FOSTER-LIS Gridded Data Products:

Observed Current Profiles and Near-Surface Water Properties from

Ferry-based Oceanographic Sampling in Eastern Long Island Sound

May 2007

Daniel L. Codiga

Graduate School of Oceanography, University of Rhode Island

Narragansett, RI 02882

GSO Technical Report 2007-01

Funded by EPA Assistance Agreement No. LI-97106001(-2)
Long Island Sound Study

Full citation:

Codiga, 2007. FOSTER-LIS gridded data products: Observed current profiles and near-surface water properties from ferry-based oceanographic sampling in eastern Long Island Sound. Technical Report 2007-01. Graduate School of Oceanography, University of Rhode Island, Narragansett, RI. 14pp.

FOSTER-LIS Gridded Data Products:
Observed Current Profiles and Near-Surface Water Properties from Ferry-based Oceanographic Sampling in Eastern Long Island Sound

Daniel L. Codiga
Graduate School of Oceanography, University of Rhode Island
Narragansett, RI 02882

Contents

1. Summary

This report explains and describes the content and format of the gridded data products (GDPs) resulting from the FOSTER-LIS (Ferry-based Observations for Science Targeting Estuarine Research in Long Island Sound) project. The water velocity GDP consists of (1) a single datafile containing the grid parameters; (2) a datafile containing the water velocities for each ferry transit during which they met quality control standards; (3) a datafile containing auxiliary parameters for each such ferry transit; and, (4) for each of 18 fixed locations along the ferry route, a datafile containing a time-sequence of values drawn from all the individual transits. Similarly, the near-surface water property GDP consists of (1) a single grid parameters datafile (the same file as for the water velocity dataset), (2) a datafile containing the near-surface temperature, salinity, and chlorophyll values and auxiliary data for each ferry transit during which quality control standards were met, and (3) for each of 18 fixed locations along the ferry route, a datafile containing a time-sequence of values drawn from all the individual transits. All GDP data files are available for download at the project website, www.gso.uri.edu/foster, in both “ascii” format and “mat” format for use in Matlab. The website also facilitates preview of GDP data. Raw and intermediate data, and/or data otherwise not included in the GDP, are available on request from the author (d.codiga@gso.uri.edu; 401-874-6212).

2. Background

A ferry is equipped with an acoustic Doppler current profiler (ADCP) and a water property sampler in order to carry out ferry-based oceanographic sampling in eastern Long Island Sound (LIS). Vertical profiles of absolute water velocity, and near-surface water temperature, salinity, and chlorophyll fluorescence, are collected nominally 8 times daily along a transect between New London, CT and Orient Point, NY. This dataset is made possible by Cross Sound Ferry Services, who facilitate sampling from their vessel *MV John H* as an in-kind contribution. The project was established as part of a 2-year research project funded by the Connecticut Sea Grant College Program in 2002 ("A ferry-based observation system for Long Island Sound: Application to physical influences on hypoxia", PI Codiga, while at University of Connecticut). The base program of the Long Island Sound Study office of the Environmental Protection Agency provided support from 2004 to 2006.

One of the primary goals for the GDPs described here is to be of use to hydrodynamic modelers in calibrating, validating, and contraining their outputs. By improving hydrodynamic models, the water quality models they underlie are strengthened and thus water quality management decisions better supported. Water quality models play a central role in total maximum daily load (TMDL) management methods now being applied to address the issue of hypoxia in LIS from a nutrient-load perspective. While hypoxia tends to occur in the western and central parts of LIS, the bulk of the exchange of water between LIS and the coastal ocean occurs at the eastern end, which makes ferry-based sampling in eastern LIS important to improve understanding of LIS-wide dynamics. In the event that more information than is accessible with the GDPs at the project website (www.gso.uri.edu/foster) is desired, all raw and intermediate data are available on request from the author at d.codiga@gso.uri.edu.

3. Water velocity profiles GDP

a. Methods and quality control

The ADCP is a 600 kHz RD Instruments Broadband Workhorse/Mariner installed in a custom-designed sea chest in the hull of the vessel located about 2/3 the way toward the stern. The sea chest includes a plexiglass plate that encapsulates fresh water adjacent to the ADCP transducers, which very effectively limits damage to them due to biofouling. In the pilot house, a satellite compass provides navigational information (position and heading) to the control PC. The primary factor in quality assurance and quality control is that in very rough sea states, increasing amounts of bubbles are swept under the hull and can interfere with the acoustic signal.

For quality assurance, prior to being averaged and mapped to fixed grid locations, raw data from each individual transit are plotted, inspected visually as a group, and assigned a quality value from 5 to 1. Only transits assigned quality flags of 4 or 5 have been included in the GDP. The meanings of the quality flags are as follows:

5. Unblemished. Near-surface and/or near-bottom outliers are absent or nearly so, there are very few if any vertical profiles for which data are missing, and data values vary smoothly across the transit.

4. Near-perfect. Transect includes as many as several near-surface and/or near-bottom outliers, and/or as many as 10-15 missing vertical profiles, and data values vary smoothly across the transit.

3. Sparse. Transect includes numerous near-surface and/or near-bottom outliers, a substantial fraction of vertical profiles missing, and/or most data that are present vary smoothly across the transit but others vary irregularly, indicating they are of poor quality.

2. Compromised. The majority of the vertical profiles are missing and/or a substantial fraction of the data present vary irregularly indicating they are of poor quality.

1. Unacceptable. Data are predominately missing and/or the majority of the data that are present vary irregularly indicating they are of poor quality.

Very conservative (high) estimates of instrument uncertainty in individual GDP velocities are +/- 3 cm/s, based on manufacturer specifications and averaging described below.

b. Grid characteristics

The grid to which observations are mapped consists of 18 longitude-latitude locations, or "virtual stations" (Figure 1). They are referred to as virtual stations because the ferry does not necessarily pass directly over each of them on a given transit; rather, there is a range of locations in which a given ferry transit will fall, and the virtual stations are placed at positions along a linear path chosen to lie near the center of that range (Figure 1).

There are 17 stations spanning the eastern end of LIS, including The Race, with one-kilometer spacing. They are denoted R01-R17 (R indicating race) from north to south. One station lies in Plum Gut, the gap between Plum Island and Orient Point on Long Island, and is denoted G01 (G indicating gut).

The method by which data along the particular vessel track during a single transit are assigned to virtual stations will now be described briefly. It is motivated mainly by the need to properly estimate along-estuary volume transport using water velocities. Each virtual station is taken to lie at the center of a "virtual channel" defined by an along-estuary direction (Figure 1). For R01-R17, the along-estuary direction is 10 degrees counterclockwise from east, which is representative of the along-estuary direction of the entire LIS estuary. For G01, the along-estuary direction is 35 degrees clockwise from east, which is appropriate to the local bathymetry and coastline. ADCP data are collected each 2 seconds, and 16 such single-ping values are averaged to form an "ensemble mean" representing each 32 second interval; typically there are 4-5 such ensemble-mean values taken within each virtual channel. For each transit, all ensemble-mean values from within a given virtual channel are averaged and the result assigned to its corresponding virtual station. If there are fewer than two such ensemble mean values in a virtual channel, the data are rejected. In Figure 1, positions of ensemble-mean data from alternating virtual channels are shown in black/white. Figure 2 shows an example of an individual crossing; the plan-view upper portion of the figure shows small green/blue dots for each ensemble-

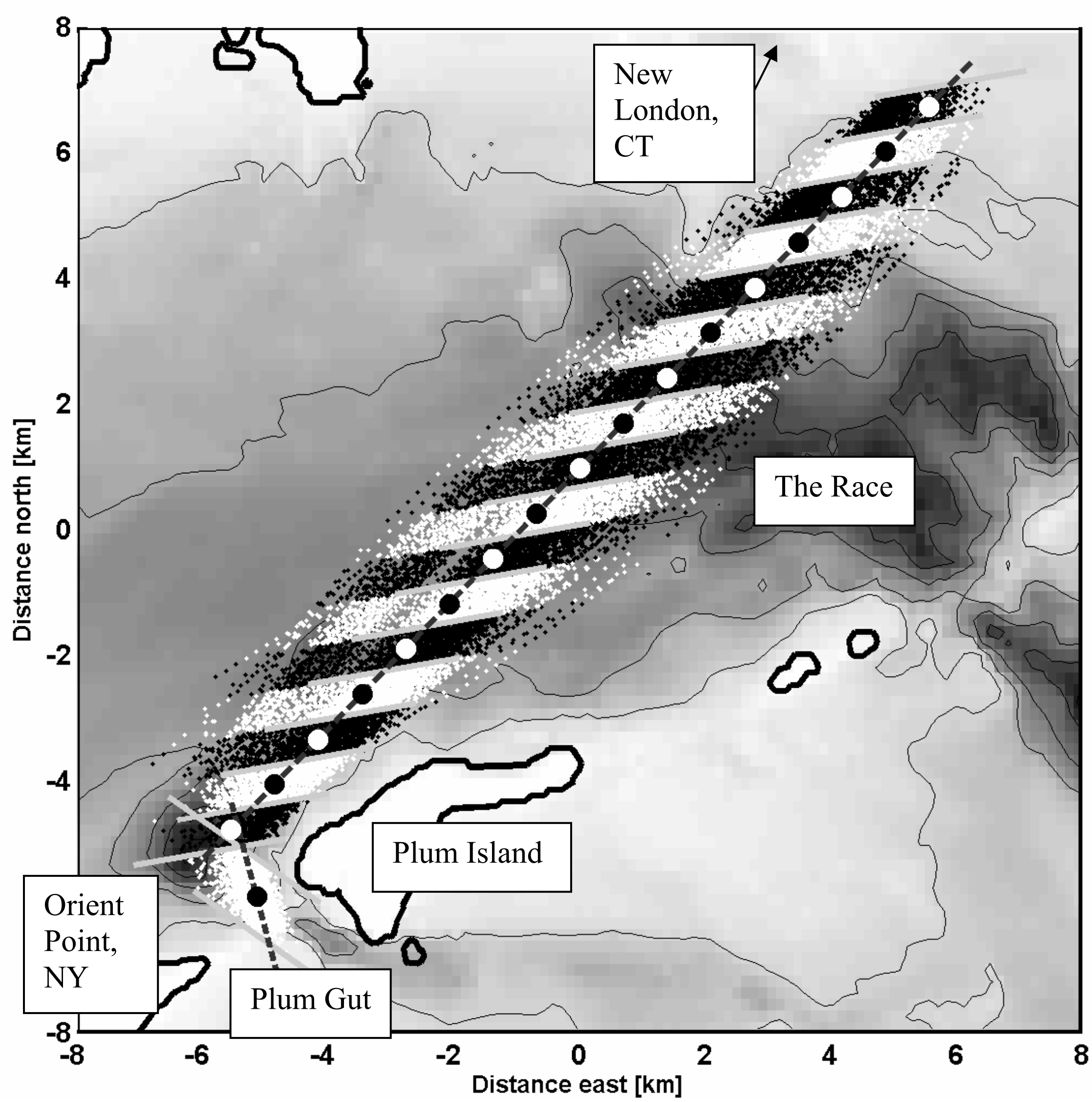

Figure 1. Plan view showing the 18 virtual stations (large dots, alternating black and white) along a linear track (dashed black line) chosen to fall near the center of the range of vessel positions from typical transits (each a small black or white dot; a few months of data are shown). Virtual stations R01-R17 are oriented north to south in The Race and station G01 lies in Plum Gut, farthest south. Each virtual station is centered within a virtual channel (boundaries shown in gray) aligned along-estuary. Bathymetry is shown

mean location where the actual ADCP samples were made, while the virtual stations to which they were assigned lie at the tails of the vertical-mean velocity vectors shown.

In the vertical, at each virtual station the grid has 47 levels, spanning from 8 to 100 meters below sea level at 2-meter intervals. The grid extends deeper than the seafloor at all virtual stations, to accommodate the range of different depths to which the sampling extends at various locations across the estuary.

Bathymetric depths are sampled by the ADCP using bottom tracking. Variations in the vessel path from transit to transit mean that data from a range of different bathymetric depths are mapped to a given virtual station. Ensemble-mean bathymetric depths are assigned to the virtual stations just as water velocities are. Therefore, for each virtual station both a mean seafloor depth, and the standard deviation about that mean, is computed using multiple transits and considered part of the grid information. As a result, water velocity samples from an individual transit can lie deeper than the multiple-transit mean bathymetric depth at a given virtual station, particularly for virtual stations where bathymetry varies strongly in the direction perpendicular to the average vessel path.

Times of ensemble-mean velocities in a virtual channel are averaged and assigned to velocities of the virtual station; results are not mapped to an evenly-spaced grid.

The data values characterizing the grid are contained in a single datafile. The 'readme' information distributed with the grid datafile is as follows:

```
1. zgrid (47 x 1)
  Depths, in meters, corresponding to the centers of depth
intervals to which velocity values apply. Mean sea level is
zero and increasingly deeper depths are more negative.
2. lat_vc & lon_vc (each 1 x 18)
  Latitudes (degrees North) and longitudes (degrees West;
negative) of the centers of the virtual channels (explained in
report) from which velocity measurements are drawn.
3. vc_name (1 x 18)
  Descriptive names (3-character strings) of virtual channels.
4. bathy_mean & bathy_stdev (each 1 x 18)
  Seafloor depth, meters, from bottom-track measurements. Mean
and standard deviation, based on ~1000 ferry transits, of the
average depth values which were each computed using all
16-second ensemble-mean measurements within a virtual channel.
```

b. Velocity observations – Individual transits

For each individual ferry transit of acceptable quality, there is a single datafile containing water velocities, named with the format "gridvel_YYYY_MM_DD_HH", where YYYY is the year, MM is the month, DD is the day, and HH is the hour (00-23) in Eastern Standard Time (EST). (Local time is the same as EST, except during daylight savings time. During daylight savings time, which occurred between the first Sunday in April and the last Sunday in October in years through 2006, local time is one hour ahead of EST.)

Velocities are provided as eastward and northward components, in cm/s, at each of the 18 x 47 grid locations as described above, where available. A time, corresponding to when the data were collected in EST, is given for each virtual station based on an average of the times of the ensemble-mean times that were assigned to the virtual station.

Actual depths of 2-meter ADCP bins are used to assign the ensemble-mean velocities to the grid depths by shifting them less than 1 meter. ADCP values shallower than 8 m have been omitted as they are within the blanking distance of the instrument. Ensemble-mean ADCP data deeper than 90% of the bottom-track bathymetric depth for its corresponding

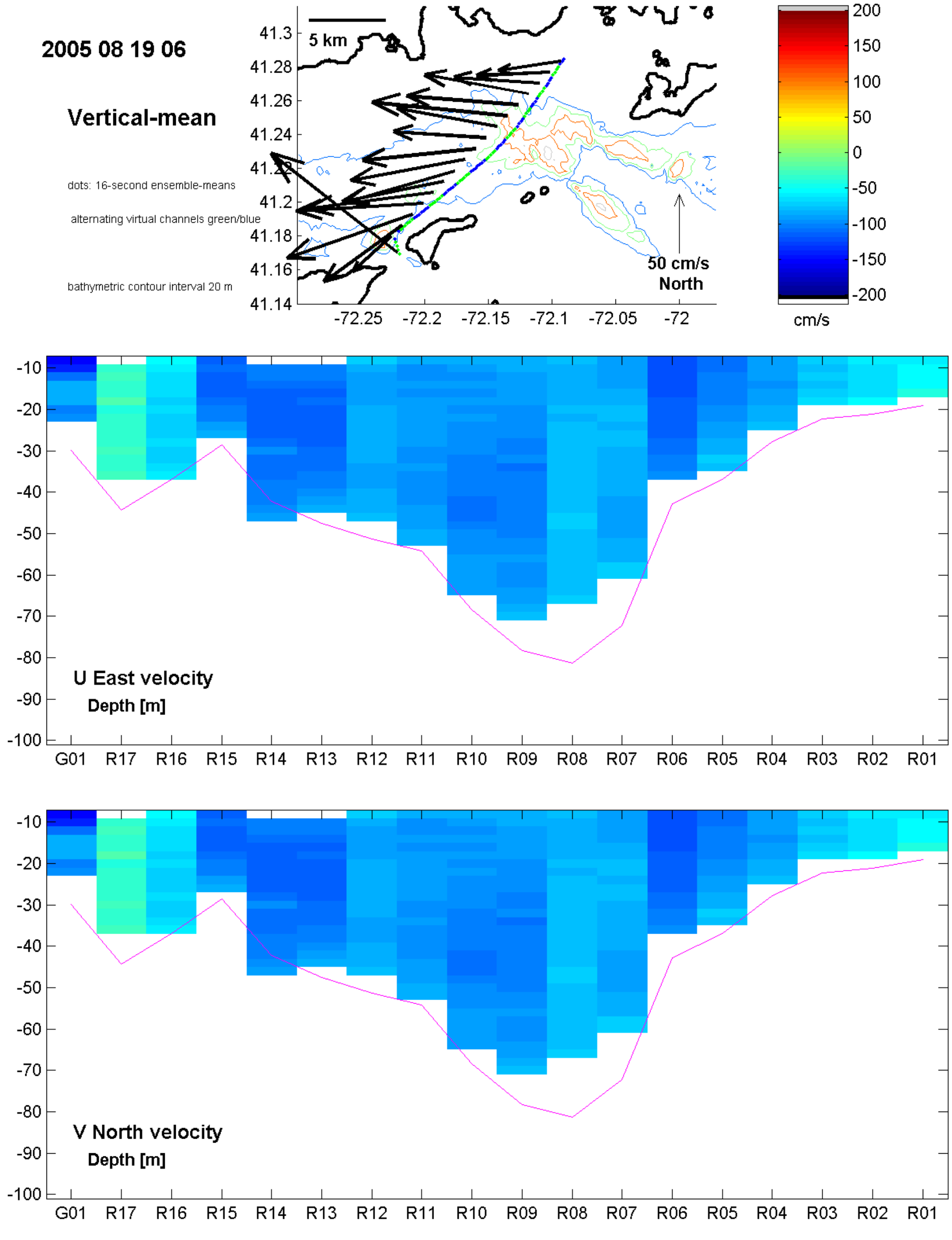

Figure 2. Example of data from an individual transit. See text for discussion.

ensemble-mean have been omitted; this is a conservative choice as the instrument specification is to keep 94%. In the event that fewer than 2 ensemble mean values are available to compute the average for a virtual station, the data are omitted.

The 'readme' information distributed with the individual-transit datafiles is as follows:

```
1. vel_east & vel_north (each 47 x 18)
  Velocity components in the directions toward the east and
north, in cm/s. Each is a matrix with values corresponding to
depths "zgrid" and virtual channel locations "lat_vc"/"lon_vc"
in the grid file.
2. t_yr, t_mon, t_dy, t_hr, t_min, & t_sec (each 1 x 18)
  Each of these variables consists of one value for each
virtual channel, the average over data collected there. They
are times in Eastern Standard Time (EST).
```

In addition, auxiliary data for each individual transit is provided in a second datafile, named with format "gridvel_YYYY_MM_DD_HH_aux". The auxiliary data consists of a matrix the same size as those for velocities, but holding the number of ensemble-mean values used to compute gridded velocities at each virtual station at each grid depth. It also contains the mean longitude/latitude for ensemble-mean samples from each virtual channel for the transit. Finally, for each virtual station it contains a mean bathymetric depth based on the ensemble-mean samples in that virtual channel, as well as the standard deviation about that mean, and number of values used in computing the mean.

The 'readme' information distributed with the individual-transit auxiliary datafiles is as follows:

```
1. numavg_vel (47 x 18)
  Number of 16-second ensemble-mean velocity measurements
used in computing the velocities. A matrix of values at
depths "zgrid" and virtual channel locations "lat_vc"/"lon_vc"
in the grid file, just as the velocities.
2. lat_ae & lon_ae (1 x 18)
  Average latitude and longitude of all 16-second ensemble-
mean velocities within each virtual channel, from this
individual crossing.
3. bathy_mean, bathy_stdev, and bathy_numavg (each 1 x 18)
  Average, standard deviation, and number of values averaged,
for all 16-second ensemble-mean bottom-track measurements of
bathymetric depth within each virtual channel, from this
individual crossing. Units of meters.
```

3c. Velocity observations – Time sequences at fixed locations

For each of the virtual stations, the velocities from all individual-transit datafiles have been aggregated in to a single time-sequence datafile. It is referred to as a time-sequence, not a time-series, because the times are irregularly spaced. The time-sequence files are named for their virtual station, for example "R01vel", "R02vel", etc. The 18 datafiles are not the same size because useful data may not have been collected from a subset of the virtual stations during a single transit, and if this is the case there is no data from that transit included in the time-sequence files for those virtual stations.

The ‘readme’ information distributed with the time-sequence datafiles is as follows:

```
1. vel_east & vel_north (47 x ntimes)
  Eastward and northward velocity, in cm/s, at 47 grid depths,
one value per ferry transit.
2. t_yr, t_mon, t_dy, t_hr, t_min, & t_sec (each 1 x ntimes)
  Each of these variables consists of one value for each
transit, the times in Eastern Standard Time (EST).
```

4. Near-surface water properties GDP

4a. Methods and quality control

The water properties measured are temperature, salinity, and total chlorophyll concentration from in vivo fluorescence calibrated by acetone extraction. They are made by a YSI model 6820 sonde, in a sampling chamber through which water is continuously pumped from an intake on the vessel hull. The intake is about 2-3 m deep and is part of the main vessel cooling system. Water passes through a screen with mesh size of 3 mm at the hull intake point. The sampling chamber is about 4 meters from the intake, and water is pumped to it prior to entering, or being influenced by, the cooling system.

The sensors record water property values nominally each 30 seconds. The transit time of the water from the intake to the sensor is about 30 seconds and the response times of the sensors are nominally between a few seconds and 30 seconds. Latitude-longitude values assigned to the data are therefore GPS (global positioning system) values, recorded by the satellite compass operating with the water velocity system, from one minute earlier than the times at which the water property sensors record water properties.

For quality assurance and quality control, data from each transit are plotted and inspected visually as a group, together with the data from the entire day. On this basis, each transit is assigned a quality level of either acceptable or unacceptable. The unacceptable data is identified unambiguously by sharp deviation of its temperature values from recent and subsequent (earlier/later that day) data values, which results from a lack of proper pumping action. This can result from the pump not being turned on, which occurs on occasion in connection with other activities carried out by the vessel crew, or from the pump losing its prime, which occurs on rare occasions. Transits designated unacceptable have been excluded from the GDP.

Among the transits designated acceptable, after mapping to virtual stations as described below, there are a small number (about 0.5%) of the salinity values that are unambiguously outliers. These are removed on the basis of exceeding 2.5 standard deviations from the mean of the values from all 18 virtual stations during that transit.

The temperature and salinity sensors require no manual calibration and the factory calibrations have been accepted. For chlorophyll, manual three-point calibrations using the acetone extraction method with natural water samples collected at Avery Point, near the northern end of the ferry transect, have been carried out and found to be stable throughout the GDP sampling period, with the exception of one calibration during winter months. That calibration is considered of lower reliability, due to low chlorophyll levels during that time of year. The averaged non-winter calibration results (slope 0.5538 and offset -0.0431) have been applied to arrive at micrograms per liter of Total Chlorophyll.

Based on manufacturer specifications, accuracies are +/- 0.15 degrees C for temperature values and +/- 1% for salinity values; resolutions are +/- 0.01 degrees C, and +/- 0.01 PSS, respectively. For chlorophyll values, the manufacturer specifies +/- 0.1 µg/l resolution, and based on experience of ours and our colleagues with the in vivo fluorescence method calibrated by acetone extraction, the estimated accuracy is +/- 30%.

4b. Grid characteristics

The horizontal grid (virtual stations, to which values are assigned by averaging samples in the associated virtual channel) for the near-surface water properties GDP is the same as that for the velocity data, as described above. The water properties samples are all drawn from a water intake in the ferry hull, 2-3 m deep. Regarding a time grid, as with the velocities the temporal sampling by the ferry at a given virtual station is irregular and therefore the sampling times are provided without being mapped to a regular grid. The grid data file is the same as for the velocities (see section 3b above).

4c. Near-surface water properties -- Individual transits

For each individual ferry transit with near-surface water property data of acceptable quality, there is a single datafile named with format “gridswp_YYYY_MM_DD_HH”, where YYYY is the year, MM is the month, DD is the day, and HH is the hour (00-23) in Eastern Standard Time (EST). (Local time is the same as EST, except during daylight savings time. During daylight savings time, which occurred between the first Sunday in April and the last Sunday in October in years through 2006, local time is one hour ahead of EST.) EST lags UTC (Coordinated Universal Time) by 5 hours. An example of an individual transit is given in Figure 3.

The ‘readme’ information distributed with the individual-transit datafiles is as follows:

```
1. salinity
  Salinity in Practical Salinity Scale units. A vector of
 values corresponding to virtual station locations "lat_vc"/
 "lon_vc" in the grid file (from south to north; G01, R17,
 R16, ... R01).
2. temperature
  Temperature in degrees Celcius. A vector (as salinity).
3. chlorophyll
  Total Chlorophyll (Chl a and others) in micrograms per liter
 based on in vivo fluorescence calibrated by acetone
 extraction. A vector (as salinity and temperature).
4. t_yr, t_mon, t_dy, t_hr, t_min, & t_sec
  Each of these variables is a vector holding one value per
 virtual station (see technical report): an average over data
 collected in the corresponding virtual channel. They are
 times in Eastern Standard Time (EST).
5. aux.numavg_temp, aux.numavg_sal, aux.numavg_chl
  Number of ~30-second interval measurements used in computing
 the gridded values. Each a vector just as 1-3 above.
```

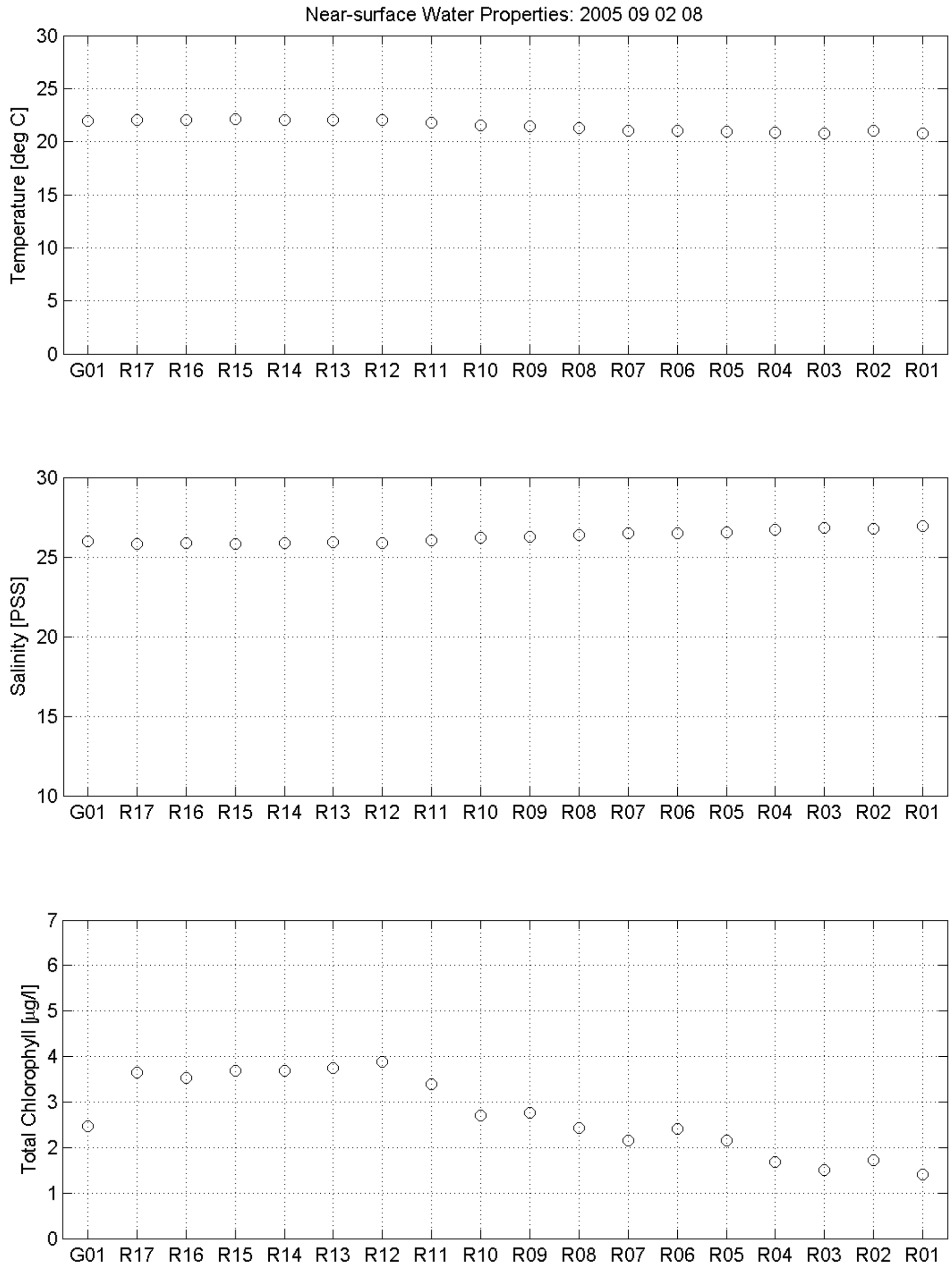

Figure 3. Example of near-surface water properties measured during a single ferry transit.

4d. Near-surface water properties -- Time sequences at fixed locations

For each of the virtual stations, the near-surface water properties from all individual-transit datafiles have been aggregated in to a single time-sequence datafile. It is referred to as a time-sequence, not a time-series, because the times are irregularly spaced. The time-sequence files are named for their virtual station, for example “R01swp”, “R02swp”, etc. The 18 datafiles are not the same size because useful data may not have been collected from a subset of the virtual stations during a single transit, and if this is the case there is no data from that transit included in the time-sequence files for those virtual stations.

The ‘readme’ information distributed with the time-sequence datafiles is as follows:

```
1. temperature, salinity, chlorophyll (each is 1 x ntimes)
 Each is a vector containing one value per ferry transit,
calculated as the mean of all measurements within the virtual
channel corresponding to this grid location (virtual station).
See grid info file for lat-lon coordinates of grid locations.
Temperature (degrees C), Salinity (PSS), and Total Chlorophyll
(micrograms/liter) from in vivo flourescence calibrated by
acetone extraction.
2. aux.numavg_vel, _sal, and _chl (each is 1 x ntimes)
 Number of ~30-second interval measurements used to calculate
the respective mean values in the three variables of 1 above.
3. t_yr, t_mon, t_dy, t_hr, t_min, t_sec (each 1 x ntimes)
 Each of these variables consists of one value for each
transit, times given in Eastern Standard Time (EST).
```

5. Data coverage periods, up to December 2006

a. Water velocity profiles GDP

The first reliable water velocity sampling was on Nov 22, 2002. The most recent sampling that has undergone data processing and is included in the GDP is December 15, 2006. During this interval, ADCP data was collected on 6,774 ferry transits. Among these, 4,632 passed quality control and are included in the GDP.

There are intervals when no sampling occurred. Some are due to the ferry not running, as a result of planned or unplanned vessel maintenance. Others are due to malfunction of the ADCP data collection and logging system.

Some of the longer periods when data was not collected because the ferry was not operating are as follows: between March 7 and April 18, 2003; between October 14, 2003 and March 3, 2004; and between January 6 and February 18, 2006.

Some of the longer periods when data was not collected due to instrument malfunction are as follows: between January 9 and June 15, 2005; and between September 6 and October 18, 2006.

b. Near-surface water properties GDP

The first near-surface water property samples in the GDP are from June 15, 2005 and the last are from September 5, 2006. There are a total of 2,255 ferry transits with acceptable quality during this interval.

Sampling from dates between February 1 and June 15, 2005 was also carried out and these data are available on request; they have not been included in the GDP because, as noted above, during this interval the water velocity sampling was not operating and latitude longitude positions were not recorded. These data are of good quality and could be of used for analyses that do not require accurate assignment of lat-lon positions, for example situations where mean values across the entire ferry transit would suffice. Sampling from dates prior to February 1, 2005 was also carried out and is available on request. It has not been included in the GDP because just prior to February 1, 2005 a different pump was installed, as a result of erratic behavior of the previous pump, which had been serving well for some time but then became unable to hold its prime as a result of changed vessel protocols for testing cooling system pumps.

Gaps in coverage in the near-surface water properties GDP, as for the water velocities GDP, result from either periods when the ferry was not operating or periods of instrument malfunction. Some of the longer periods when the vessel was not operating are listed above. Some of the longer periods when the instrument malfunctioned were as follows: between February 18 and March 8, 2006; between June 19 and July 1, 2006; between August 20 and September 4, 2006.

6. Complete subsets of both GDPs, bundled in a single compressed file

For ease of distribution, it is helpful to identify a minimal set of files containing only one non-redundant set of the GDP data values. Such a set is formed by the grid information file (described in section 3b), the 18 time-sequence files for water velocity profiles (described in section 3d), and the 18 time-sequence files for the near-surface water properties. This subset of files is not as extensive as the full collection presented at the project website, as it omits the individual-transect data files and the preview files, but it is nonetheless a complete distillation of the dataset.

A bundle containing this subset of 37 files, together with a PDF of this technical report, can be compressed to a single zipfile of modest size. Such zipfiles, containing only ASCII and Matlab "mat" format files, respectively, are available at the project website under the names

FOSTERLIS_CompleteSubset_Ascii_May2007.zip (about 15 megs), and
FOSTERLIS_CompleteSubset_Matlab_May2007.zip (about 22 megs).

The Matlab "mat" files include some created by Version 6 and some created by Version 7; users will need Matlab Version 7, which is capable of reading both types.

7. Acknowledgements

Funding to establish FOSTER-LIS came from the Connecticut Sea Grant College Program through award R/ER-21 in 2002. Continued support from 2004 to 2006 was provided via the Long Island Sound Study through cooperative agreement LI-97106001(-2) from the U.S. Environmental Protection Agency.

These datasets exist largely due to the generosity of Cross Sound Ferries, Inc. Thanks are due particularly to John and Adam Wronowski for their early and sustained support, in the form of in-kind use of the ferry for oceanographic sampling and providing the research team access to the vessel. Technical assistance from Dick Sise and the captains and crews of the *MV John H.* has been essential to project success.

Dirk Aurin managed sampling instrumentation, and developed data reduction and analysis methodology. Dennis Arbige installed and maintained system electronics, including navigation and real-time data display equipment. Gary Grenier designed and built custom mounting hardware for the ADCP sea chest, and the pump system for the water properties sensor system.

John St John and colleagues at Hydroqual Inc. provided suggestions on the GDP format.