMs)
SBE26+ Wave/Tide recorder.
ADCP (currents and waves)

A path to some required functions is automatically provided within the mfiles described below. Here it is just in case:
/data01/pisco/ucsb/shared-files/transfer/Chris/mfile_library/

SBE37 moored CTD processing
Data path = /mcrlter/internal/research/monitoring/physo/processing/ctd/*site*/)
mfile = /ctd_preprocessing/process_CTD_asc_files.m

The raw data are supplied as SBE .asc files with calibration information in the header. Time base is assumed to be GMT but check notes at top of header for discrepancies.

Copy these raw *CTD*.asc files into a temporary directory on the file server for preprocessing. They should also be copied to the data path above and subdirectory /raw_asc/ for permanent storage.

The mfile Process_CTD_asc_files.m sequentially loads these files and prompts the user to click on the start and end times of the deployment by inspecting the pressure time series. Toward the top of this mfile the user must provide the path to the temporary directory where the files are residing. The cropping mechanism within this mfile is ancient and crude requiring you to click zoom on the plot, then zoom into the deployment region, unzoom and hit return to activate the cross-hairs for choosing the start time. Reset to original view and repeat the process for the recovery time. The out-of-water regions are cropped and 4 (unfiltered/uninterpolated) variables are saved out to a text file with the same name as the input file but with a (.txt) extension.

col(1) – Matlab serial time, GMT
col(2) – Pressure, db
col(3) – Temperature, C
col(4) – Conductivity, S/m

P, T, and C are then 1 hour low pass filtered and Salinity and Density derived using sal.m and swstate.m respectively.

The variable 'outdata' is saved out to a .mat file of the same name (.mat extension) with column designations:

col(1) – Matlab serial time, GMT
col(2) – Pressure, db
col(3) – Temperature, C
col(4) – Conductivity, S/m
col(5) – Salinity, ppt
col(6) – Density, sigma-theta

Both the .txt and .mat files are archived here:
/mcrlter/internal/research/monitoring/physo/processing/ctd/*site*/final_mat/

SBE39 moored thermistor processing
Data path = /mcrlter/internal/research/monitoring/physo/processing/thermistor/*site*/)
… and /btm/*site*/
mfile = /thermistor_preprocessing/process_thermistor_asc_files.m

Raw *WTM*.asc, *WTP*.asc, and *BTM*.asc,files need to be copied to a temporary directory on the file server for processing. All files WTM, WTP, and BTM files can be processed together. There was some uncertainty about the BTM files and if they would be included in the monster files. This is why, as you'll see below, the processed BTM files are stored in their own directory /btm/.

The raw data are supplied as SBE .asc files with calibration information in the header. Time base is assumed to be GMT but check notes at top of header for discrepancies. Three file types WTP (temperature with pressure) and WTM (w/o pressure) and BTM (temperature sensor on the bottom) are available for input. All output files have a pressure column but WTM and BTM files will have missing value codes there.

As with the moored ctd data, the user is prompted to click on the start and end times. Once the deployment endpoints have been determined the time series are cropped and a text file of the same name with a (.txt) extension is saved out. No filtering nor interpolation in this text file either.

col(1) – Matlab serial time, GMT
col(2) – Temperature, C
col(3) – Pressure, db (all NaNs for WTM and BTM files)

After saving the text file, T and P are 1 hour low pass filtered. A matlab data file with the same name containing the variable 'therm' and with a (.mat) extension is saved out with the same column designations.

The WTM and WTP .txt and .mat files are archived here:
/mcrlter/internal/research/monitoring/physo/processing/thermistor/*site*/final_mat/
while the BTM files are archived here
/mcrlter/internal/research/monitoring/physo/processing/btm/*site*/final_mat/

SBE26+ wave/tide recorder processing
Data path = /mcrlter/internal/research/monitoring/physo/processing/wvt/*site*/)

(note: Run Windows Task Manager and view 'performance' during these steps. They are highly processor limited and you may not be aware that anything is happening if you don't have this to look at)

The first 3 steps are performed using SBE Seasoft for Waves (v 1.13) on your local machine. The large raw .hex files need to be downloaded to a local temporary directory. You will also need the SBE software configuration files with the .psa extension located at the /wvt/ level of the data path above.

The required seasoft modules used (and associated .psa files) are:
- Convert hex
- Merge barometric pressure (mcr_Merge26W.psa)
- Process wave burst data (mcr_ProcessWave26.psa)
- Create reports (mcr_CreateReports.psa)

The raw (.hex) file are converted to (.tid) tide files and (.wb) wave burst files using the Convert Hex module. Calibration coefficients are stored within the instrument and need not be supplied at this step. See http://www.seabird.com/pdf_documents/manuals/26plus_007.pdf for the full explanation.
NOTE: Atmospheric pressure needs to be subtracted from this time series and, at this writing, the FAA airport data is most reliable and contiguous. All data and mfiles used to create a properly formatted .bp file can be found here: (this will be moved and subversioned)
/home/pisco/gots/Chris_area/svn/matlab_tools/trunk/src/matlab/pisco_toolbox/contrib/UCSB/MCR_LTER/met_data/FAA_met/
Once corrected for atmospheric pressure the (.tid) tide file can be archived here:
/mcrlter/internal/research/monitoring/physo/processing/wvt/*site*/asc_tid_r26/

The .tid file columns are:
col(1) – scan number
col(2) – dd/mm/yyyy
col(3) – HH:MM:SS
col(4) – Pressure, db
col(5) – Temperature, C
col(6) – Conductivity, S/m

(again: Run Windows Task Manager and view 'performance' during these steps. They are highly processor limited and you may not be aware that anything is happening if you don't have this to look at)

The (.wb) wave burst file is subsequently processed using the SBE Process Wave Burst Data module. There are a lot of options available here. The easiest way to keep track of and repeat what you've done is to save and reuse the .psa setup file (currently mcr_ProcessWave26.psa but you can name it anything you like) along with the output data. You can save out a number of different output files. So far I'm just saving out the (.was) auto spectrum analysis file and the (.wts) zero crossing analysis file. A (.rpt) report file is also output and gives a summary of some of the options you've chosen.

Now that you have the (.was) and (.wts) files you can move on to SBE Create Reports and derive a variety of variables. The (.psa) setup file mcr_CreateReports.psa is a handy way of always outputting the same variables every time. The final output file has the appended extension (.r26) and contains the variables:

col(1) – date/time, dd mmm yyyy HH:MM:SS
col(2) – Significant Wave Ht., m, from time series stats
col(3) – Dominant Wave Period, sec, from time series stats
col(4) – SWH, m, from auto spectrum analysis (not included in monster file any more)
col(5) – DWP, sec, from auto spectrum analysis

Put this (.r26) file as well as the (.rpt), (.was), and (.wts) in the same directory as the previously create (.tid) file as these are all at about the same processing step.

Now move on to the Matlab processing steps.

The mfile /wvt_preprocessing/WVTascii_2_mat.m is used to combine the (.r26) and (.tid) files into Monster importable (.mat) files.

The 20 minute (.tid) tide data (mtime, P,T, and C) are 1 hour low pass filtered and salinity is derived from these filtered variables using sal.m. The 2 hour (.r26) wave burst data are NOT filtered but are interpolated onto the (.tid) files 20 minute time base. We want to be able to NaN out the interpolated values for SWH and DWP after monsterfication so we also include a logical column (flag) identifying the interpolated values to be NaNed out later.

A matlab .mat file with the same name containing the variable ‘wave’ and with a (.mat) extension is saved out with the column designations:

col(1) – Matlab serial time
col(2) – Conductivity, S/m
col(3) – Temperature, C
col(4) – Pressure, db (remember: needs to have atm pressure subtracted)
col(5) – Salinity, ppt
col(6) – Density, sigma-theta
col(7) – SWH, m (contains all interpolated values. Time series analysis)
col(8) – DWP, sec (also contains the interpolated values)
col(9) – SWH, m (unused in monster)
col(10) – DWP, sec
col(11) – flag, logical (1 used toNaN interpolated values, 0 = keep)

There is a call to a separate function, save_WVT_textfiles.m, which is used to generate unfiltered, uninterpolated (.txt) text files after the start and end times are determined for both the (.tid) 20 minute data and the (.r26) 2 hour wave burst data. Two output files are generated. Data from the 20 minute (.tid) tide file are output as (.tid.txt) files. Data from the 2 hour (.r26) wave burst file are output as (.r26.txt) files. All these files should be archived as follows:

/asc_tid_r26/ (.was .wts .tid .rpt .r26)
/final_mat/ (.mat .txt)
/raw_hex_wb/ (.hex .wb)

Wave measuring ADCP processing
Data path = /mcrlter/internal/research/monitoring/physo/processing/adcp/*site*/)

The first few steps here require moving the raw files (.000 extension) to and from your local machine. They are often several 100 MB and so you should plan ahead for delays based on your internet connection speed.

Step 1 is to process the big .000 files with RDI WavesMon to process the wave burst data and to isolate current ensemble data into a separate file. Waves data outputs will have a .wvs extension. Current data will have a .000 extension or a .PDO extension depending on the version of your software.

Within WavesMon you’ll set a variety of options described below:

[deployment info] set transducer height to 0.4m

[advanced processing]
fft length = 2048
freq bands = 128

lft = 0.025
uft = 0.35, 0.65, 0.20

[data screening] set percent good to 75%

[what2process] check all boxes except time series output (default)

[advanced file outputs] might be good to set max current size to, say, 200MB to avoid having to concatenate the files later.

Note that processing [control][start] can take several hours and will produce one or more .wvs and .000 or .PDO files.

The next step is to open each of the newly create files containing current ensemble data with RDI WinADCP and check them for data quality and note the ensemble numbers and times of the start and stop for each deployment. There is a period of time where diver bubbles will interfere with the current measurements so don't go by pressure to determine when the units are happily 'on station'. The beam intensity profile is a good diagnostic showing when the instrument is undisturbed.

Once you've noted good start and stop ensembles for each current ensemble file you can upload all these files back to a temporary directory on the file server for more processing. (note: no need to upload the raw .000 files as they are already on the server. Just cp them over to the archive directories noted below)

The last step to do on your local machine is to open each of the .wvs wave files in RDI WavesView and inspect the Hs, Tp, Dp, and water level time series data. We save out these time series as text files by:

- right clicking on the upper time series plot
- check boxes for Hs, Tp, Dp, and water level if not already
- click [save text file] box and save with proper file name format
- in a text editor (J-edit?) open each .20.1.txt file and remove all but one header line and any data lines recorded while out of water (have values of –32768). (note that clipping data at the beginning of the file may require you to change the file name if you clip into the next day)

Before you upload all these files to the file server add in the yyyymmdd_yyyymmdd deployment and recovery times from the raw file onto the .wvs and .wks file names. The .20.1.txt, .wks and .wvs files are good to go at this point and can be uploaded to: /mcrlter/internal/research/monitoring/physo/processing/adcp/*site*/wave_wvs_txt/

The last step in processing the current ensemble files requires using Matlab mfile located here: /monster_processing/adcp_preprocessing/. First you have to update good_ensemble_list.m with the information from these deployments. Next, run ADCP_batchonedir_premonsterfication.m to create the .20.1.mat and .30.1.mat files.

Copy these files to their archive directories but maintain them in the temporary directory for a while as you may have to refer back to them for error investigating purposes.

MONSTERFICATION

All oceanographic sensor data collected at one 'site' and during one specific 'year' are included in an annual Monster file. Monster files are flat files made up of a 20 minute interval rows starting Jan 1, year 00:00:00 through Dec 31, year 23:40:00 GMT, and 151 columns (at this writing). All preprocessed data from all the instruments are interpolated onto this 20 minute time stamp into designated columns using MCR_monster_batch.m and MCR_monster_generator_beta.m. MCR_monster_batch is a front-end that calls the _generator_beta.m function. The data paths probably won't need to be messed with since they refer to the widely used directory structure. The output directory (outdir) can be set to a temporary directory so you can inspect the monster files before replacing the archived versions. Let me describe two switches in MCR_monster_generator_beta.m that can be set (on/off 1/0) depending on what you want to do. The output from these scripts are a .mat and a .txt file. The .txt file takes a tedious amount of time to write out so for the initial runs and inspections I set the text_switch = 0 for faster creation of the .mat monster files. The tide_switch turns on or off a function that calculates the velocity component due to the tides for the adcp data. Sometimes this function fails (especially with short time series of velocity data) and so you can turn it off here and proceed with monsterfication. Ok, now back to MCR_monster_batch.m. Once you're happy with the monster files you're creating you will want to update the monster files (.mat and .txt) in the archive and concatenate all the years together. To accomplish this:

- make sure outdir is set to the archive root directory
- be sure you are working on the latest years data (Data doesn't extend to end of year)
- set concat_switch to 1 (this cuts the text file off at the last data record)
- set the revision string to the next value depending on what you've done.

The column designations in both the .mat and .txt monster files are:

col	data description	instrument
1	Matlab serial time	Time in Matlab format
2	year (yyyy)	strings
3	month (mm)	
4	decimal day (dd.ddddd)	

(Note: bin distance from bottom increases as column number increases. There have been recent changes in the adcp setups requiring some recoding. When the units were set to 0.5m bins we averaged groups of two bins creating 1.0m bins. Now we're collecting data at 1.0m bins so no grouping is necessary)

5:24	U velocity (20 bins)(m/s)	ADCP
25:44	V velocity	ADCP
45:64	Average beam intensity	ADCP
65:84	Percent good (PG3)(%)	ADCP
85	Temperature (c)	ADCP
86	Pressure (db)	ADCP
87	Sig. Wave ht. (Hs)	ADCP
88	Dom. Wave pd (Tp)	ADCP
89	Wave direction (degrees true)	ADCP
90:91	unused	

(distance from bottom increases as column number increases)

92:111	Temperature(s)	SBE39 (allow for 20 levels)
112:131	Pressure(s)	SBE39
132	20m BTM	SBE39
133	10m BTM	SBE39
134	unused	
135	Pressure	SBE37 (deeper unit)
136	Temperature	SBE37
137	Conductivity	SBE37
138	Salinity (ppt)	SBE37
139	Density (sigma-theta)	SBE37
140	Pressure	SBE37 (shallow unit)
141	Temperature	SBE37
142	Conductivity	SBE37
143	Salinity	SBE37
144	Density	SBE37
145	Pressure	SBE26+
146	Temperature	SBE26+
147	Conductivity	SBE26+
148	Salinity	SBE26+
149	Density	SBE26+
150	Sig. wave height (m)	SBE26+ (from time series stats)
151	Dominant wave period (sec)	SBE26+

Monster files are saved out as Matlab data files with a (.mat) extension. An example file name = FOR01_monster_2008.mat. Missing values in the .mat version are NaN and in the .txt version are 9999.

Updating the summary figures

There are several mfile within /summary_figures/ that quickly update some visually useful plots once new data is available. Here are some of them:

mcr_data_coverage.m (line/timeseries plot showing the presence of various sensor data from year to year and site to site)

annual_timeseries_summary.m (temperature contour and some wave timeseries)

temperature_line_timeseries_allsites.m (simply plots all temperature data for all sites)

Pax_currents_map_allSites.m (plots a Moorea map and principal axes currents for sites with adcp)