Ocean Sci. J. (2013) 48(1):105-116
http://dx.doi.org/10.1007/s12601-013-0009-y

Article

Available online at http://link.springer.com

SpringerLink

Ocean Science Journal

A New Method for Estimating High-Frequency Radar Error Using Data from Central San Francisco Bay

Maxwell Hubbard[1], Donald Barrick[1]*, Newell Garfield[2], Jim Pettigrew[2], Carter Ohlmann[3], and Matthew Gough[4]

[1]*CODAR Ocean Sensors, Ltd., 1914 Plymouth Street, Mountain View, CA 94043, USA*
[2]*Romberg Tiburon Center for Environmental Studies, San Francisco State University, Tiburon, CA 94920, USA*
[3]*University of California Santa Barbara, 522 University Road, Santa Barbara, CA 93106, USA*
[4]*Naval Postgraduate School, 1 University Way, Monterey, CA 93943, USA*

Received 6 October 2012; Revised 3 January 2013; Accepted 28 January 2013

Abstract – This study offers a new method for estimating High-Frequency (HF) radar surface current velocity error in data comparisons with other types of instrumentation. A new method is needed in order to remove the zero-mean random spatial and temporal fluctuations present in surface-current measurements from all sensors. Conventional methods for calculating radar error when comparing with another instrument have included their root mean square differences and scatter plots that provide correlation coefficient and slope/intercept of the regression line. It seems that a meaningful estimate of radar error should attempt to remove both sensors' zero mean random fluctuations, inasmuch as possible. We offer and compare a method that does this. The method was tested on data collected in the Central San Francisco Bay, where GPS surface-drifter deployments were conducted within the coverage of four 42 MHz radars over six days in October of 2008. Drifters were continuously deployed in these areas over the sampling days, providing 525 usable drifter measurements. Drifter and radar measurements were averaged into thirty-minute time bins. The three-day long-term averages from the sampling areas were then subtracted from the thirtyminute averages to remove biases associated with comparisons done with short, disjoint time-sample periods. These were then used to develop methods that give radar error or bias after the random fluctuations have been removed. Results for error estimates in this study are commensurate with others where random fluctuations have been filtered, suggesting they are valid. The estimated error for the radars in the SF Bay is low, ranging from -7.57 cm/s to 0.59 cm/s.

Key words – HF radar, remote sensing, surface currents, sensor calibration, drifter, instrument bias

*Corresponding author. E-mail: don@codar.com

1. Introduction

The acceptance of HF radar as a valid instrument for measuring surface currents has been achieved through extensive validation studies (Paduan and Rosenfeld 1996; Chapman et al. 1997; Ullman et al. 2003; Kaplan et al. 2005; Paduan et al. 2006; Ohlmann et al. 2007). Generally, the validation studies involve comparing the measurements of the HF radar to that of another instrument, most commonly, the Acoustic Doppler Current Profiler (ADCP) or the surface drifter. Over time, the methods used in these instrument-toinstrument comparisons have evolved. This evolution is a response to the unavoidable truth that trying to compare measurements from different instruments in order to find the error of one is nebulous, and – although straightforward – is not well described as a root mean square difference (RMSdiff) or correlation coefficient. The issue is that the HF radar (HFR), ADCP, and surface drifter do not measure the same processes over the same spatial or timescales. Currents vary spatially across a given area due to turbulence near the water's surface. In addition, there is always sensor noise to contend with. Some common causes of turbulence include wind stress, breaking waves, and geographic influences such as headlands and underwater sills. HFR does not have the temporal or spatial resolution to determine details of the small-scale turbulence. When relying on a correlation coefficient or RMSdiff, differences in the measurements caused by natural, zero-mean random short-term current variability,

which we posit should not be considered instrument error, are essentially incorporated into the estimation of error of one instrument or the other, and generally tend to dominate the desired comparison metrics.

This study takes a step forward in the instrument-to-instrument comparison evolution by defining a method to isolate the error in HF radar measurements when comparing them to the measurements of GPS-tracked surface drifters in the Central San Francisco (SF) Bay. We start in the next section by examining the nature of sensor differences, based on the desired sensor output. We will define a persistent, long-term difference as a bias error, and that is the quantity we seek here. Then one can focus on reducing or removing this bias. The method devised here takes into account and removes the natural random variability in the current field that will be measured differently by the radar and drifter, and then defines a bias error estimate for the radar. The results of the new bias error estimate are presented alongside the results from conventional statistical comparison methods. The error estimates for the HF radars used in this study are low, and vary minimally between the individual radar stations The major objective of this manuscript is to present, exemplify, and justify our new metric for HF radar error.

2. Meaning of Sensor Differences and Bias

A basic approach to establishing sensor accuracy is to compare the output of one instrument to another's, with the assumption that one is more accurate than the other one under examination, and therefore serves as the "standard." Tools that are used for this purpose have been: (i) The difference between the outputs of the two sensors at the same points in space and time; (ii) A scatter plot of these points where the vertical and horizontal axes are the measured values from each sensor; (iii) The correlation coefficient between these points, which often accompanies the scatter plot. If the measurements agree perfectly, then: (i) Their difference is zero; (ii) The scatter-plot points fall on a straight 45° line; (iii) Their correlation coefficient is unity.

Before examining and evaluating these three comparison metrics, let's consider why the measured output data points may not agree, in order to guide discussion of what we mean by sensor accuracy.

Random fluctuations in sensor measurements

Considering as our example HF surface current measurements, there are two sources of zero-mean random errors in sensor measurements: (i) Turbulence at the surface induced by wind gustiness / wave breaking, and the random nature of the heights of the scattering waves themselves. These are sometimes referred to as sub-grid-scale random variability. (ii) Noise that gets added to the sensor signals, e.g. external atmospheric and/or radio noise, or receiver thermal noise. In nearly all cases, these fluctuations in each sensor's signals are uncorrelated with each other, because of the different natures of the two instruments.

Such random fluctuations are not desired by the user, and contaminate the geophysical quantity that the sensor was intended to measure. Ocean observers using HF radar and other current sensors are concerned about processes with time scales greater than a half hour, and space scales of a kilometer or larger. Although random fluctuations can be reduced by averaging over space and time, tradeoffs dictate upper limits to such averaging, otherwise required geophysical information is filtered out. Hence, there will be some level of random fluctuations remaining in the sensor outputs. Any comparison will be imperfect due to these residual, uncorrelated, small-scale sensor fluctuations.

Long-term bias or error in the sensor measurement

Comparisons that are meant to establish accuracy in the ability to observe a geophysical quantity of interest normally focus on assessing a bias or persistent error in the sensor's output. The present manuscript focuses on this type of error. For example, does imperfection in or deterioration of the instrument introduce a bias in its output, so that the sensed quantity consistently differs from the true value? In order to make this assessment, the comparison method must remove as much of the random fluctuations as possible, which are not desired nor are part of the long-term bias.

Which metrics capture the desired bias?

The correlation coefficient is almost meaningless in capturing the bias. When considering the scatter plot, it is a measure of the "point scatter" or dispersion about the best-fit regression line. Consider an example. Two instruments measure tidal currents, with no random fluctuations in either signal. One sensor is biased so that it measures the same tidal temporal pattern, but its amplitude is low by 50%. The correlation coefficient will be unity (perfect), but the bias is unacceptable, i.e. 100% error by one measure. Likewise, the scatter plot will show no point dispersion

because they all fall along a single line. Here, however, the line will not lie along the unbiased 45° position, but its slope departure from this (as well as its non-zero intercept) is a measure of the bias. Sensor data difference metrics will also capture the bias; a long-term mean difference sees only the bias, whereas RMS difference (RMSDiff) sees bias as well as the turbulent fluctuations that are not part of the bias. The main reason for this is that the zero-mean random fluctuations are squared along with the bias in the RMSDiff before the averaging, making it harder for them to be averaged out.

As another example, if there is zero-mean random variability in the sensor outputs but no bias between them, then the correlation coefficient may be poor, and RMS difference may be large, but these poor metrics will have nothing to do with sensor bias, which is zero. Based on these considerations and examples, we develop two metrics in this paper that separate the instrument bias from the random signal components that may be present and are "noise" as far as the desired geophysical quantity, but that must be dealt with in another way (spatial and/or temporal averaging/filtering).

Sources of bias in HF radar

The three observables in HF radars for surface current measurement are range, radial velocity (from Doppler), and echo bearing. The dominant source of error in HF radar systems has long been recognized to be bearing measurement; the other two observables are precisely determined. Estimating bearing in HF radars is done one of two ways: phased-array beam forming (BF) and scanning, and direction finding (DF) among the signals from multiple receive antenna elements. The source of bearing error for both is distortion of the antenna pattern. Such distortion is caused by near-field obstacles and imperfections near the receiving antenna elements (e.g. buildings, poles, feedline cables not isolated from the antennas, corroded cable connectors, etc.). When this distortion is not measured, or it changes over time due to system hardware degredadation, then errors creep in. An error in echo bearing results in an error in current velocity, because a correct vector is not placed in the proper position on the map. All bearing determination methods are based on an algorithm that assumes a known antenna pattern.

There has been an extensive literature on antenna pattern measurement with DF types of HF radars (e.g. Kohut and Glenn (2003), Paduan et al. (2006), Laws et al. (2010)). In terms of outcomes for velocity errors, these are typically shown to be reduced by several centimeters per second by proper measurement of antenna patterns and its calibrated use in bearing determination. The reader is referred to these works, and we do not go into details here regarding this subject.

There has been clear evidence that pattern distortions are equally significant with phased-array beam forming (BF) (Teague 1986). That community of users, however, is considerably smaller than DF radar operators. As a result, methods have not been tested or widely accepted for reducing these biases in BF phased arrays, although significant biases (Mariette et al. 2006) suggest this is a major source of error for these systems also.

3. Instrumentation

HF radar

The HF radars used in this study are located in the Central SF Bay. Due to the dynamic tidal current regime and short distance scales involved, the 42 MHz short-range/high-resolution SeaSonde® sensor manufactured by CODAR (Coastal Ocean Dynamics Applications Radar) Ocean Sensors was chosen for deployment in the Bay. The four radar stations used in this study are located at Crissy Field, on a National Oceanic Atmospheric Administration (NOAA) pier (CRIS, center frequency = 43.69 MHz); at the Sausalito water treatment facility (SAUS, center frequency = 44.21 MHz); on the Tiburon Peninsula at the Romberg Tiburon Center for Environmental Studies (RTC1, center frequency = 41.48 MHz); and on the northwest side of Treasure Island (TRES, center frequency = 40.75 MHz) (see Fig. 1). Hereafter the radars are referred to by their four letter site codes; CRIS, SAUS, RTC1, and TRES.

Each HF radar station has an individual coverage area within which radial currents are measured (Fig. 1). Radial currents are observed traveling toward and away from the radar, in a polar coordinate system. The radial measurements provide near surface-current speeds that are representative of current velocities in the top ~0.5 m of the water column. The array in the San Francisco Bay has an approximate coverage area of 10 km by 12 km, the typical range of 42 MHz systems being within 5-12 km. The 42 MHz radars have a 400 m radar range-cell bin resolution where a radial velocity is calculated every 30 minutes (fifty-five minute average current velocity). *The thirty minutes is centered on fifty-five minutes of spectral data, which is standard data averaging times for 42 MHz SeaSonde systems.* The radial

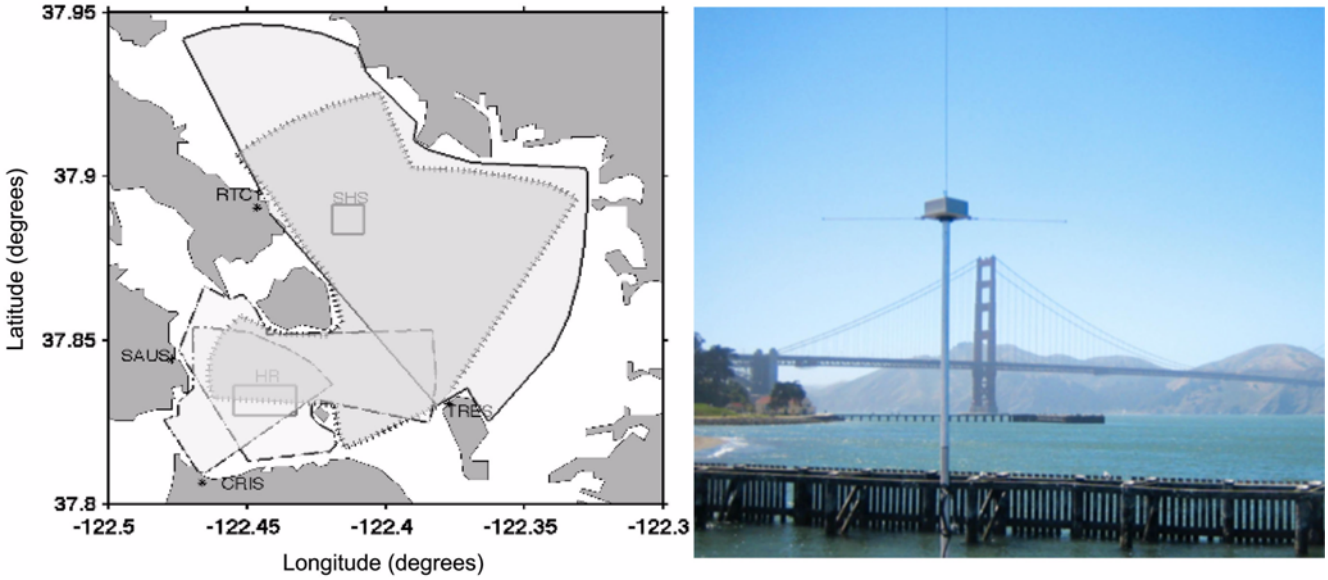

Fig. 1. Right side (a), Central San Francisco Bay. Black stars represent radar stations starting from the bottom at Crissy Field (CRIS), then moving clockwise to Sausalito Water Treatment facility (SAUS), Romberg Tiburon Center for environmental studies (RTC1), and Treasure Island (TRES). Grey filled areas represent radar measurement locations for each radar. Black boxes represent the two areas of drifter sampling, Southampton Shoal (SHS) to the North and Harding Rock (HR) in front of the Golden Gate. Left side (b), photograph of the radar antenna at Crissy Field

bins are spaced at 1° angular bearing steps around the antenna over annuli that define range cells.

Surface drifters

The drifters used in this study are Microstar® surface drifters from Pacific Gyre. They were outfitted with a GPS receiver that recorded latitude and longitude every 10 minutes. During the SF Bay deployments the drifters were tethered to a diamond shaped nylon or canvas drogue that was suspended at one meter depth. With the drogue at this depth, the drifters are thought to accurately portray the movement of a water parcel near the surface similar to the radar measurements, as discussed in Ohlmann et al. (2005).

4. Methods

SF Bay is a busy commercial estuary with large ships using the major deep-water channels and high-speed ferries crossing the Bay on multiple routes. Drifter deployment was constrained in order to avoid ferry and commercial shipping routes. As a result, sampling boxes were positioned around Southampton Shoal (SHS) and Harding Rock (HR) (see Fig. 1). The SHS sampling box location includes the radial coverage areas of the RTC1 and TRES radars, while the HR sampling box includes the CRIS and SAUS radar coverage areas. TRES and RTC1 had 20 and 41 radial bins within the SHS sampling box respectively (see Fig. 2). The HR sampling box contained 52 radial bins from the CRIS radar and 87 bins from the SAUS radar.

In October of 2008, up to 17 drifters were deployed over six separate days from three San Francisco State University research vessels. During the six days of sampling, a total of 525 10-minute drifter velocity observations were used from the total collected. Each site was sampled for three days; sampling days at SHS were October 8, 9, and 21; and October 15, 16, and 22 for HR.

One of the main goals for drifter deployment was to obtain as many measurements across the same group of radar radial bins as possible in order to understand better the statistical relationships between the two measurements. To achieve this goal, drifters were deployed so that the current would move them through the sampling boxes; when drifters exited the sampling box, they were captured by the research vessels and redeployed along the up-current boundary of the sampling box, as described in Ohlmann et al. (2007). To account for the multiple radar bins that the drifter movement covered within the sampling boxes, the midpoint between two drifter track readings was paired with the closest radial bin center point.

Individual drifter tracks were visually inspected for

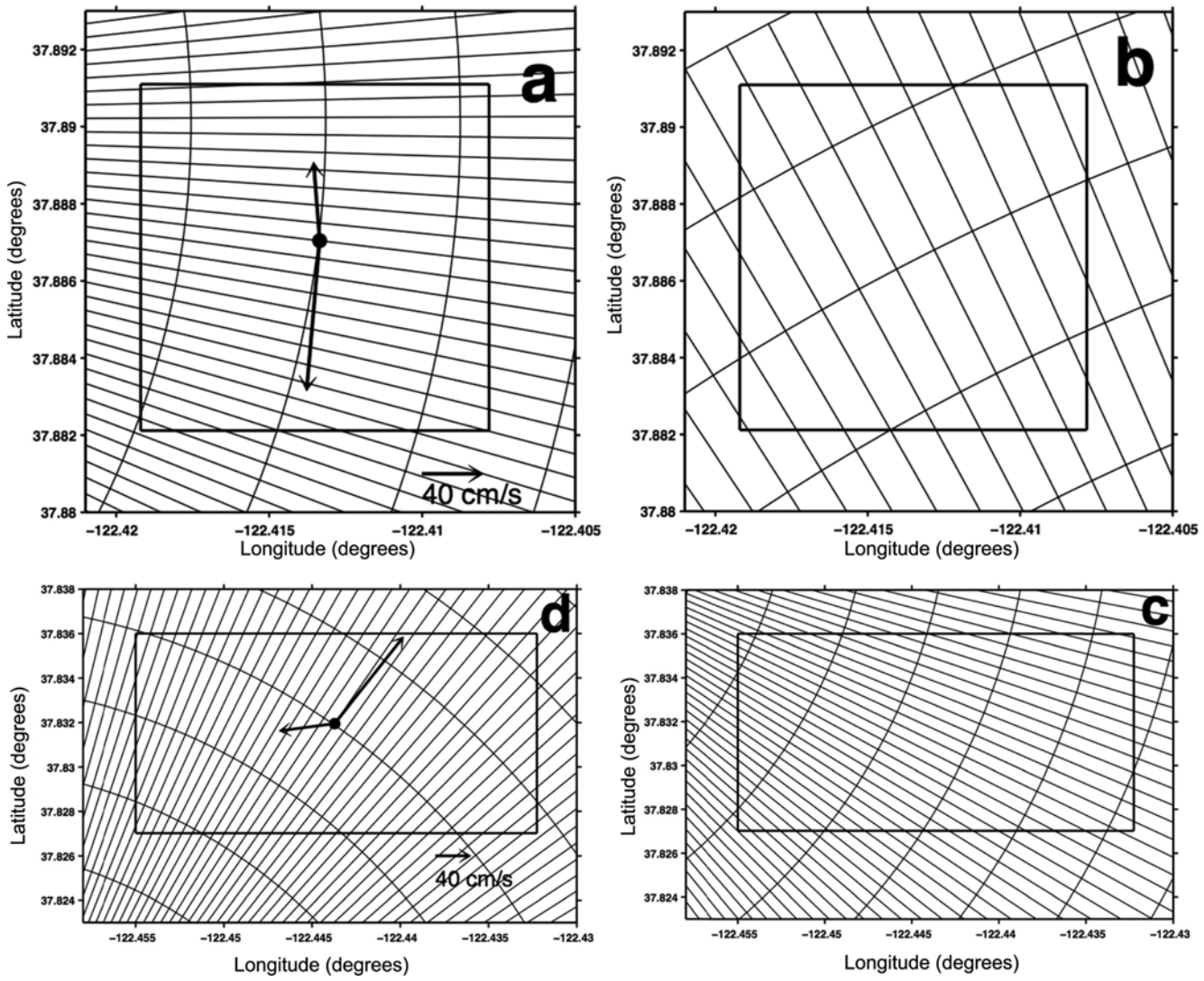

Fig. 2. Radar bins used for comparison within the sampling boxes. Clockwise from top left; (a) RTC1 (SHS sampling box), (b) TRES (SHS sampling box), (c) SAUS (HR sampling box), and (d) CRIS(HR sampling box). Black arrows in the boxes to the left (a and d) represent the average current flow in both the ebb and flood phases of the tide during sampling times for each sampling box. Average flows were taken as the top third of currents during the flood and ebb phases during the study days

measurements made within the sampling box. Suspect drifter measurements were identified based on abrupt angle changes, measurements with the opposing velocity, or dramatically increased magnitude compared to the average measurement. Measurements with these attributes meant that the drifter was most likely being transported on a boat while still logging position. During the experiment, the upper limit set for radial current measurements on the Bay radars is 200 cm/s; drifter speeds above this limit were discarded. A first level of quality control on the radar data happens within the software. No post-processing quality control was done on the radar data for this comparison, and all radar measurements that were paired with a drifter measurement meeting the criteria outlined above were used.

Drifter velocities were decomposed into vr^d, the radial component corresponding to the radar bin measurement with which the drifter track was being compared. Drifter velocity u (east-west) and v (north-south) components were transformed into the radial velocities using

$$vr^d = u\sin\theta + v\cos\theta \quad (1)$$

where u and v are the original east-west and north-south velocity components of the drifter movement, and θ

Table 1. Drifter comparison results for CRIS. Statistics in the table are calculated from the half-hour averages over the three-day deployment periods: slope of the best-fit line; y-intercept; r^2; root-mean-square difference; radar mean radial velocity and mean standard deviation (mean standard deviations were calculated over entire sampling day for both radar and drifter); radar maximum radial velocity; radar minimum radial velocity; drifter mean radial velocity and mean standard deviation; drifter max radial velocity; drifter minimum radial velocity; and number of paired observations. All radial velocities are in cm/s

Date	Slope of Best Fit Line	y-intercept cm/s	r^2	RMSDiff cm/s	HF Mean & Std. Dev. cm/s	HF max cm/s	HF min cm/s	Drifter Mean & Std. Dev. cm/s	Drifter max cm/s	Drifter min cm/s	No. of half-hr avgs.	No. of total obs.
10/15/08	1.2	17.71	1.0	17.03	-8.63±17.43	73.94	-106.3	-22.03±7.11	50.09	-99.26	6	41
10/16/08	1.05	10.32	0.84	16.18	66.78±12.48	0.31	-116.4	-73.69±7.68	-12.28	-122.73	11	76
10/22/08	1.03	2.45	0.75	9.06	24.13±8.53	55.12	-1.23	21.15±15.4	45.4	0.36	9	98
3-Day Total	1.03	7.84	0.95	14.36	-21.89±12.23	73.94	-116.4	-28.94±7.93	50.09	-122.73	26	215

Table 2. Drifter comparison results for SAUS. Statistics are the same as Table 1

Date	Slope of Best Fit Line	y-intercept cm/s	r^2	RMS-Diff cm/s	HF Mean & Std. Dev. cm/s	HF max cm/s	HF min cm/s	Drifter Mean & Std. Dev. cm/s	Drifter max cm/s	Drifter min cm/s	No. of half-hr avgs.	No. of total obs.
10/15/08	1.18	3.39	0.98	9.28	-12.71±12.43	44.84	-91.72	-13.6±5.96	34.41	-79..09	6	42
10/16/08	1.03	-4.11	0.81	13.89	-55.89±11.21	3.04	-102.05	-50.42±5.91	3.49	-76.68	10	81
10/22/08	0.70	9.97	0.88	8.77	37.64±10.27	59.89	9.17	39.57±7.81	69.74	11.24	10	109
3-Day Total	1.03	-2.42	0.96	11.12	-9.95±11.16	59.89	-102.05	-7.31±6.74	69.74	-79.09	26	232

Table 3. Drifter comparison results for TRES. Statistics are the same as Table 1

Date	Slope of Best Fit Line	y-intercept cm/s	r^2	RMS-Diff cm/s	HF Mean & Std. Dev. cm/s	HF max cm/s	HF min cm/s	Drifter Mean & Std. Dev. cm/s	Drifter max cm/s	Drifter min cm/s	No. of half-hr avgs.	No. of total obs.
10/8/08	1.53	-4.64	0.89	6.62	18.7±7.79	32.27	-8.63	15.26±3.89	24.75	0.71	9	63
10/9/08	0.89	3.86	0.95	6.8	4.75±4.72	33.9	-27.71	1.0±5.03	32.87	-35.61	7	94
10/21/08	0.76	6.3	0.88	10.49	15.43±3.99	37.96	-30.48	12.01±3.65	39.87	-41.14	14	136
3-Day Total	0.83	5.26	0.88	8.67	13.92±5.31	37.96	-30.48	10.41±4.06	39.87	-41.14	30	293

Table 4. Drifter comparison results for RTC1. Statistics are the same as Table 1. * Indicates a low value due to an abnormal outlier in the HF data. To stay consistent and not bias the results all HF data points that were matched with a drifter point were used in the analysis

Date	Slope of Best Fit Line	y-intercept cm/s	r^2	RMS-Diff cm/s	HF Mean & Std. Dev. cm/s	HF max cm/s	HF min cm/s	Drifter Mean & Std. Dev. cm/s	Drifter max cm/s	Drifter min cm/s	No. of half-hr avgs.	No. of total obs.
10/8/08	1.38	-3.03	0.82	3.47	5.38±7.77	12.27	-8.72	6.11±4.74	10.9	-3.09	8	60
10/9/08	0.56	1.29	0.67	6.98	8.14±9.60	18.71	-1.33	12.15±3.51	24.23	-7.1	7	90
10/21/08	0.21	3.54	0.02*	13.57	4.05±6.56	35.44	-11.68	2.39±3.60	13.43	-15.41	14	136
3-Day Total	0.44	2.87	0.16	9.89	5.4±7.63	35.44	-11.68	5.77±3.90	24.23	-15.41	29	286

represents the radial direction measured clockwise from 0° north (this radial direction corresponds directly to the radar bin at that location). Due to the mixed semi-diurnal tides of the area, radial current velocities were sampled both in the negative (moving away from the radar) and positive (moving towards the radar) direction for all the radars in the study. Linear regressions along with basic statistics between the drifter half-hourly velocity means and radar half-hourly velocity means were calculated on a day-to-day basis as well as for the entire sampling period at each individual radar site (Tables 1-5). Fig. 3 shows all the drifter velocities captured on October 22 for the CRIS radar and the corresponding radar bins used for comparison, as well as a time series of all the measurements from October 22 before they were averaged into half-hourly bins.

5. HF Radar Error Estimate Definitions

A fundamental difference between the drifter and radar data is that the radar is measuring backscattered signals off

Table 5. Drifter comparison results for the entire experiment. Statistics are the same as Table 1

Date	Slope of Best Fit Line	y-intercept cm/s	r^2	RMSDiff cm/s	HF Mean & Std. Dev. cm/s	HF max cm/s	HF min cm/s	Drifter Mean & Std. Dev. cm/s	Drifter max cm/s	Drifter min cm/s	No. of half-hr avgs.	No. of total obs.
Entire sampling period	0.98	1.81	0.93	11.17	-2.29±8.8	73.94	-116.42	-4.17±5.44	69.74	-122.73	111	1026

Table 6. Results for MD1 and MD2 in cm/s

	CRIS	SAUS	TRES	RTC1
MD1 (cm/s)	-7.05	2.64	-3.51	0.37
MD2 (cm/s)	-7.57	-0.95	-4.55	0.59

the ocean surface, while the drifter is measuring differences of GPS positions. Another difference encountered between the two measurements is the variability of the currents in a given area due to turbulence caused by breaking waves, wind stress, and geographic influences. This is where the difference in time and space scales becomes apparent between the two measurements, especially in a complex tidally driven area like the San Francisco Bay. Since it is impossible to align the drifter and radar data in space and time so that they observe exactly the same current, these two sets of measurements will have different variations around their means, the variations in this case being produced by the currents of the Bay. When analyzing the measurements of one instrument independently, these variations are not normally separated from the mean, and are considered to be part of the measured current. However, if left in the equation when determining the error of one instrument based on the measurements of the other, they have been assigned as error in one or the other instrument. This section describes the process we propose for removing the natural variability of the surface currents measured by the radar and drifter, thereby leaving an improved estimate of radar error or bias when compared to the drifter. Two methods outlined below are based on the idea that a portion of the measurements from the drifter and radar reflect the short-term variations of the flow field over time.

The mean of the radar measurements and the mean of the drifter measurements within the sampling box for each 30-minute sampling interval over the three-day period is given by:

$$Mn_n^d = \frac{1}{I}\sum_{i=1}^{I} vr_j^d \tag{2}$$

$$Mn_n^r = \frac{1}{J}\sum_{j=1}^{J} vr_j^r \tag{3}$$

where I represents the number of drifter estimates of radial velocity and J represents the number of radar estimates of radial velocity in each sampling box during each half-hourly time interval. This provides a time-series of half-hourly means over the three days of sampling, subscript n

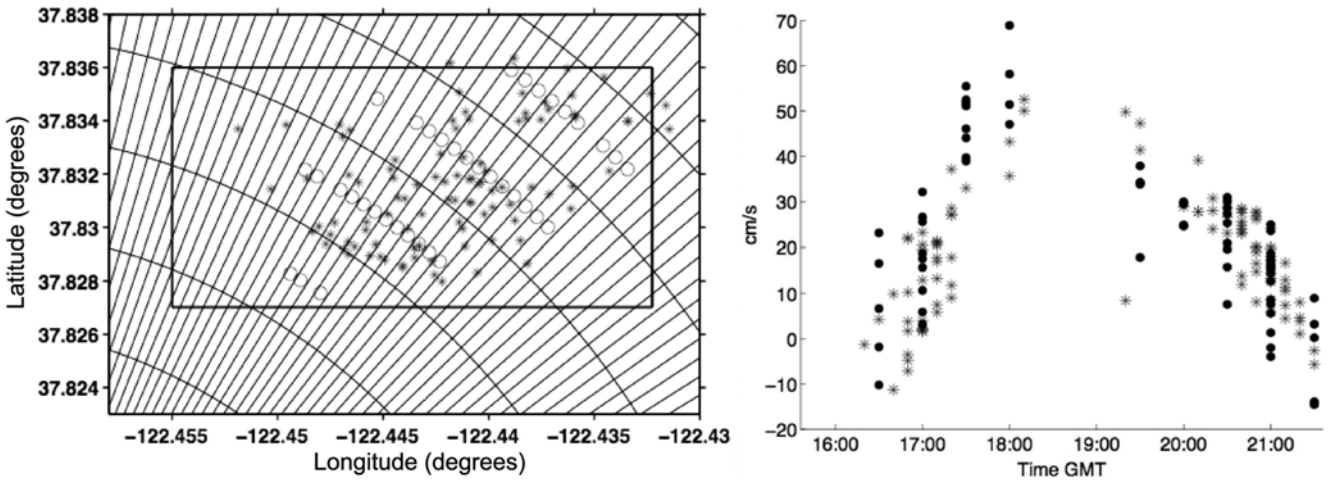

Fig. 3. Left (a), locations of drifter radial velocity measurements locations (star) and corresponding radial bins (circles) used for comparison at CRIS, in the HR sampling box on the 22nd. Right (b), time series of drifter radially directed measurements (star) and radar radial measurements (circles) for the 22nd before averaging

being the index for the specific half-hourly average. Here, *vr* represents radial velocity, superscript *d* denotes the drifter while superscript r denotes the radar.

We note that tides and other trends are normally removed by performing a tidal analysis on a long, (e.g. several week) continuous time series of measurements, using a technique similar to T_TIDE (Pawlowicz et al. 2002). This is not possible in comparisons such as this, where drifter deployments are conducted for a few daytime hours, and then only over several days. In fact, what one ends up with are disjoint pieces of three time-series sections of perhaps five hours each over this period. An example is shown in Fig. 3(b). As a result, we have devised the method below to remove bias inherent from analysis of the three short disjoint pieces, which do not even comprise one complete tidal cycle.

To deal with the bias introduced by needing to deal with these piecewise disjoint data samples, we define an average of the half-hour samples over the entire three-day period. We then subtract these means from the individual half-hourly samples, defined in Eq. (4) below. To verify that this reduces the piecewise disjoint bias, we show a scatter plot in Fig. 4 of radar vs. drifter data from CRIS: (a) represents all points before half-hour averaging. Panel (b) is the reduced data-sample set after the half-hour averaging of Eq. (2) and (3). Finally, (c) shows the scatter after we subtract the three-day mean, defined in Eq. (4) below, from the half-hour samples. Note that for (c), the regression (dash) line is closest to the ideal, no-bias 45° solid line, meaning the piecewise bias of (b) has been minimized. We represent this bias removal as:

$$\beta_n^d \equiv Mn_n^d - \frac{1}{N}\sum_n^N Mn_n^d;\;\; \beta_n^r \equiv Mn_n^r - \frac{1}{N}\sum_n^N Mn_n^r \qquad (4)$$

Again, the running index every half hour is n, and the total number of half-hour intervals in the three-day measurement period is N. Note that the tidal variations, along with the zero-mean half-hour fluctuations, are still inherent to our β_n above.

Let us now decompose the fluctuating components, β_n^d, β_n^r into a common time-varying portion μ_n (e.g. tides) and a zero-mean departure from this common term $\delta_n^{d,r}$:

$$\beta_n^d = \delta_n^d + \mu_n;\;\; \beta_n^r = \delta_n^r + \mu_n\,. \qquad (5)$$

Next, let us express the difference between our two sets of measurements for every interval (e.g. one-half hour here):

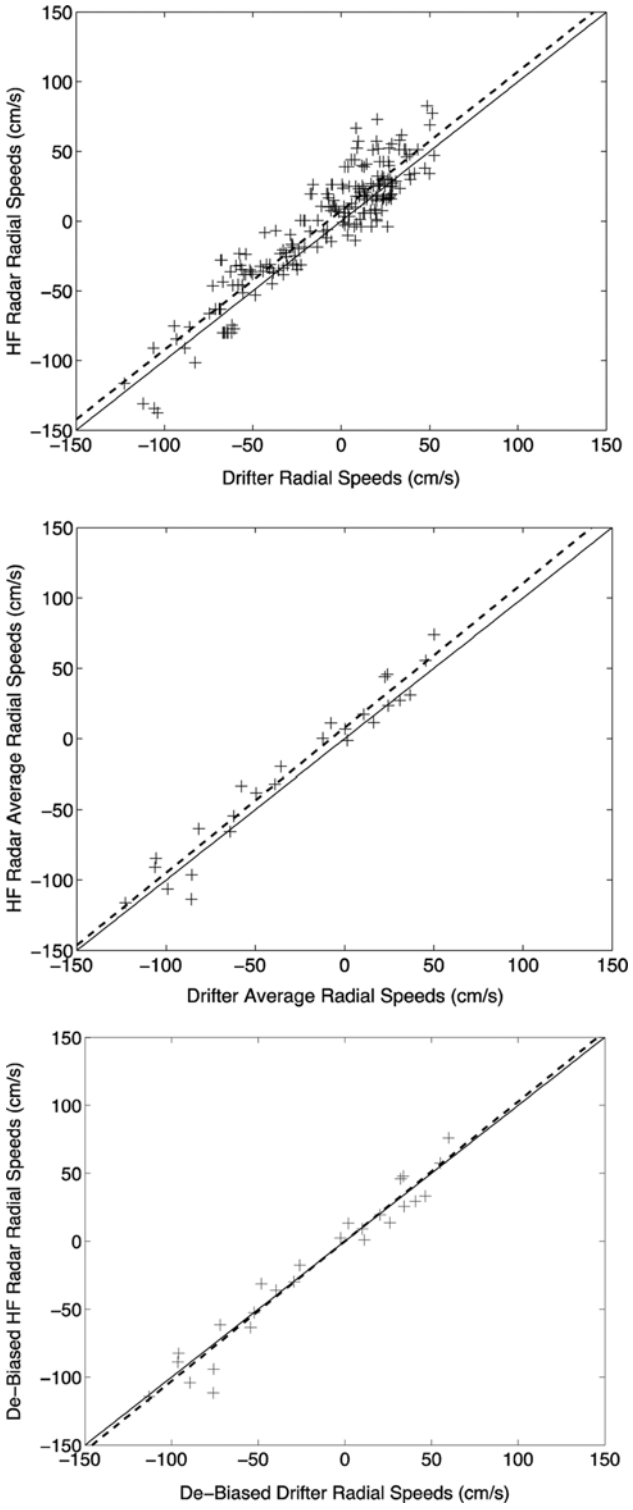

Fig. 4. Comparisons for the entire three days of sampling at CRIS. Best fit lines are dashed and the ideal one-to-one lines are solid. Top (a), scatterplot and best fit line for raw unaveraged drifter and HF radar radial velocity pairs. Middle, scatterplot and best fit line for the drifter and HF radar half-hourly averages. Bottom, scatterplot and best fit line for the drifter and radar half-hourly values that are de-biased using Eq. (4)

$$D_n \equiv \beta_n^d - \beta_n^r + Md_N^{(l)} = \delta_n^d - \delta_n^r + Md_N^{(l)}. \quad (6)$$

or

$$D_n = \Delta_n + Md_N^{(l)} \quad (7)$$

where we define a zero-mean random component including both drifter and radar contributions to be:

$$\Delta n \equiv \beta_n^d - \beta_n^r = \delta_n^d - \delta_n^r, \quad (8)$$

and $Md_N^{(l)}$ is the mean difference – or bias – between the two measurements, independent of time. The superscript l will denote one of two methods that we will define below to estimate this mean bias.

Method 1. Definition for mean bias

Our first definition of bias error is merely the mean difference between all of the N half-hourly points over a long time, which here is the three-day period measurement span – a rather obvious estimate by considering Eqs. (2) and (3) and simply subtracting.

$$Md_N^{(1)} \equiv \frac{1}{N}\sum_n^N (Mn_n^d - Mn_n^r) \quad (9)$$

Squaring and averaging the difference between Eqs. (2) and (3), i.e. the parenthetical expression in the summation of Eq. (9) above, defines the mean-square difference of these half-hourly points. This is given below in Eq. (10).

$$MSDiff_N \equiv \frac{1}{N}\sum_n^N [Mn_n^d - Mn_n^r]^2. \quad (10)$$

This will be used in subsequent calculations. It has been the conventional metric for sensor error that we want to replace because it inherently contains the zero-mean random fluctuations that we argue should not be included as an instrument bias.

Method 2. Definition for mean bias

As a second estimator of mean error or bias, we start with the textbook definition of mean-square difference, which we will then equate to Eq. (10) above. This is based on simply partitioning any random variable into two parts: a zero-mean fluctuating quantity Δ_n and a mean over the N samples. The latter mean is in fact our second definition of mean error or bias, which we denote as superscripted with (2). This partitioned random variable is the quantity in square brackets of Eq. (11) below. Thus the known quantity on the left side, i.e. the mean-square difference, is known from Eq. (10).

$$MSDiff_N \equiv \frac{1}{N}\sum_n^N [\Delta_n + Md_N^{(2)}]^2. \quad (11)$$

Thus the mean-square difference on the left side of Eq. (11) above is a known quantity that we get from applying Eq. (10) to the N half-hourly data samples. The zero-mean random fluctuations are also known, and defined in Eqs. (4) through (8). The unknown is our second definition of mean error or bias, denoted by the $Md_N^{(2)}$ term within square brackets of Eq. (11). Its solution is developed below into a quadratic equation that is solved thusly by expanding the square-bracket summation quantity to get:

$$MSDiff_N \equiv \frac{1}{N}\sum_n^N \Delta_n^2 + \frac{2Md_N^{(2)}}{N}\sum_n^N \Delta_n + (Md_N^{(2)})^2 \quad (12)$$

which is a quadratic equation with the standard form:

$$x^2 + bx + c = 0 \quad (13)$$

the unknown, x, being our second estimator of mean difference, $Md_N^{(2)}$. All of the remaining quantities in Eq. (12) have already been derived from the data. And so our unknown mean-difference estimator can be solved with

$$x = Md_N^{(2)} = \frac{-b \pm \sqrt{b^2 - 4c}}{2} \quad (14)$$

where

$$b \equiv \frac{2}{N}\sum_n^N (\beta_n^d - \beta_n^r) = \frac{2}{N}\sum_n^N \Delta_n, \quad (15)$$

and

$$c \equiv \frac{1}{N}\sum_n^N \Delta_n^2 - MSDiff_N, \quad (16)$$

providing a solvable quadratic equation with two possible outcomes. Because we have a separate, earlier definition of mean/bias, $Md_N^{(1)}$, consistency suggests we select the outcome closer to that result. We discuss next why our two estimates of bias are in fact different from each other when dealing with a finite number of samples, N.

6. Reconciling Different Sensor Bias Definitions

We raise and examine here some fundamental questions related to our approach to examining mean error or bias.

Question #1: Why have we derived two estimates for mean error or bias?

Question #2: Why do the two definitions differ?

Question #3: A third method is possible if we express the biases in terms of infinite ensemble averages rather than N-sample averages, and they are much simpler. What are the differences, and why don't we use the simple infinite-ensemble definition?

In answer to the first two questions, we note that in the first case of $Md_N^{(1)}$, we used straight definitions of means over different time periods, comprising N samples, to get rid of the short-term, smaller-scale zero-mean fluctuations. In the second case leading to the quadratic solution for $Md_N^{(2)}$, we used the mean-square difference of the two sensors' N samples. Both of these methods are easy to apply to the measured data, and both lead to bias estimates that appear different. As we will show, this difference is due to the N-finite-sample averaging we must do with a countable set of measurements.

To understand the above statement, let us define mean error or bias in terms of infinite-ensemble averages (rather than N-sample averages). We will denote this infinite-ensemble average of p using braces, as $<p>$. Then in place of Eq. (11) we would write:

$$MSDiff_\infty = \langle(\Delta_n + Md)^2\rangle = \sigma_d^2 + \sigma_r^2 + Md^2 , \qquad (17)$$

where we use Eq. (8) and define infinite-ensemble variances of the two zero-mean sensor fluctuations as: $\sigma_d^2 \equiv \langle\delta_d^2\rangle$ and $\sigma_r^2 \equiv \langle\delta_r^2\rangle$. Averages of cross-product terms in expanding the squared quantity in Eq. (17) are zero, i.e. $\langle\delta_d\delta_r\rangle$, because the zero-mean random fluctuations observed by the two sensors are uncorrelated when an infinite number of samples are averaged, due to the different natures of their measurements. Then we can define an infinite-ensemble mean error or bias by solving Eq. (17) to get:

$$Md = \sqrt{MSDiff_\infty - \sigma_d^2 - \sigma_r^2} . \qquad (18)$$

So now we have yet another definition of mean error or bias, valid when we have an infinite number of samples. Why not use this, along with or instead of $Md_N^{(1)}$ and $Md_N^{(2)}$? When we examined use of this for sample sets where N was not that large, as in our cases here involving data taken over several hours during three days, we found that occasionally the argument of the radical in Eq. (18) was negative, meaning the result was imaginary. This is a consequence of using an expression for an infinite-ensemble average when the number of samples is finite. It turns out that the argument in the radical of Eq. (14) for $Md_N^{(2)}$ can never be negative for any number of samples, so this quandary does not occur. Hence, we decided that it was more accurate and appropriate to stay with the finite-ensemble-averaging approach we introduced here.

With this background on infinite-ensemble averaging, the relationship and commonality among our definitions can now be clarified. Examining Eq. (15), we see that -- when considering an infinite number of samples -- b becomes zero and c becomes $\sigma_d^2 + \sigma_r^2 - MSDiff_\infty$. Therefore, we have $Md_{N\to\infty}^{(2)} \Rightarrow Md$, from consideration of Eq. (18). We see that it is the finite number of samples, N, in our averaging that leads to a small degree of uncertainty in these estimates of bias, which disappears as the number of samples becomes increasingly large. And finally, if the zero-mean random fluctuations disappear so that $\sigma_d = \sigma_r = 0$ and $MSDiff_N = (Md_N^{(2)})^2$, i.e. the mean-square difference is merely equal to the mean difference squared, then comparison of Eqs. (10) and (11) with (9) shows that $Md_N^{(2)} \to Md_N^{(1)}$. Hence, all of our definitions of mean error or bias are mutually compatible, and agree perfectly in the limit as $N \to \infty$.

7. Results

We observe that HF radar error -- or mean bias -- for this study is low. This is true when the error is either calculated with both the traditional mean difference, Eq. (9), as $Md_N^{(1)}$, or as the one introduced in this paper, $Md_N^{(2)}$, in Eq. (11). Table 6 -- which is the principal result of this study -- summarizes the traditional mean difference; CRIS = -7.05, SAUS = 2.64, TRES = -3.51, and RTC1 = 0.37; while the quadratic method of determining error gives: CRIS = -7.57, SAUS = -0.95, TRES = -4.55, and RTC1 = 0.59. Although these mean differences or biases could potentially be due to either radar or drifter, let's assume here it is due entirely to radar (worst case for radar); a rationale might be that the mean velocity of the drifters (after removal of their associated zero-mean randomness) is a more true, acceptable measure of actual current flow. The results in Table 6 cover the three-day sampling periods at each radar. Of the two radars positioned to look just inside of the Golden Gate (HR), SAUS has the lower error. At the SHS sampling area RTC1 has the better results for radar error. These slight differences

in the measured error between the different radar stations suggest better radar antenna patterns due to a cleaner near-field environment around the radars.

In addition to our new HF radar error estimations in Table 6, statistics for daily sampling periods are shown in Tables 1-5 for the individual radar stations. Note the traditional RMSdiff values previously used for estimating instrument error in Tables 1-5 for the entire three-day sampling periods are significantly higher than our new error/bias estimates. This is to be expected because our new method for determining error has removed the random zero-mean short-term fluctuations which we maintain are not really sensor errors. As previously stated, this allows us to focus on the real mean bias or difference between currents measured by the individual instruments.

8. Conclusions

A new, more meaningful estimate of HF radar error is defined herein. The motivation for such a method is the fact that local short-term zero-mean random fluctuations due to small-scale non-relevant upper-ocean turbulence measured independently by two different sensors is not an error in either instrument, and should be excluded from any attempt at meansuring error or bias. Our metrics for mean differences or biases based on the methods outlined above fall between -7.57 cm/s and 0.59 cm/s for all of the radars used in this study (Table 1). This is based on radial current speeds with a range of approximately -140 cm/s to +80 cm/s (Table 5). The bias for CRIS (-7.57 cm/s) is nearly twice that of the remaining three radars. We feel this larger bias is explained by its highly distorted antenna pattern caused by the nearby obstructions necessitated by the very small space permitted by authorities for the antenna mounting (Fig. 1); even though transponder pattern measurements remove most of this error, severe distortion will always leave a residual bias, which we are seeing in this case.

This suggests that the SF Bay radars are functioning properly and that their geometry and measured patterns at the time of the study were acceptable, with that at CRIS known and shown to be the worst of the three. The fact that both mean differences or biases in Table 6 come out to be nearly the same indicates that the new method works well. The small differences are due to the finite ensemble averaging employed, in contrast to an infinite ensemble. Future work lies ahead in understanding better and improving the HF radar error, the work presented here being just one more step in the evolution of instrument-to-instrument comparisons.

In general, radial velocity bias errors depend on bearing angle. In this study, different bearing angle sectors were viewed among the four sites. However, a full bearing span for each radar's field of view was not analyzed here; this constitutes a more extensive measurement campaign. That should be a future goal. The present study offers perhaps a representative sampling over bearing among the four radars inside San Franciso Bay and discusses physical reasons for the larger biases.

Acknowledgment

Special thanks to Regan Long for the picture of the Crissy Field radar and her contribution to my learning. We would also like to thank the Geosciences department at San Francisco State University and the Marine Operations Crew of Romberg Tiburon Center for Environmental Studies.

References

Chapman R, Shay L, Graber H, Edson J, Karachintsev A, Trump C, Ross D (1997) On the accuracy of HF radar surface current measurements: Intercomparisons with ship-based sensors. J Geophys Res **102**:18737-18748

Kaplan D, Largier J, Botsford L (2005) HF radar observation of surface circulation off Bodega Bay (Northern California, USA). J Geophys Res **110**:1-25

Kohut J, Glenn S (2003) Improving HF radar surface current measurements with measured antenna beam patterns. J Atmos Ocean Tech **20**:1303-1316

Laws K, Paduan J, Vesecky J (2010) Estimation and assessment of errors related to antenna pattern distortion in CODAR SeaSonde high-frequency radar ocean surface current measurements. J Atmos Ocean Tech **27**:1029-1043

Mariette V, Cochin V, Thomas N, Gurgel K-W (2006) Surveillance littorale operationnelle (SURLITOP) - Operational coastal monitoring. In: Radiowave Oceanography Workshop (ROW6), Hamburg, Germany. http://www.satlab.hawaii.edu/row/ follow login instructions: row2006, hamburg

Ohlmann C, White P, Sybrandy A, Niiler P (2005) GPS-cellular drifter technology for coastal ocean observing systems. J Atmos Ocean Tech **22**:1381-1388

Ohlmann C, White P, Washburn L, Terrill E, Emery B, Otero M (2007) Interpretation of coastal HF radar-derived surface currents with high-resolution drifter data. J Atmos Ocean Tech **24**:666-680

Paduan J, Rosenfeld L (1996) Remotely sensed surface currents in

Monterey Bay from shore-based HF radar (CODAR). J Geophys Res **101**: 20669-20686

Paduan J, Kim K, Cook M, Chavez F (2006) Calibration and validation of direction-finding high-frequency radar ocean surface current observations. IEEE J Oceanic Eng **31**:862-875

Pawlowicz R, Beardsley B, Lentz S (2002) Classical tidal harmonic analysis including error estimates in MATLAB using T_TIDE. Comput Geosci **28**:929-937

Teague C (1986) Multifrequency HF radar observations of currents and current shears. IEEE J Oceanic Eng **11**:258-269

Ullman D, O'Donnell J, Edwards C, Fake T, Morschauser D, Sprague M, Allen A, Krenzien B (2003) Use of Coastal Ocean Dynamics Applications Radar (CODAR) technology in US Coast Guard search and rescue planning. US Coast Guard Research and Development Center, Report No. CG-D-09-03