2011

Improvements to a Real-Time, Satellite-Derived Surface Current Product (OSCAR) and Evaluation in the Intra-American Sea

Mindy Jo Robinson
University of South Florida, mindyr@mail.usf.edu

Improvements to a Real-Time, Satellite-Derived Surface Current Product (OSCAR) and Evaluation in the Intra-American Sea

by

Mindy Jo Robinson

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science
Department of Physical Oceanography
College of Marine Science
University of South Florida

Major Professor: Gary T. Mitchum, Ph.D.
Robert H. Weisberg, Ph.D.
Mark E. Luther, Ph.D.

Date of Approval:
October 10, 2011

Keywords: gridded altimetry, linear predictive coding, geostrophic component velocity, Ekman component velocity, moored ADCP measurements, shipboard ADCP measurements

Table of Contents

List of Tables

List of Figures

Abstract

A method of creating real-time height fields using available gridded altimetry data combined with extrapolated values computed using linear predictive coding for use in the Ocean Surface Currents Analysis, Real-time (OSCAR) is developed and tested. This method is implemented by Earth and Space Research (ESR) to produce operational height fields that are in turn used to calculate the geostrophic component velocity of OSCAR. The gridded altimetry product used in OSCAR is evaluated against sea level measured by tide gauges while the operational total near-surface velocity from OSCAR (geostrophic plus Ekman component velocity) is evaluated using ocean velocities from moored and shipboard acoustic Doppler current profiler (ADCP) measurements. The evaluation is focused in the Intra-American Sea (IAS), having complex currents due its topography and bathymetry, and proves to be a challenging place for OSCAR to accurately reproduce ocean velocity. Issues are related to the upper limit in frequency that altimetry can capture, accuracy of altimetry near to land, the scale of the spatial smoothing across relatively narrow currents and geographical differences in the effective smoothing of the gridded altimetry used for OSCAR. These problems highlight the need for more and better ocean observing and modeling systems to provide data that can supplement the use of gridded satellite-derived products, especially in the IAS.

1 Introduction

1.1 Motivation

The development of a method to create the operational SSH product from OSCAR is for the expansion of the original OSCAR project that used a method for creating a surface current product using past altimeter data, validated in the tropical Pacific region. The OSCAR method has been evaluated and improved and is now applied globally. For the OSCAR global surface current product to be real-time, past SSH data must be used to develop an appropriate method of extrapolating future SSH, as no actual real-time satellite dataset is available due to the time necessary for data processing. After the extrapolation method is set and a real-time geostrophic surface current product is created, it must be evaluated against data sources to determine the validity of the method and utilized to show the product can yield scientific results.

The IAS is chosen for evaluation of the OSCAR product because it is both scientifically interesting and socioeconomically important to the Central and Southeastern United States. Altimeter-based SSH and OSCAR velocities have previously been evaluated in the open ocean, and the IAS provides a more stringent test given its complex topography and bathymetry. The IAS is the source of the Gulf Stream, an important mechanism in global salt and heat transport. The locations of the Loop Current and Loop Current Eddies are important to forecasting the intensity of tropical cyclones

that affect the Gulf of Mexico region and change on seasonal and inter-seasonal time scales. Variability in the circulation in the Gulf of Mexico greatly influences fisheries and the transport of pollutants (Mooers and Maul 1998). Inter-annual variations in the IAS can be associated with changes in atmospheric moisture content and can partially affect the amount of precipitation that falls in the Central U.S. (Mooers and Maul 1998).

1.2 Background

1.2.1 History of OSCAR

In-situ observations of ocean properties from ships and buoys are irregular and sparse in both time and space. The OSCAR project began with the desire to take advantage of the unprecedented temporal and spatial coverage provided by data from a satellite altimeter in order to better describe the state of the ocean at any given time and place. The tropical Pacific was chosen as a test location for the development of a way to estimate surface circulation using satellite-derived SSH and wind stress fields via geostrophic and Ekman relationships. The pilot study combines gridded SSH derived from Topex Poseidon (T/P) along-track data and Special Sensor Microwave/ Imager (SSM/I) and QuikScat winds. Geostrophic currents, inferred from SSH gradients, are found by combining the T/P height fields with climatological sea surface temperature (SST) and salinity, then calculated on a 1-degree square grid at 5-day intervals. The process uses the surrounding ~10 days of data. Updated surface current maps for the pilot study were provided every week with 8 to 14 days latency in time (i.e. the output was 8 to 14 days old upon release). (Johnson, et. al. 2007)

Proven valid, the OSCAR product can be used as a check against other models and observational datasets. The funding proposal for the second generation of OSCAR (K. Dohan et al., ESR) suggests that results may also be applied to many practical problems like determining optimal trajectories for ships and tracking hazardous materials adrift. Applications to fisheries and biological issues include larval transport and migration of marine animals. In shipping and transportation, optimal departure times and paths of travel can be determined based on the direction of current flow.

The calculation method for the original OSCAR currents is documented in Lagerloef, et. al. 1999 (hereafter referred to as LMLN) and includes a formulation for momentum balance and the datasets assimilated into the analysis for generation of the final product. The method aims to estimate geostrophic and Ekman components of tropical Pacific surface circulation representative of data obtained by 15-meter drogue drifters by applying simple dynamics to satellite-derived sea level and wind fields. The resulting currents are evaluated by comparing them to currents obtained by actual drifters and current moorings in the tropical Pacific. Geostrophic velocities are calculated using an f-plane at higher latitudes and a β-plane approximation within ~ 1 equatorial Rossby radius of the equator. The transition between these two regimes is approached using a Gaussian weight function, where the total geostrophic velocity is a weighted sum scaled with latitude of f- and β-plane estimates. Ekman velocities are estimated using a depth scale and drag coefficient for a purely wind driven response of the sea surface to wind stress.

Linear momentum balance representing average drifter motion in the surface layer with a thickness h and including u and v components of both geostrophic and Ekman velocities yields the complex vector form of the geostrophic balance equation:

$$if\mathbf{U}_g = -g\mathbf{Z} \qquad (1)$$

where f is the Coriolis parameter, $\mathbf{U}_g = u_g + iv_g$, the complex form of the geostrophic velocity found using the usual f plane calculation, g is gravity and $\mathbf{Z} = \partial\zeta/\partial x + i\partial\zeta/\partial y$, the complex form of the sea surface height gradient. Near the equator, the geostrophic velocity must be computed using the β plane approximation, $f = \beta y$. The derivative of (1) with respect to y gives:

$$\beta\mathbf{U}_b + \beta y(\partial/\partial y)\mathbf{U}_b = ig(\partial/\partial y)\mathbf{Z} \qquad (2)$$

where $\mathbf{U}_b$ is the geostrophic velocity found using the β plane approximation. The two different forms of the geostrophic velocity are used in a weighted sum to find the best estimate at each latitude. The total estimated velocity field is found by the sum of the geostrophic and Ekman components at each location on the grid.

Through comparison to observational data documented in LMLN, several limitations of this method for producing the OSCAR product are revealed. Statistical comparison of satellite-derived and in-situ drifter velocity components is done by calculating the standard deviation of the differences and the correlation coefficient, both functions of latitude. For current meter moorings, mean differences for standard deviation of the differences and correlations are computed at each mooring site. Looking at these results, the method gives good agreement for the zonally averaged meridional current (wind-driven) and the mean zonal current south of the equator (mostly

geostrophic). A noted bias in the Northern Hemisphere is due to underestimation of the magnitude of major zonal currents. Errors are smaller in higher latitudes and increase equatorward due to local accelerations acting on the drifters which are not included in the linear formulation for motion. Smaller-scale features like the Equatorial Undercurrent meandering and instability waves not resolved by the 1^{o} grid can also account for some of the inconsistencies. The comparison analysis shows this method of creating surface current maps inferred from altimeter-derived SSH fields can be used to monitor monthly circulation patterns. Further analysis is needed to look at features on shorter time scales.

1.2.2 Needed improvements

Bonjean and Lagerloef 2002 (hereafter referred to as BL) gives specific solutions to the issues found with the OSCAR test product and documents changes in the formulation for the second generation of the OSCAR product. The same data from LMLN is used to directly infer surface velocity, but with an improved model for characterizing SSH and motion. The BL model is validated using drifter records and current data from moorings just as in the original LMLN study. There are three surface variables for the horizontal (zonal *u* and meridional *v*) currents: SSH, wind stress and SST. The vertical distribution of velocity in the surface layer is characterized with two parameters: a depth scale *H* and a coefficient of vertical diffusion *A*. The formulation uses quasilinear steady flow in the surface layer where the horizontal velocity in complex form $\boldsymbol{U}(x,y,z,t) = (u + iv)$ is proportional to the pressure gradient plus vertical shear and mixing:

$$if\boldsymbol{U} = -1/\rho_m \nabla p + A\boldsymbol{U}'_z \qquad (3)$$

where $\nabla \equiv \partial/\partial x + i\partial/\partial y$, $\boldsymbol{U}'_z$ is the vertical shear, ρ_m is the characteristic density and A is the parameter characterizing eddy viscosity. From this equation, BL derives the relation for the velocity averaged between the ocean-atmosphere interface and the velocity at some arbitrary depth:

$$if\,\bar{\boldsymbol{U}} \equiv (if/h)\int_{-h}^{0} \boldsymbol{U}(z)\,dz = -g\nabla\zeta + (h/2)\nabla\theta + (\tau - A\boldsymbol{U}'(-h))/h \qquad (4)$$

where ζ is the displacement of the ocean-atmosphere interface, h is an arbitrary depth below the interface, θ is buoyancy, $\tau = \tau^x + i\tau^y$ is the surface wind stress divided by the characteristic density and $\boldsymbol{U}'(-h)$ is the bottom shear. This relationship includes pressure gradient force, the buoyancy gradient and the net drag force from vertical diffusion through a layer of arbitrary thickness. Net drag force depends on wind stress and buoyancy gradient. This differs from LMLN, where the buoyancy gradient was taken to be = 0 and the drag coefficient constant. Having $\boldsymbol{U}'$ depend on depth means there is a higher-order turbulence closure parameterization in this formulation than in LMLN and it accounts for much of the Northern Hemisphere bias found in the equatorial region for the LMLN comparison to in-situ currents. The layer velocity $\bar{\boldsymbol{U}}$ (an average of sorts between surface velocity and velocity at depth) is defined as:

$$\bar{\boldsymbol{U}} = \boldsymbol{U}_o + (1/h)\int_{-h}^{0}\int_{0}^{z} \boldsymbol{U}'(Z)\,dZ\,dz. \qquad (5)$$

This relationship is better than equation (4) in practice because it is defined everywhere, including at the equator where y = 0. Combining equation (4) and an expression for $\boldsymbol{U}'$, the surface velocity $\boldsymbol{U}_o$ is:

$$if\,\boldsymbol{U}_o = -g\nabla\zeta + (1/H)\boldsymbol{q}(H/\hbar_e)\tau + ((H/2)/\boldsymbol{q}(H/2\hbar_e))\nabla\theta \qquad (6)$$

where $\boldsymbol{q}(\xi) = h_e/\tanh(h_e) = 1 + h_e^2/3 - h_e^4/45 + \ldots$, H is the depth scale and $h_e \equiv$

$(A/if)^{1/2}$. This shows that $\boldsymbol{U_o}$ is made up of a sum of three terms related to the gradient of ocean-atmosphere interface displacement, wind stress and the buoyancy force gradient. For the case of the equatorial limit, making $f = 0$ in equation (6) and minimizing the resulting equation determines the depth scale H. Because of momentum terms not included here and measurement uncertainty in the observed currents, a singularity can still occur at the equator in equation (6). This remains a problem for this type of calculation for derived surface currents. The surface velocity $\boldsymbol{U}_o$ also must have some special consideration near the equator. LMLN chose to deal with this problem by applying a weight function and a regression fitting technique. BL estimates the velocity near the equator using a linear combination of orthogonal polynomials which satisfies equation (6). In the transition region between near-equator and regular latitudes, a linear weight function is used in computing the velocity. This formulation better reproduces equatorial features not achieved by the LMLN method. Overall, the BL method provides corrections for bias and discrepancies found by LMLN in the estimated versus the observed surface currents in the tropical Pacific, largely due to the addition of vertical shear and buoyancy terms along with an improved procedure for characterizing surface velocity near the equator. These changes are applied in the operational real-time product to which part of this project contributes the daily SSH fields.

Other general improvements made in the second generation OSCAR product include expanded spatial coverage from only the tropical Pacific to global—all major ocean basins and mid and high latitudes—as well as bringing the product into real-time. The generation of real-time OSCAR requires that two altimeter-based SSH products (the

Archiving, Validation and Interpretation of Satellite Oceanographic data (AVISO) product and the Naval Research Laboratory (NRL) product, both discussed below in detail) be combined and that a forecast method be developed.

1.2.3 SSH Products from Altimeter Data

AVISO and NRL SSH products are combined to create the heights product developed as part of this project. The AVISO altimeter products, with support from Centre National d'Études Spatiales (CNES), are produced by the Data Unification Altimeter Combination System (DUACS) as part of the multi-mission ground segment (SSALTO). All the following information is taken from the SSALTO/DUACS Handbook (2007). Satellite missions included in generating the data products are Jason-1, Topex/Poseidon (T/P), ENVISAT, Geodetic Satellite (GEOSAT) Follow-On (GFO), Earth Remote Sensing (ERS) 1 and 2 and GEOSAT. Output products available are delayed mode (DM) (also referred to in documentation as delayed time) and near real-time (NRT). The grid used is a global 1/3°, so-called "MERCATOR" projection, $M_{I,J} = (X,Y)$. The formulation for generating the AVISO grid is:

$X = X_{min} + \Delta X * I$, where I goes from 0 to I_{count}.

$Y = 180/\pi * \arcsin [\tanh (\Delta Y_{rad} (J + J_{eq}))]$, where J goes from 0 to J_{count} and

$J_{eq} = ½ * \ln(1 + \sin (Y_{rad}) / 1 – \sin (Y_{rad})) / \Delta Y_{rad}$.

X_{min} and Y_{min} are specified in the product files in degrees.

$Y_{rad} = Y_{min}$ in radians.

ΔX, ΔY = X and Y step specified in the product files in degrees.

$\Delta Y_{rad} = \Delta Y$ in radians.

I_{count} = the number of values for X (the longitude dimension of the file).

J_{count} = the number of values for Y (the latitude dimension of the file).

For the AVISO product files being used in this project: $X_{min} = 0^o$, $Y_{min} = -82^o$, ΔX and $\Delta Y = 1/3^o$, $I_{count} = 1080$ and $J_{count} = 915$. These values yield a global grid running from -82^o to (approximately) 82^o in latitude.

Both the AVISO DM and NRT products are available as gridded maps of absolute dynamic topography. This representation of the sea surface requires the addition of a mean field—the mean dynamic topography (MDT)—to the sea level anomaly computed from altimetry. The MDT is the part of the sea surface height due to permanent currents, specifically, the mean sea surface height minus the geoid. The sea level anomaly is referenced relative to a 7-year mean profile (1993 – 1999) to get the absolute measurements of the ocean dynamic topography. The mean dynamic topography must be consistent with the physical content of the altimeter data and so must correspond to the mean of the ocean circulation including geostrophic, barotropic and baroclinic circulation. The MDT is computed using a two-tiered method. First, an estimate of the large-scale circulation is computed by averaging the altimeter heights over the 7-year period, removing the geoid based on a model using Gravity Recovery and Climate Experiment (GRACE) data. Second, heights are computed using temperature and salinity profiles from in-situ measurements—Argo floats and conductivity, temperature and depth (CTD) sensor data. Global geostrophic velocities are constructed from drifter data. A multivariate objective analysis is then used to combine the height and velocity estimates

to improve the first guess MDT in the initial calculation. The MDT is validated against other researchers' estimates of the MDT.

The AVISO DM maps are provided on the global MERCATOR grid on a weekly basis with a substantial (on the order of months) delay from present due to processing. This merged-mission dataset is available for dates from October 1992 to the most recent release (ongoing). Missions included in some or all parts of the processing are T/P, Jason-1, ERS1/2, GFO and ENVISAT. The DM is more precise than the NRT product because of the intrinsic better quality of the precise orbit ephemeris used by the Geophysical Data Record (GDR) in processing the altimeter data and because the product is computed on an optimal time window. The height field being computed is centered with 6 weeks of data before and after it. All available along-track data from up to 4 satellites at any given time is used as input for an optimal interpolation (OI) method including reducing long-wavelength errors to generate height data. This inter-calibrated, highly accurate, long time series makes the DM product very useful in determining a long-term mean height field.

The AVISO NRT maps are on the global MERCATOR grid and released daily with a 7 to 10 day latency in time, available for August 2001 to present. This product is generated by merging the same altimeter data as for the DM product but with only a few days of delay between acquisition of the data and release of the mapped product. Due to this near real-time status, the data window is not centered on the height field being computed, and only the previous 6 weeks of data are used in processing for the height maps. To compensate for the data window not being optimal, some improved corrections

are used in the processing such as an ocean tide model to correct the altimeter data, an improved orbit error reduction method and an improved mapping method.

The initial motivation for obtaining the NRL height product was that the AVISO NRT daily product, though only having a 7 to 10 day delay due to processing, was kept proprietary by AVISO for approximately 6 weeks. Eventually, AVISO began releasing NRT product without restriction, shortening the length of the time series required from NRL. According to Jacobs, et. al. 2002, NRL maps of SSH anomaly (SSHA) come from the Naval Oceanographic Office at Stennis Space Center in Mississippi, with the product developed for use in Navy operations for fleet support and as a contribution to global ocean modeling done by the Navy and others. The SSHA maps are provided daily on a global $1/8^{o}$ grid from -80^{o} to 80^{o} in latitude. Missions merged to create the product are T/P, ERS2, GFO, Jason-1 and ENVISAT. The height data is obtained using the same basic methods for processing altimeter data as most other altimeter processing, deriving SSH from the GDR. Along-track SSHA is used in an OI procedure to create a mesoscale product rapidly enough for daily release of maps. The OI procedure is designed to reproduce features with the length scales of eddies (20 km to 100 km) and includes in-situ observations of eddy characteristics to make the interpolation on those scales as accurate as possible. The previous 60 days of altimeter data is used in generating the final daily SSHA product. Note that no time delay between the release date of the SSHA maps and present is mentioned in documentation by Jacobs, et. al. 2002, and has been determined in work associated with this project to be 5 days (see Section 2.1.2).

1.2.4 The Linear Prediction Method

As discussed in Section 1.2.3, there is an unavoidable time delay between the most recently available altimeter data and real-time. To extrapolate across this gap, the linear prediction method (*Numerical Recipes*, 1st Edition, pp. 444-446), a very high-order autoregressive (AR) model, is used. Linear prediction can be especially useful for extrapolating smooth, oscillating (not necessarily periodic), modulated signals. Many cycles of the signal can be accurately produced. The details of this concept are described in this section.

First consider a set of measured data values that are from some underlying set of true values. The best estimate of the true value at any point is a linear combination of the known, measured values (with noise) and some discrepancy between the measured and true values. There exists a set of coefficients that minimize the discrepancies between the measured and the true values and they are determined by the covariances. Covariances must be estimated carefully, as the linear prediction results are very sensitive to their estimation. The recommended method of estimating the covariances is the use of the Burg algorithm such that each coefficient is re-estimated based on the previous one in order to minimize the residual at each point. If the data points are equally spaced then a set of some number of consecutive data values can be used to predict the next value(s). Stationarity is assumed such that the autocorrelation only depends on the difference between points and not on the points individually. Noise is assumed to be uncorrelated between consecutive points.

The linear prediction must be stable, meaning the characteristic polynomial for determining the linear prediction coefficients must have all of its roots inside the unit circle. This is not necessarily expected for a finite-length dataset of randomly varying amplitudes. The preferred computational "trick" is to reflect any particular root lying outside the unit circle onto or inside the unit circle, then recompute the modified coefficient for that point. Reflection is the preferred method because it preserves the amplitude of a series driven by a sinusoidal set of residuals. Inherent bias in the linear prediction should be removed in practice. This is done by subtracting, on a point-by-point basis, a mean that is weighted based on the autocorrelation values. The linear prediction is then performed and the values are then added back onto the results.

Linear predictive coding (also described in *Numerical Recipes*) is a set of computing methods related to the linear prediction method that can be used to speed up the calculation of the necessary linear prediction quantities and was applied for the extrapolation calculation. The signal, due to an inherent amount of redundancy in it, can be "compressed" for computational purposes. Any particular estimated value of the signal is predictable from previous values (having a minimal residual discrepancy) from relatively few linear prediction coefficients. The number of coefficients is determined determined by experimentation, finding how many previous data points must be used to minimize the residual discrepancy of each data point following. In computing, only the number of linear coefficients M, their M values and the first M data points are kept while determining the residual discrepancy for each desired data point. The value of M is the

model order of the AR calculation. The specifics of the application of this method in this project are discussed in Section 2.4.1.

1.2.5 Characteristics of the Intra-American Sea

The Intra-American Sea (IAS) is an encompassing term used to identify the region in the Western Hemisphere that includes the Caribbean Sea, the Gulf of Mexico and the portion of the Atlantic Ocean off the coast of the southeastern United States that is home to the beginning of the Gulf Stream. The IAS is vital to many industries including oil and gas, shipping and trade, commercial cruise lines, commercial and recreational fishing and tourism. All of these high-value industries depend on accurate representation of the surface circulation, both in a long-term mean sense and for daily operations. It is important that near-surface circulation models such as OSCAR are capable of accurately representing the circulation.

Surface waters from the Atlantic flow into the Caribbean Sea through passages between the Greater and Lesser Antilles islands that form the eastern and northern boundaries of the Caribbean Sea (Richardson 2005). In a mean sense, this inflow forms the main axis of the Caribbean Current that flows westward until it encounters a change in bathymetry off the coast of Nicaragua. This shallowing forces a turn to the north and back to the west again. The Caribbean Current then forms the northward-flowing western boundary current known as the Yucatan Current, exiting the Caribbean Sea and entering the Gulf of Mexico. In the Gulf of Mexico, the Loop Current flows to the north, u-turns back to the south and then flows eastward through the Straits of Florida, becoming the Florida Current. The Florida Current follows the east coast of peninsular

Florida and travels to the north to become the Gulf Stream. This general description of the mean flow has been determined by many authors (e.g., Pratt and Maul 2000, Johns et al. 2001, Centurioni and Niiler 2003, Richardson 2005 and Alvera-Azcarate et al. 2009) using both direct measurements from instruments on buoys, moorings, drifters and ships and derived quantities like geostrophic velocities implied from satellite altimetry to characterize the flow of surface waters in the IAS.

There are major permanent features of the circulation in the IAS created by its complex topography and bathymetry. One is the Panama-Columbia Gyre, a large cyclonic circulation located off the east coast of Panama and the north coast of Columbia. The steering of the mean flow to the north by the Nicaraguan Rise and the intensification of the the flow creating the western boundary current along the Mexican coast (referred to as the Yucatan Current) are a result of variations in topography and bathymetry, as well is the mean behavior of the Loop Current in the Gulf of Mexico and the Florida Current (Alvera-Azcarate et al. 2009).

Variability in the mean flow of the IAS has also been studied by many authors, including those mentioned above. Most of that variability is the result of the propagation of eddies and meanders (small propagating features that do not have a closed circulation). The passage of these eddies and meanders occurs on different timescales. Alvera-Azcarate et al. 2009 found variability on annual, intra-annual and interannual timescales. Because OSCAR is largely based on geostrophic currents derived from altimetry (as studied by Alvera-Azcarate et al. 2009), these timescales of variability should also be identifiable. An interesting question if is OSCAR shows any major differences from a

geostrophic-only analysis, as it includes the Ekman component of near-surface circulation —that which is caused by winds. This may be especially important in the Caribbean as the Trade Winds blow steadily through the area. Because eddies and meanders are superimposed on the mean flow, many aspects of both permanent and interloping circulation features may be influenced by the addition of the wind-driven component of the circulation. A real-time model such as OSCAR has the potential to be particularly advantageous in this respect as instruments like buoys that directly measure the total current can be blown ashore or not capable of entering certain regional features of the circulation in the IAS due to the prevailing winds (Richardson 2005).

OSCAR should be expected to be able to distinguish between areas that are dominated by mean flow and which are dominated by variability, as many studies using both direct and derived observations have been able to determine (e.g., Centurioni and Niller 2003, Richardson 2005 and Alvera-Azcarate et al. 2009). Areas observed to be dominated by the mean flow include the southeastern Caribbean near the inflows through the Lesser Antilles, off the east coast of Nicaragua, through the Yucatan, the Florida Current and a relatively slower (compared to the aforementioned flows) current off of the southern coast of Hispaniola that flows westward and joins the main axis of the Caribbean Current (Centurioni and Niller 2003 and Richardson 2005). Areas with slower observed mean velocities include the northeastern Caribbean, the area just south of Cuba (Centurioni and Niller 2003 and Richardson 2005) and the western Gulf of Mexico (Alvera-Azcarate et al. 2009). The northernmost extent of the Loop Current is also

dominated by variability—the northernmost position of the Loop Current and the episodic occurrence of eddy-shedding events (Alvera-Azcarate et al. 2009).

2 Extending OSCAR to a Real-Time, Global Product

2.1 Chapter Overview

The OSCAR product is a model of near-surface currents at 15-meters depth. It has three major components as detailed in Bonjean and Lagerloef 2002: geostrophic, Ekman and buoyancy (also called the thermal or shear term). The geostrophic component can be determined by computing currents from height gradients where the heights are measured by satellite altimetry. USF provides the global, real-time height fields to OSCAR. Because the second generation of OSCAR is a global product, gridded height fields from optimally interpolated along-track altimetry from several different missions are used to determine the geostrophic currents. Available, well-known, well-documented sources include the AVISO and NRL products. Each product is evaluated and characterized based on its statistical properties and the products are combined into one time series on the same datum, resulting in a series that is 5 to 6 days short of real-time.

The remaining time gap caused by the delay from data acquisition to product release for gridded altimetry requires the use of a predictive model. The linear prediction method, which uses patterns in past data to predict future data, with a model order of 100 is chosen. Evaluation shows it outperforms persistence through the 10 days of forecasted heights needed for the USF contribution to OSCAR. The operational product as part of OSCAR runs daily at ESR. It automatically obtains the newest AVISO and NRL products, adjusts and combines them, applies the linear prediction method, makes a time

series of the global, real-time maps and makes them available for incorporation into the rest of the OSCAR system.

2.2 Preliminary Assessment

The AVISO NRT daily product has at least 7-10 day latency in time due to processing. In order to create a real-time SSH product, another data set must be used to fill in this gap from the last available AVISO daily height map to today. Daily data from the NRL has a 5 day gap between the most recently released daily height map and today. An extrapolation method must be used to bring the dataset to real-time. NRL is on a different grid than AVISO, so both the existing files and the extrapolated days of SSH must be interpolated to the correct grid.

2.2.1 Compatibility of AVISO Delayed Mode and AVISO Near Real-Time

Upon first inspection of the NRL and NRT products, it is obvious that the mean field has been removed from NRL but not from NRT, so computing a mean field is necessary. The AVISO NRT daily dataset does not go back far enough in time to create a long-term mean SSH field; however, there is the AVISO DM that goes back to 1992. If the AVISO DM and NRT are comparable datasets, the long-term mean field can be computed using the DM data. To determine if these two datasets are compatible through time, a global area weighted lagged covariance calculation is performed. Global area weighting is used because calculations using altimeter data become increasingly difficult at higher latitudes, and it is expected that these areas to be far less accurate reproductions of SSH when compared to lower latitudes. The weighting used is

$$w=d(lat)*\cos(lat)$$

where lat represents the latitudes in the AVISO grid, d is the distance in the y direction

(North-South) assigned to each grid point and cos(lat) is the distance in the x direction (East-West) assigned to each grid point so that each point has a weighted area associated with it over the entire grid. Note that this type of weighting assumes a perfect sphere for the shape of the Earth. Figure 2-1 shows that the maximum weighted lagged covariance occurs at a lag value close to 0 (to within noise). The maximum at approximately zero-lag indicates the dates associated with the height data files are consistent.

To determine if the DM and NRT datasets are compatible with respect to the height data, the ratio of their variances is taken over the common time period where both DM and NRT data are present. The distribution of the ratio of the variances given in Figure 2-2 has a maximum at 1, indicating that the variances over the common time period are comparable and represent the same data. Because it has been determined that the DM and NRT are compatible, the DM can now be used to obtain a long-term mean field with the dataset dating back to 1992—representative of past NRT data, as if it existed.

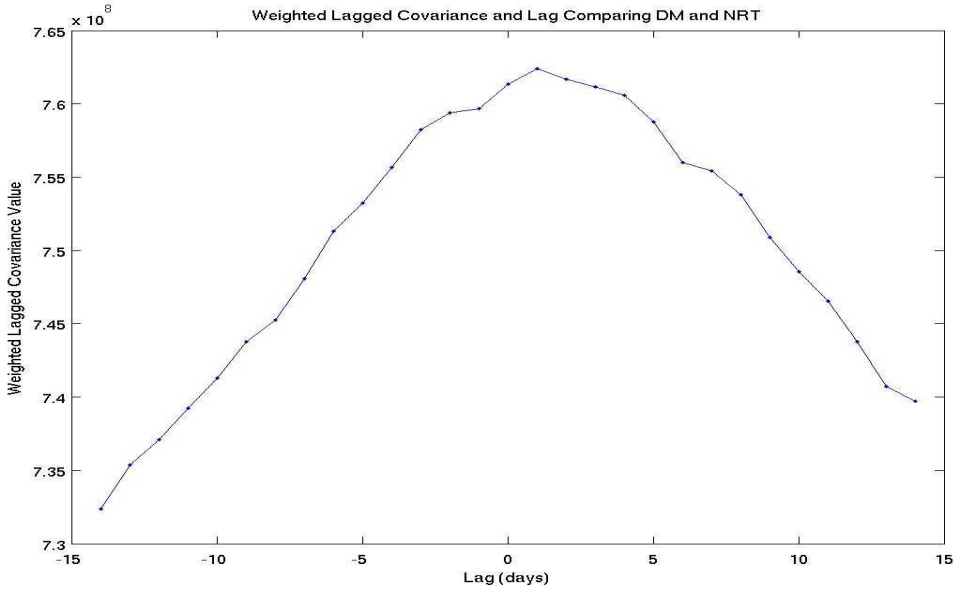


Figure 2-1. Weighted lagged covariance value vs. lag value (in days) to determine if AVISO delayed mode (DM) and near real-time (NRT) have compatible file dates. Maximum value at lag = 0 indicates the dates associated with the height files match.

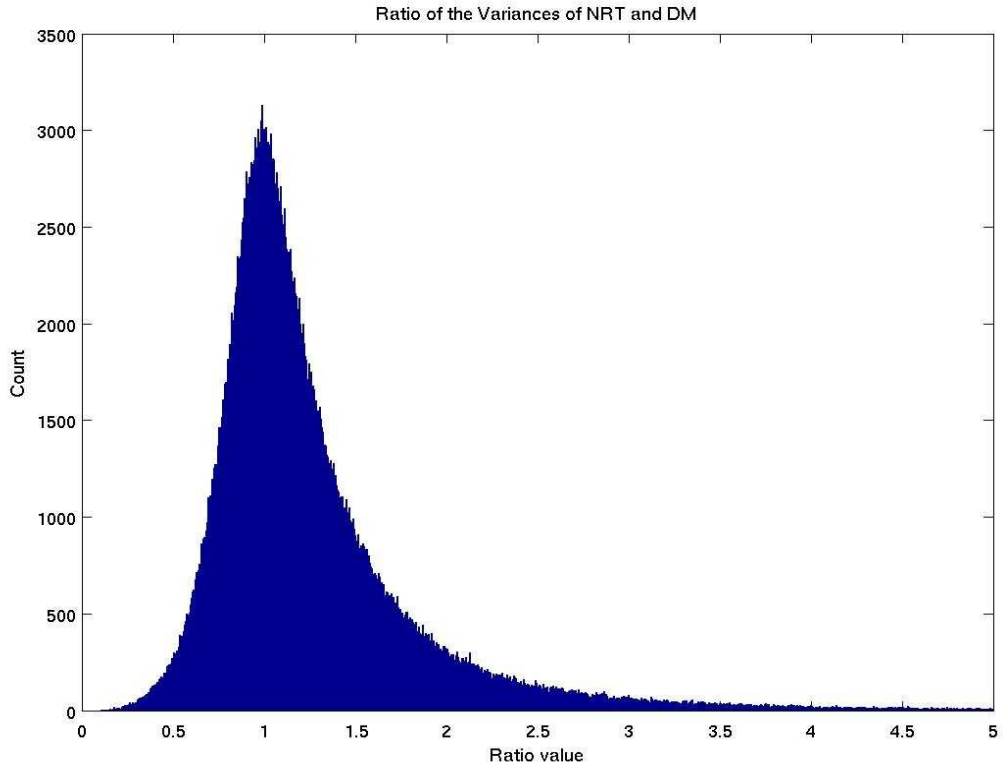


Figure 2-2. Distribution for the ratio of the variances for AVISO DM and NRT. Maximum at 1 indicates that the variability in the datasets is comparable and that the DM and NRT represent the same height data.

2.2.2 Determining the Temporal Lag between NRL Daily Data and NRT Daily Data

The first step in using the NRL daily data to fill in the gap to today left by the NRT daily data is to determine what temporal lag exists between them. Their file names are not necessarily the exact day to which the height data corresponds. The NRT data is used as the "actual" reference date and a weighted lagged covariance calculation is done using the same global area weighting described in the DM to NRT comparison. The maximum weighted covariance occurs at a lag value = -5 days (as shown in Figure 2-3), indicating the date assigned to the NRL file name leads the NRT date by 5 days. The optimal date is determined by shifting the NRL files back 5 days so that they correspond to the correct NRT date. For example, an NRT date of 01 December 2006 and an NRL date of 06 December 2006 are actually the same day's height data.

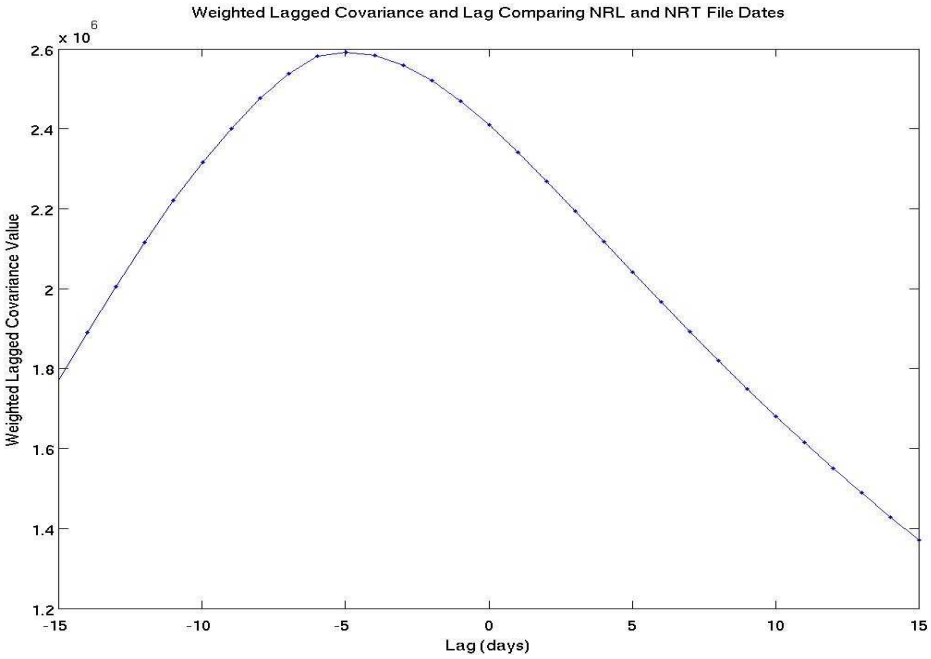


Figure 2-3. Weighted lagged covariance value vs. lag value (in days) to determine if the NRL and NRT have compatible file dates. Maximum value at -5 days lag indicates the file name associated with the NRL files leads the date on the NRT files by 5 days.

2.3 Determining Adjustments Needed to Make NRL Compatible with NRT

The next step in using NRL to fill the gap left by NRT is to make NRL look like NRT. With only a preliminary inspection, it is obvious that they are not alike. First, they are on different grids with different latitude ranges and different latitude and longitude spacing. A "nearest neighbor" interpolation is used to put NRL (with one-eighth degree grid spacing in latitude and longitude and a latitude range of -80º to 80º) onto the same grid as AVISO (with a one-third degree grid spacing in longitude and an unevenly spaced latitude grid that ranges approximately from -82º to 82º). Obviously, the highest 2 degrees of latitude North and South will be missing, but these areas are almost entirely devoid of data points due to land or permanent ice cover. Second, it appears that NRT still includes a mean field, while NRL does not. Because it has already been determined that AVISO DM and NRT are comparable datasets, the long-term DM time mean is

removed from each NRT height grid. The resulting grid is then differenced with the mean NRL interpolated to the NRT grid to try to establish a common datum. The long-term DM time mean is calculated over the date range used by NRL to remove its long-term time mean: 28 October 1992 to 21 Aug 2002. This difference (not shown) is normally distributed with a central shift of approximately -25 mm (where NRT is the larger quantity). It is advisable to check for any zonal or meridional variation in this shift by splitting the data into quarters of both latitude and longitude and looking at the distributions. Figure 2-4 shows that for all latitudes and longitudes, the distributions are not radically different. There is some spatial structure in the shift, but they are similar enough to use a global mean offset value of 25 mm. NRL is transferred to the same datum as NRT by interpolating to the NRT grid and adding the long-term DM mean field plus the 25-mm adjustment to the NRL product. This final converted grid is hereafter referred to as "NRL→NRT".

Now that the NRL and NRT products are on the same datum, the amount of smoothing for each product must be considered. This comparison is made using the datasets' standard deviations. Figure 2-5 shows the values of standard deviation for both NRL→NRT and NRT on the same scatter plot. Noted in red are the values where the standard deviation of NRT greatly exceeds that of NRL. A map of these points (Figure 2-6) shows the areas in question include the Bay of Fundy, the area of the Amazon River, the coast of the U.K., the southern coast of Australia, the Sea of Japan and others where tidal range can be very different compared to tidal ranges over the rest of the globe. This suggests that there is some inconsistency between the tidal models or corrections being used by NRL and NRT.

From the distributions of the standard deviations (Figure 2-7a), it looks like there might be more high-energy information in the NRT product, having a larger spread than the standard deviation of the NRL→NRT product. Looking at global maps of the standard deviations in Figure 2-7b, it is apparent that the NRL is more smooth than the NRT, with notably more fine-scale detail present in the NRT map. Since there is no way to "de-smooth" the NRL product, the transformation of NRL is complete. Interpolating to the NRT grid and adding the long-term mean field plus a 25-mm global offset is the "best guess" for making NRL look like NRT.

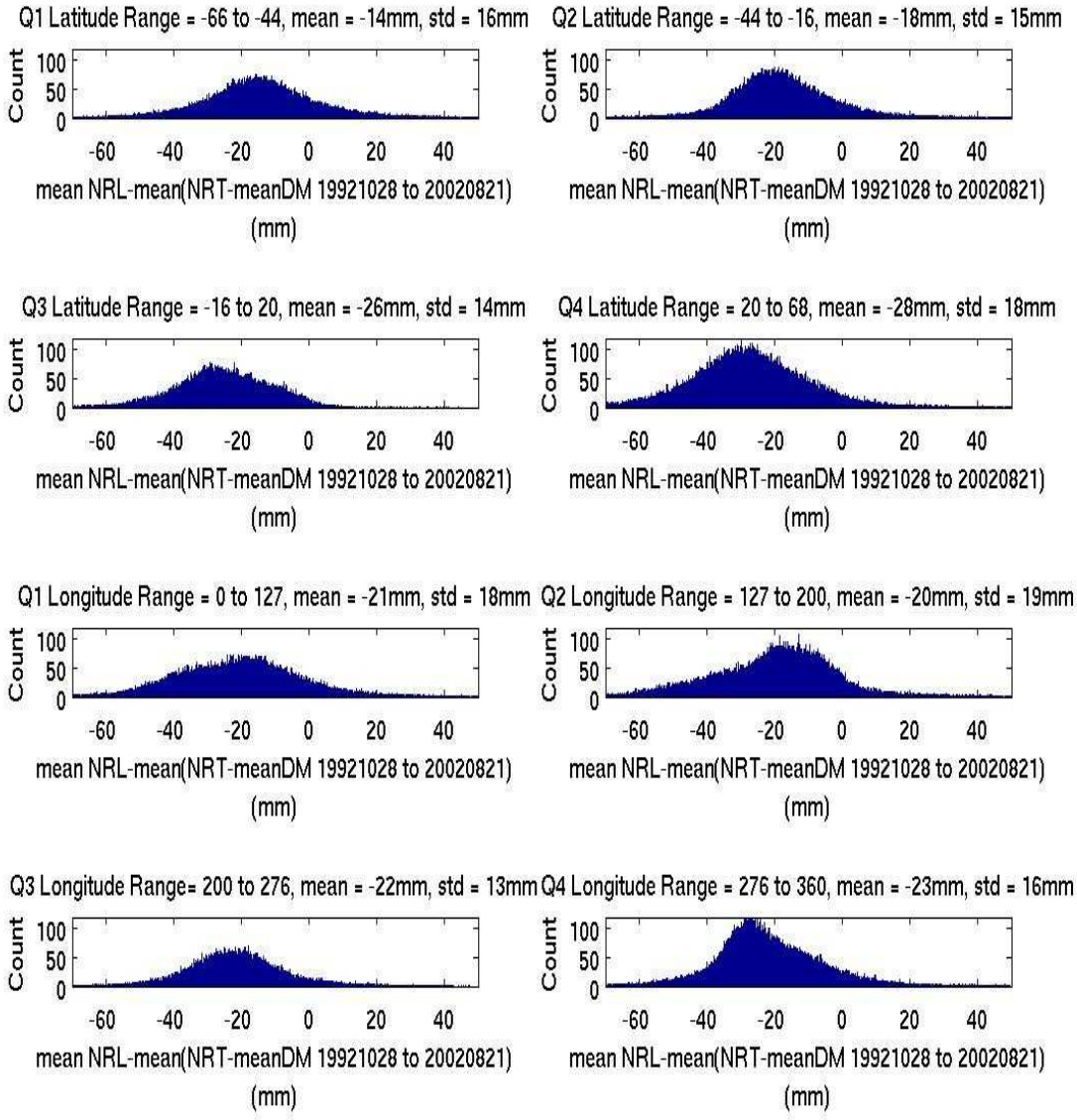


Figure 2-4. Quarters of latitude and longitude for the distribution of mean NRL – mean(NRT – mean DM 19921028 to 20020821). Used to determine if there is any major spatial variation in the global mean offset. No significant spatial change in the central value is apparent.

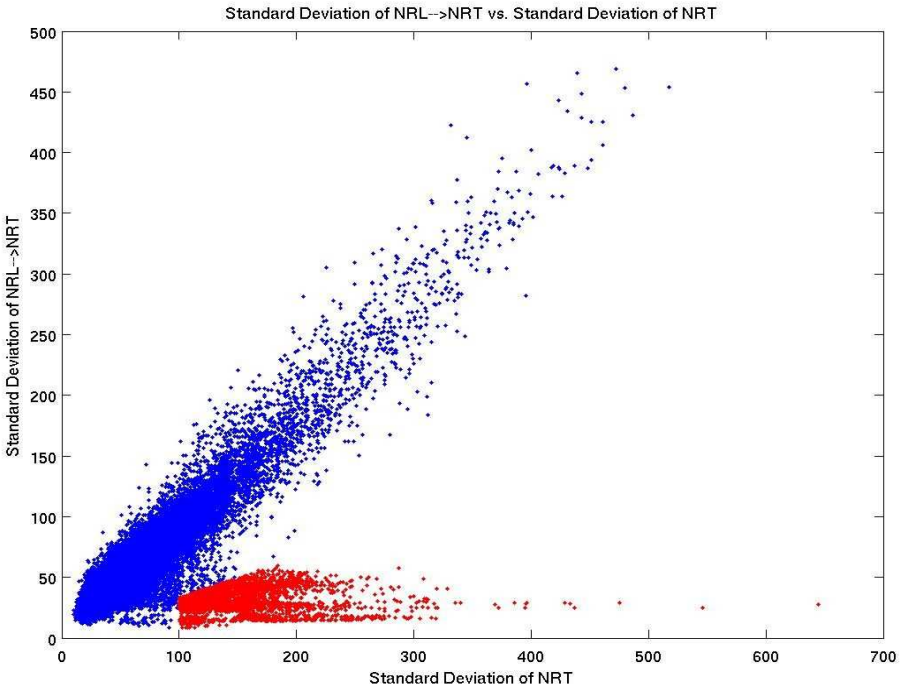


Figure 2-5. Standard deviation NRL→NRT vs. standard deviation NRT. Note in red where NRT is much larger than NRL→NRT.

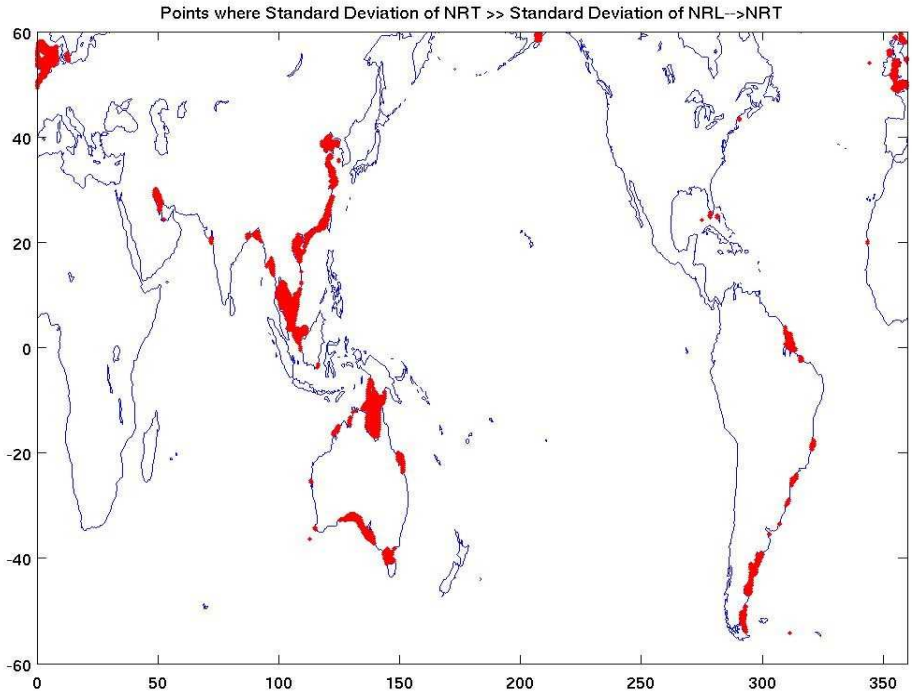


Figure 2-6. A map of points where the standard deviation of NRT is much larger than NRL→NRT, Occurs in regions where tidal ranges are very different from a typical tidal range (in a global sense).

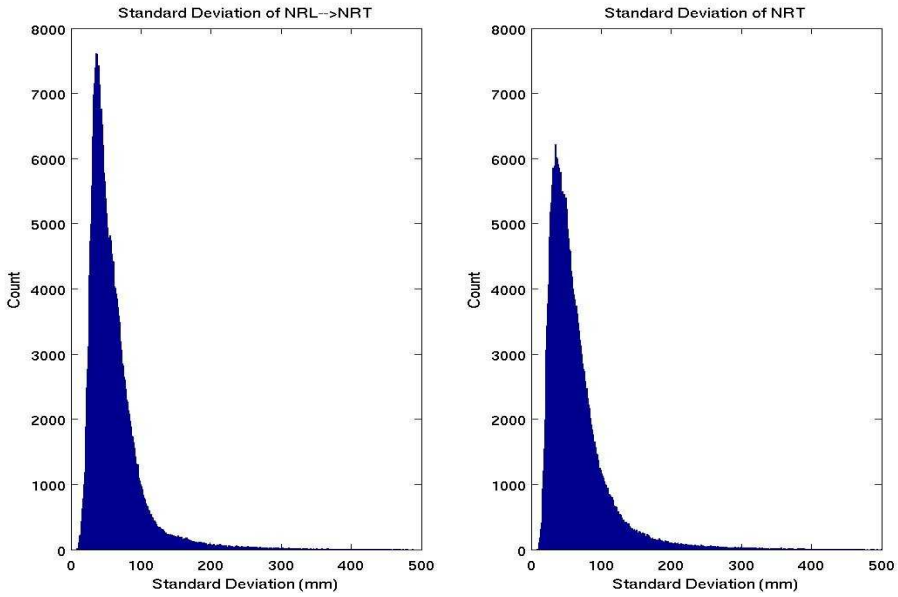


Figure 2-7a. Distributions of standard deviations of NRL→NRT and NRT forced onto the same scale. The larger spread of NRT suggests it may contain more high-energy information than NRL→NRT.

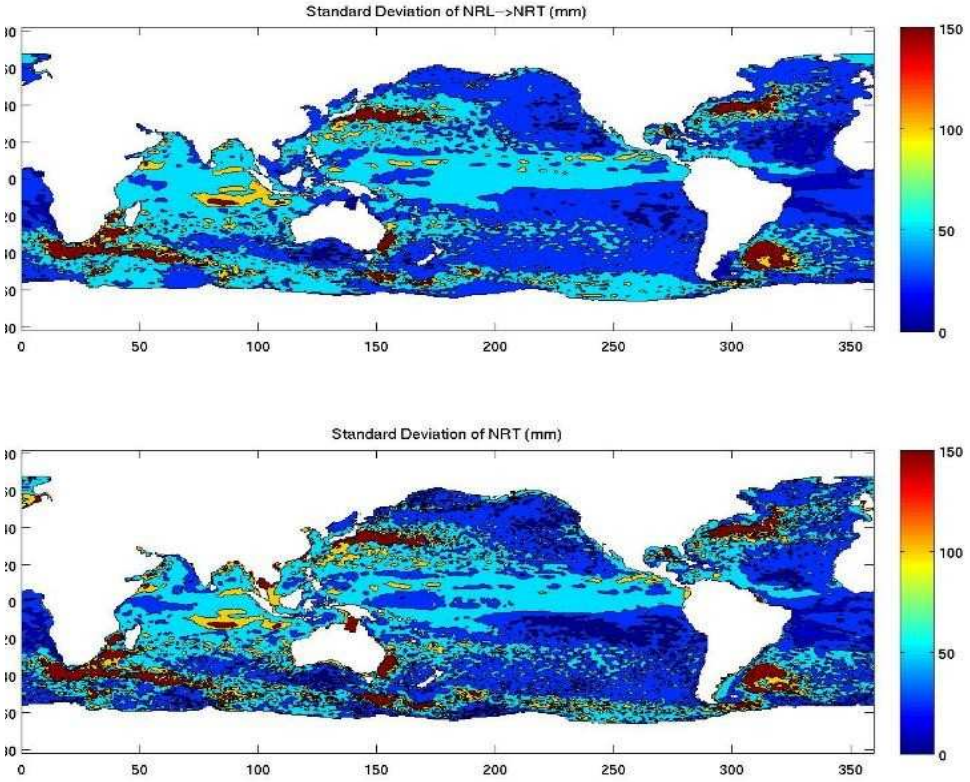


Figure 2-7b. Maps of standard deviations of NRL and NRT. Note that NRL→NRT is obviously smoother than NRT.

Another way to characterize ways in which the NRL→NRT and NRT products are different is by exploring the variation in the differences between NRL→NRT and actual NRT over time. Figure 2-8 shows where the standard deviation of the differences between NRL→NRT and NRT are large. Red points are locations where the standard deviation of the differences is greater than 50 mm, and blue points have standard deviation greater than 100 mm. The areas > 50 mm include high-energy areas where, due to the difference in smoothing, the higher wave number features are smoothed out in the NRL product. The areas > 100 mm are in coastal areas where there are likely to be tidal model and/or shallow water data discrepancies between the NRL and NRT products. These would be implicit features of the datasets and cannot be helped, but are documented here for reference.

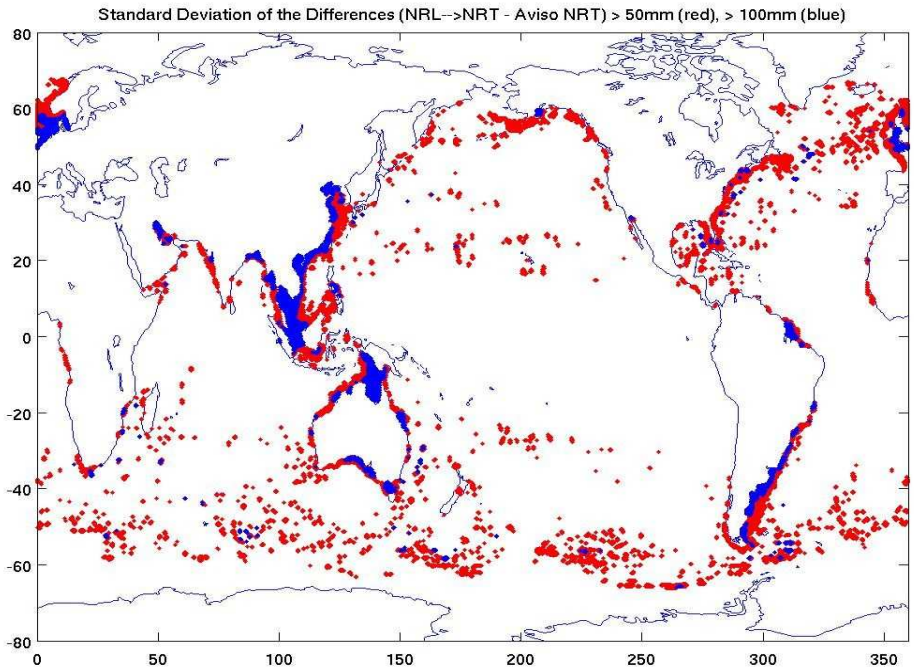


Figure 2-8. Large values of the standard deviation of the differences between NRL→NRT and actual NRT. Standard deviations >50mm mapped in red (includes high-energy areas) and >100mm in blue (coastal areas). Both could be explained by implicit discrepancies in producing the raw NRL and NRT products.

Another comparison used to explore similarities and differences between the two products is to compute the spatial correlation for each time pair of NRL → NRT and NRT using the same global area weighting described in Section 2.1.1. Figure 2-9a shows the height maps for a particular day with the time means included and Figure 2-9b shows the same pair of maps for the same day, but with the time means removed. Since high correlation is to be expected between the two with the means included, the time means are removed for the global area weighted spatial correlation calculation. Figure 2-10 shows the time series of the spatial correlation with values around 0.7, fairly high for this sort of data.

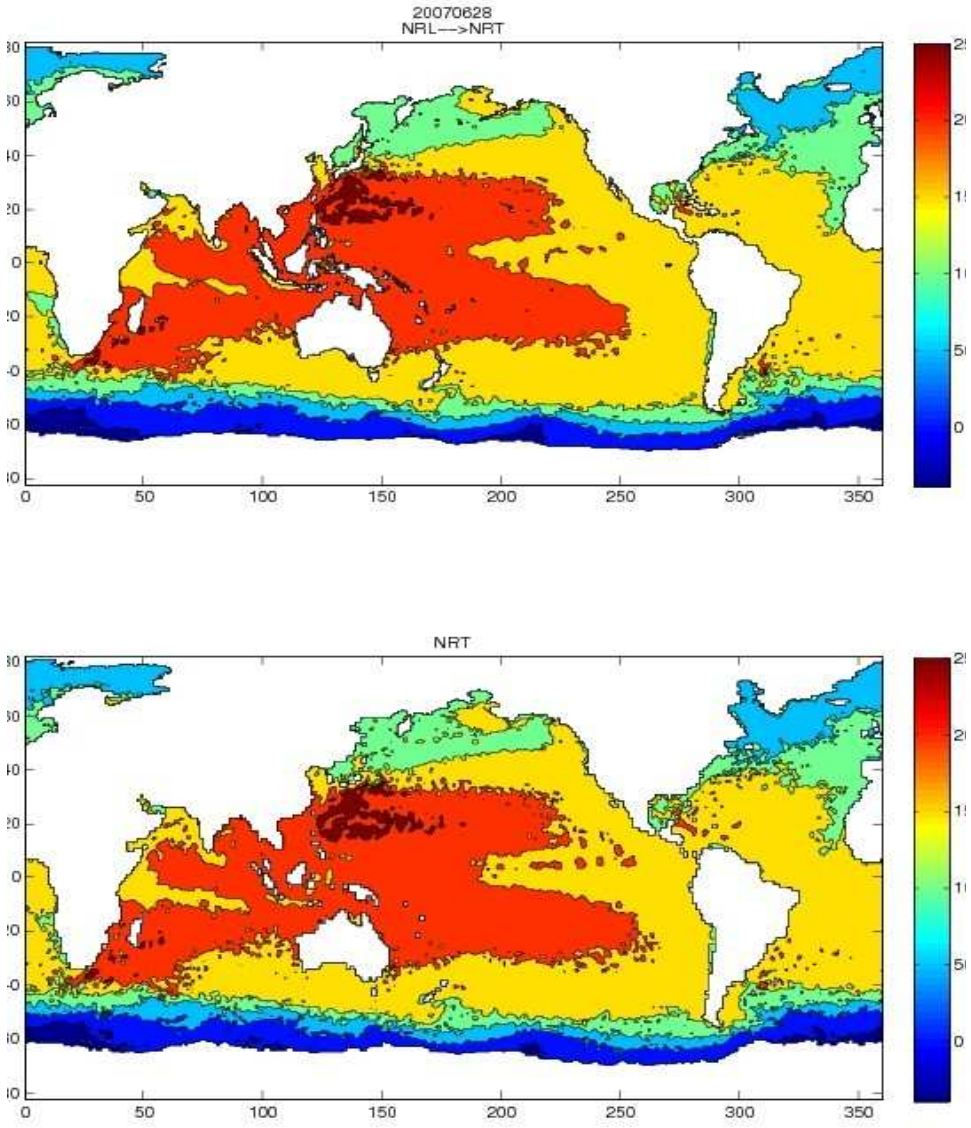


Figure 2-9a. Maps of NRL→NRT and NRT sea surface heights on a particular day with the time means included.

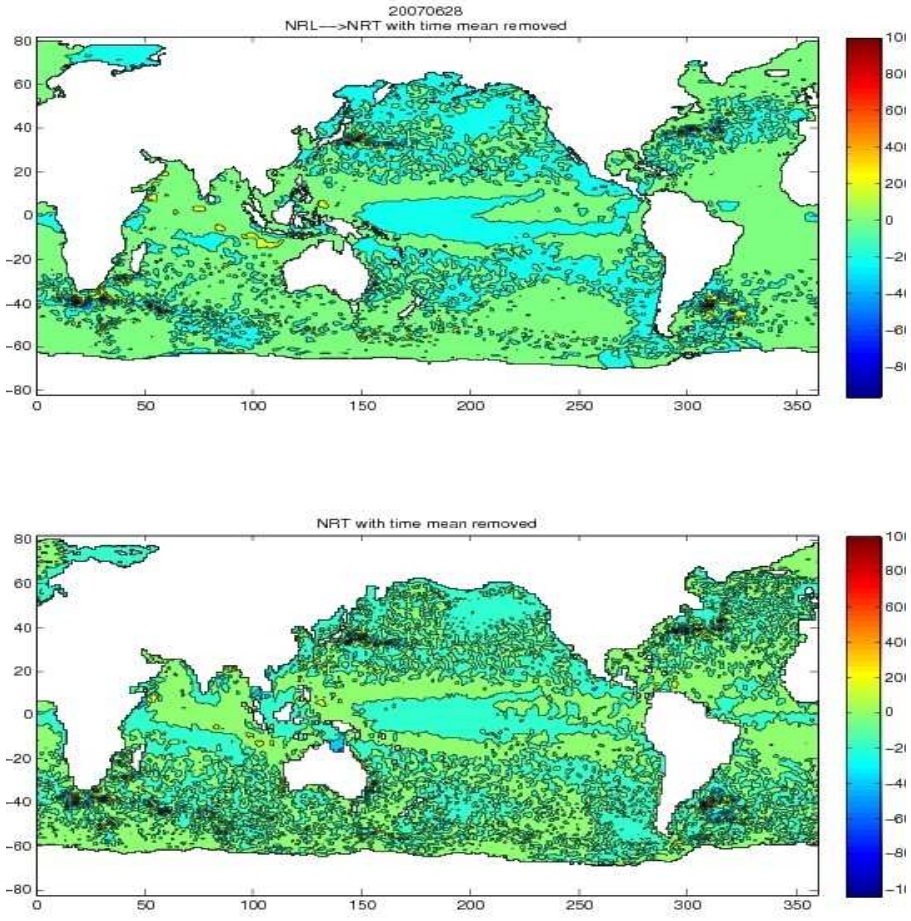


Figure 2-9b. Maps of NRL→NRT and NRT sea surface heights on the same day as Figure 2-9a with the time means removed.

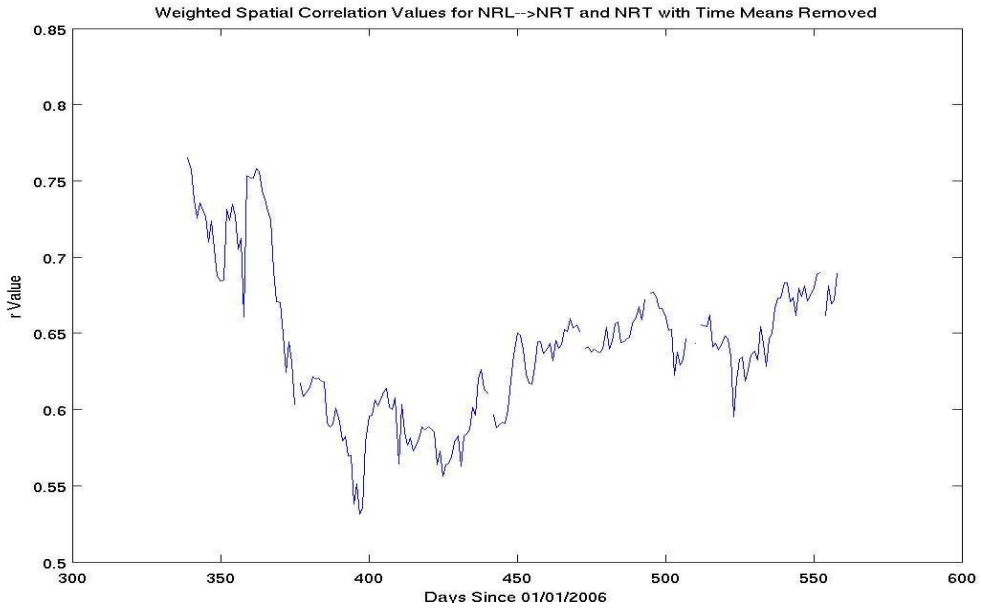


Figure 2-10. Weighted spatial correlation values for NRL→NRT and NRT vs. time with time means removed. The correlation is good given this sort of data.

To check the hypothesis based on a visual comparison that differences between the variance in the NRL→NRT product and the variance in the NRT product are due to smoothing differences in the raw height data products (NRL is smoother than NRT), the correlation is computed between NRL→NRT and NRT over the common time period, then spatially smooth the NRT at different levels to see if and how the correlation changes. The smoothing method is as follows: take a box of certain size around each point, simply find the average of the values in that box, then replace the point with that average value. Figure 2-11a shows the histogram for the correlation values at each point over the common time period for NRL→NRT and NRT before any spatial averaging occurs. Figure 2-11b gives the histogram for the temporal correlation values after using the simple smoothing described above with a 3x3-point box size used for averaging, and Figure 2-11c with a 5x5-point box size used for averaging. It is apparent that the distributions have less scatter with more correlation values closer to 1 with more smoothing.

Differences in correlation values before smoothing and after 3x3 smoothing and 5x5 smoothing are shown in Figures 2-12a and 2-12b, respectively. In 12a, the underlying noise distribution is clear based on the shape of the left-hand side of the distribution. Points outside the noise on the right-hand side of the distribution have improved correlation with smoothing. In 12b, more points move to the positive side of the distribution, indicating further improved correlation with more smoothing.

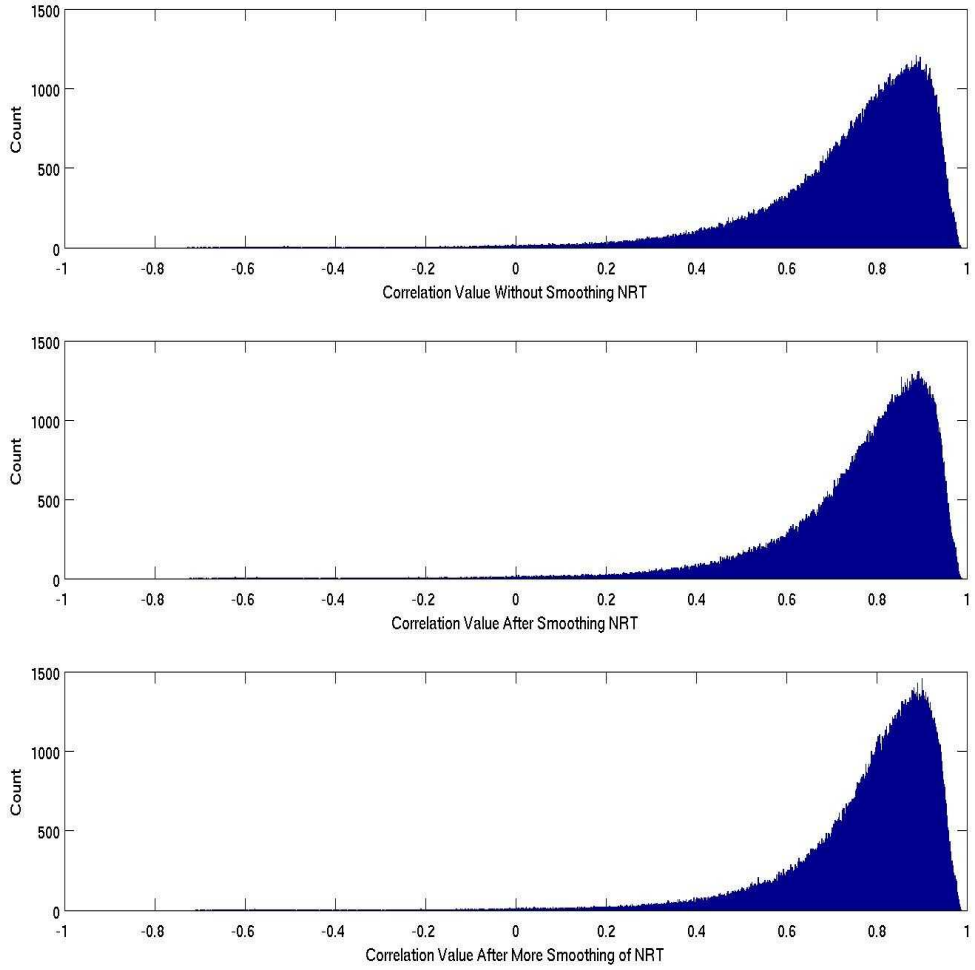


Figure 2-11. a. (top) Temporal correlation values at each spatial point for NRL→NRT and NRT over the common time period before spatial smoothing of NRT, b. (center) Temporal correlation values at each spatial point for NRL→NRT and NRT over the common time period with 3x3-point box averaging used for smoothing NRT, c. (bottom) Temporal correlation values at each spatial point for NRL→NRT and NRT over the common time period with 5x5-point box averaging used for smoothing NRT. Note the distributions have less scatter and more values are closer to correlation of 1 with an increased amount of smoothing.

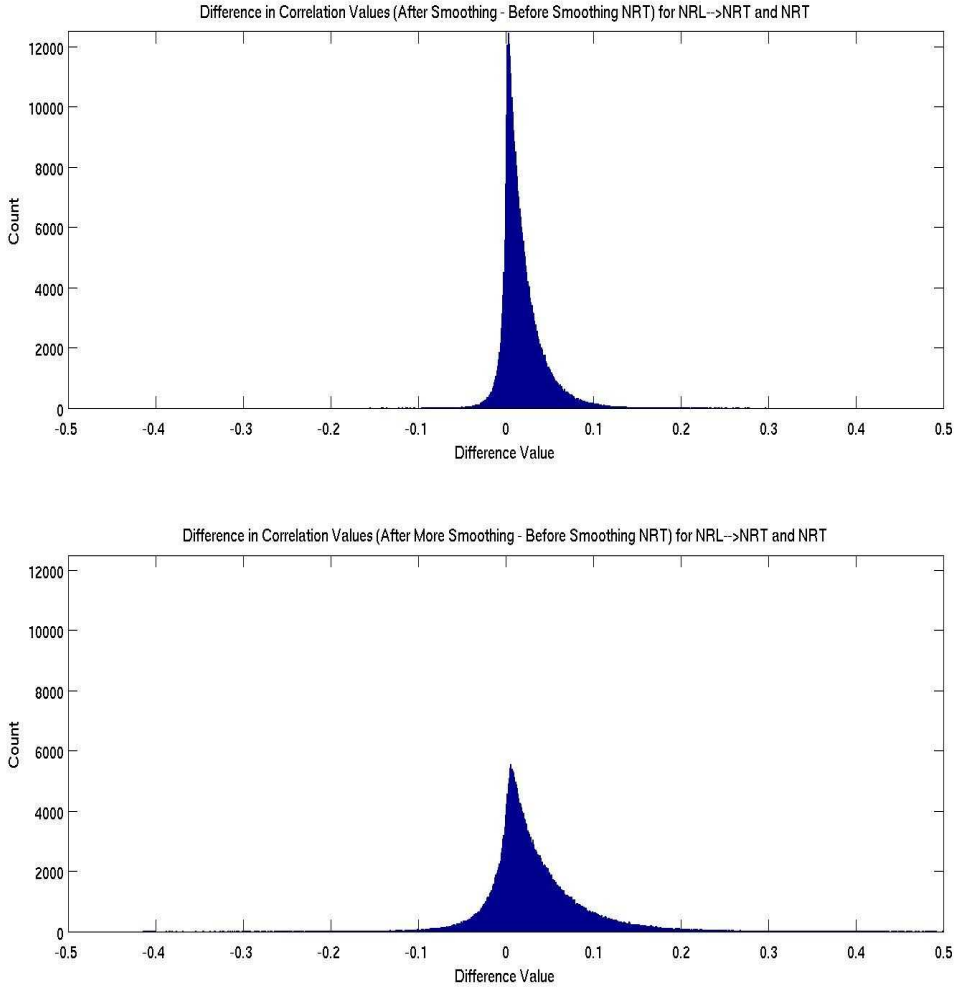


Figure 2-12. a. (top) Difference in correlation values at each point for NRL→NRT and NRT over the common time period: before smoothing of NRT – after 3x3 smoothing of NRT. Underlying shape of the noise is apparent on left-hand side of distribution, with slight shift of the distribution to positive values, b. (bottom) Difference in correlation values at each point for NRL→NRT and NRT over the common time period: before smoothing of NRT – after 5x5 smoothing of NRT. Underlying shape of the noise is apparent on left-hand side of distribution, with more points in the distribution shifted to positive values.

Figures 2-13a, 2-13b and 2-13c show the ratio of the variance in NRL→NRT to the variance in NRT before smoothing, after 3x3 smoothing of NRT and after 5x5 smoothing of NRT. Before any spatial smoothing, the peak of the ratio is shifted away from and is less than 1, indicating some deficiency of the variance in NRL→NRT compared to NRT. After 3x3 smoothing of NRT, the peak of the ratio values distribution moves toward and very near to 1, and moves just past and greater than 1 after 5x5

smoothing. These results suggest two important things. First, the smoothing of NRT, removing high-frequency/high-wave number information, makes NRL→NRT and NRT more similar. Second, an idea of the scale of the high-frequency and high-wave number information causing the differences seen in the products is revealed. A 3x3 box for averaging makes the peak of the ratio values very near to 1, and a 3x3 box is approximately 2/3º by 2/3º on the globe's surface.

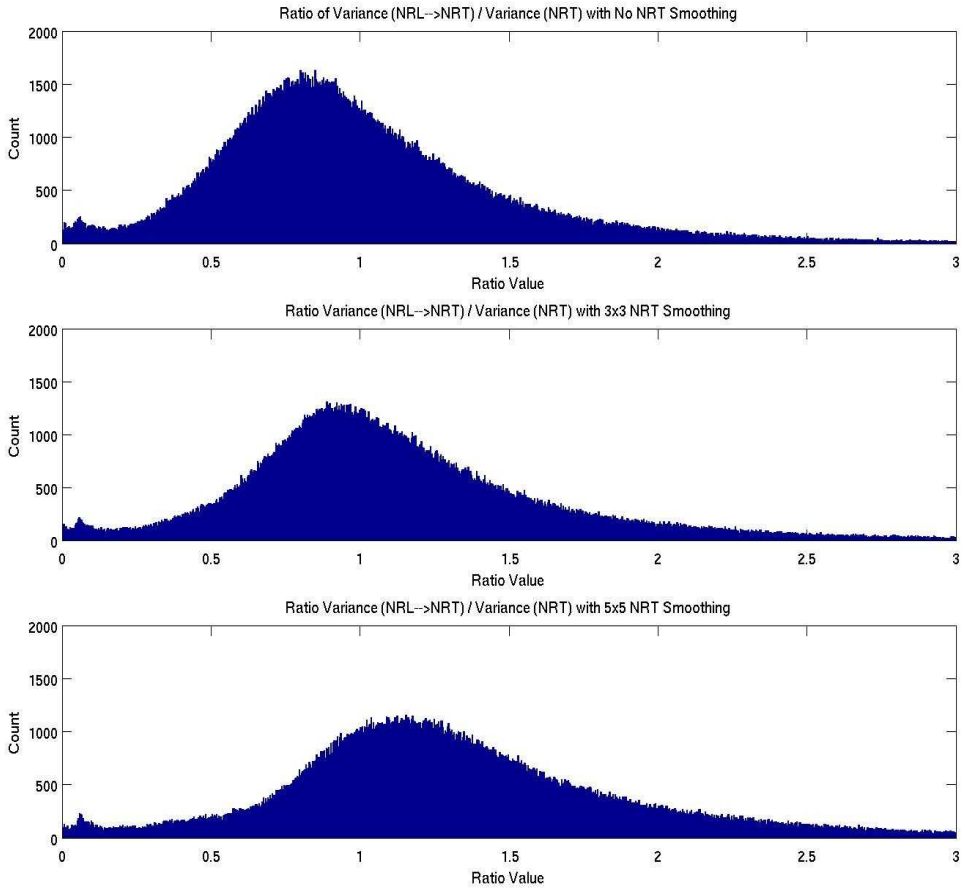


Figure 2-13. a. (top) Ratio of the variance of NRL→NRT to the variance of NRT before smoothing of NRT. Peak is shifted away from and less than 1, indicating a deficiency in variance of NRL-->NRT compared to variance of NRT, b. (center) Ratio of the variance of NRL→NRT to the variance of NRT after 3x3 smoothing of NRT. Peak is shifted toward and very near to 1, indicating smoothing NRT on this scale may account for differences in variance, c. (bottom) Ratio of the variance of NRL→NRT to the variance of NRT after 5x5 smoothing of NRT. Peak is shifted past and greater than 1, indicating smoothing of NRT may take away the variance deficiency of NRL→NRT compared to NRT.

Before drawing any further conclusions, there must be a check that the improvement in correlation seen with smoothing is not simply due to some bias that may

occur in using this method. Figure 2-14 shows the result of a 10,000-point simulation of correlation values between red noise series similar to the height data before and after smoothing. There is no bias in the distributions of the correlation before smoothing, after smoothing, or the difference in correlation values. This lack of bias in the simulation indicates the improvement seen in correlations after smoothing of the NRT product is real. Therefore, it can be said that these results support the original visual interpretation that the differences in NRL→NRT and NRT are due to smoothing differences in the raw height data products, particularly in areas containing high-frequency/high-wave number information.

To summarize, the available altimeter products from AVISO and NRL were combined to create the most up-to-date time series of global ocean heights possible. The NRL product was adjusted to look like the NRT by interpolating to the AVISO grid and adding a mean field plus a global mean offset resulting from a vertical datum difference between the DM and NRT products. Given all of the results discussed above, it has been shown that the method used here for adjusting the NRL product to make it like the AVISO NRT product works fairly well, with some caveats noted.

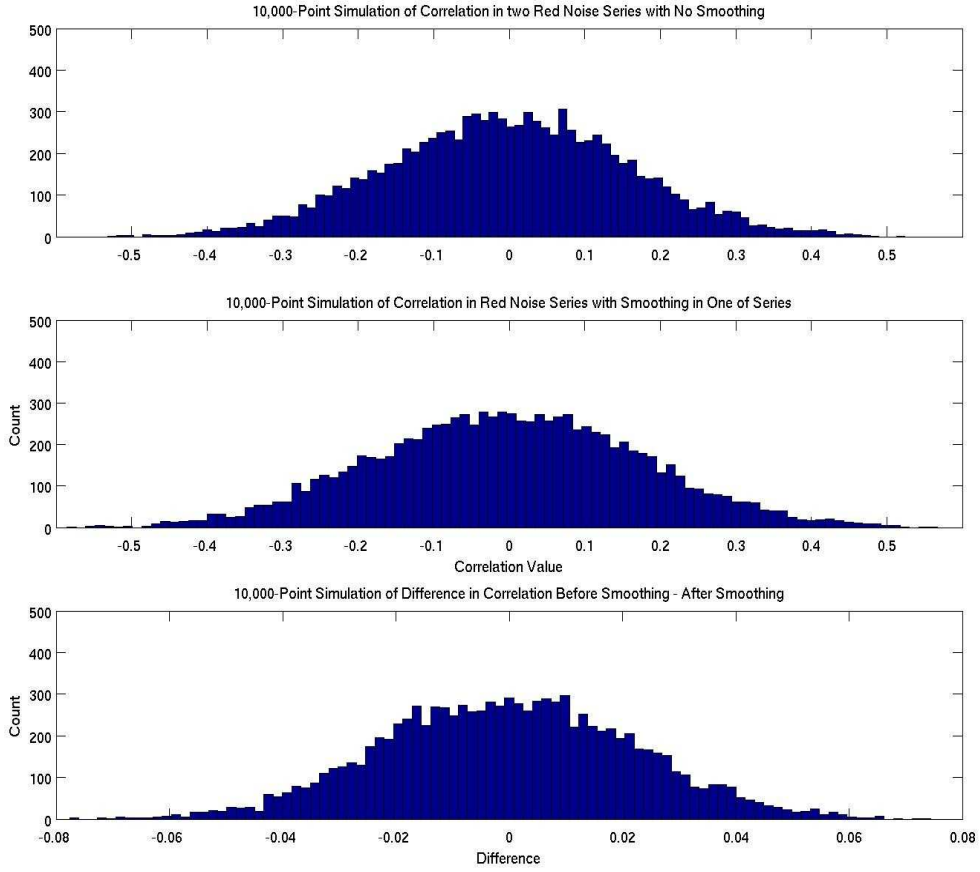


Figure 2-14. Resulting distributions from a 10,000-point simulation of correlation values. Before (top) and after (center) smoothing one of two red noise series similar to the height data time series and resulting distribution of the difference in correlation before smoothing – after smoothing (bottom). Note there is no indication of bias in the distributions due to smoothing of noise series.

2.4 Specifying an Extrapolation Method

2.4.1 10-Day Persistence Forecast and 10-Day AR Model Extrapolation

Before an auto-regressive (AR) model can be used to do an extrapolation to fill in the 5-day gap left by the most recent NRL daily data file to today, the AR model order must be chosen. Preliminary testing showed that a typical AR(1) model is not accurate in reproducing the heights time series. Instead, the linear prediction method (detailed in Section 1.2.4) using past data to predict future data is applied. Experimenting with the heights time series, it is determined the model order for producing an appropriate extrapolation is 100—using the past 100 days of each height time series to extrapolate

the next 10 days. Ten days of extrapolated data will probably not be needed, but this length of time provides a comfortable buffer in case there are any extenuating circumstances in obtaining the NRL daily data.

A 10-day persistence forecast (where forecast points 1 through 10 equal the last existing point in each height time series) is used to gauge how well the chosen AR model extrapolation is doing in predicting the next 10 days in the heights time series. The AR model extrapolation should greatly outperform the persistence forecast, especially after the first day or two, given the nature of the changes in SSH from day to day. The persistence forecast to AR model extrapolation comparison is done by applying each method to a set of 100 randomly chosen points in each time series, taking care not to have the "last point" of the test time series be any less than 100 days + 10 days from the actual end of the time series. This set of output gives 10-day "forecasts" for each method that correspond to existing data points. The results are compared using a ratio of the error in the estimated values to the variability in the time series (standard deviation) at each of the 10 days. Values that approach 1 indicate the error in the estimate becomes comparable to the variability in the data and therefore loses forecast skill. Figure 2-15a shows a map of these error values 5 days into the extrapolation. As expected, the AR model extrapolation outperforms the persistence forecast and has reasonably low error values in most areas. Even in the "worst case scenario" at day 10 of the extrapolation (maps shown in Figure 2-15b), the AR model extrapolation still shows a fair amount of skill in many areas while the persistence forecast generally fails to produce accurate height values.

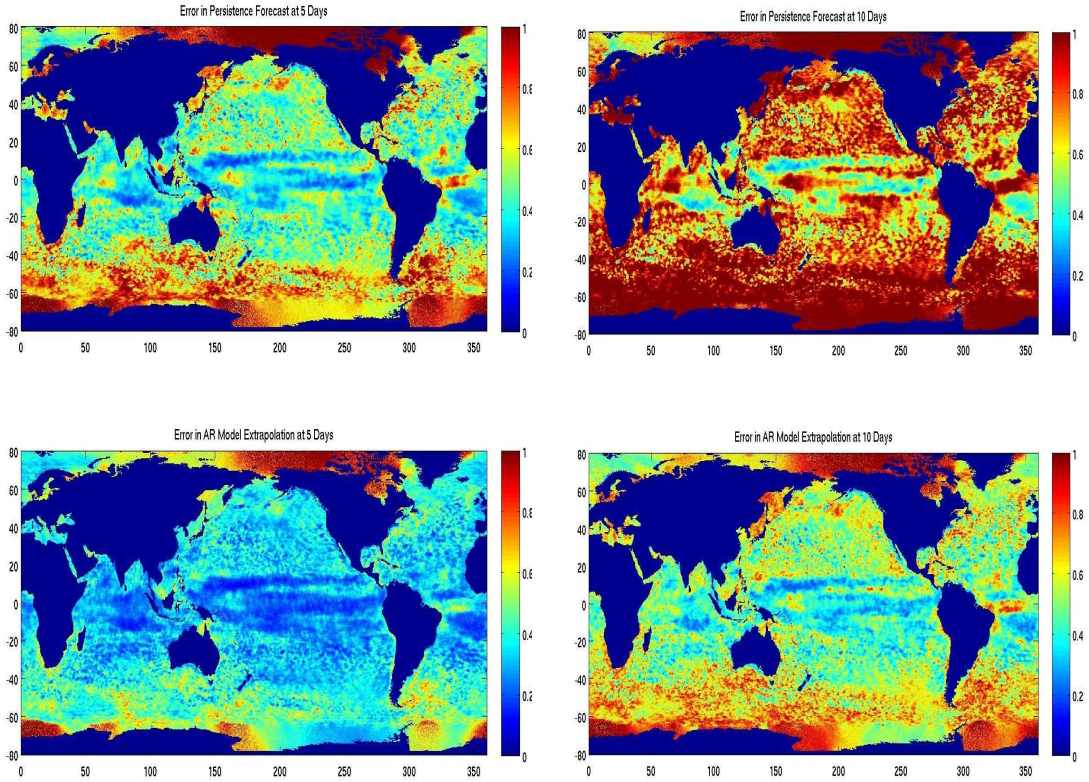


Figure 2-15. a. (left) Maps of error in persistence forecast (top) and AR model extrapolation (bottom) on day 5 of the predictions. Note values close to 1 indicate the error in the estimate becomes comparable to the variability in the data and results in loss of forecast skill.

Figure 2-15. b. (right) Maps of error in persistence forecast (top) and AR model extrapolation (bottom) on day 10 of the predictions. Note the AR model extrapolation still shows a fair amount of skill while persistence fails to reproduce height values.

It now becomes necessary to get an idea of where the AR model does and does not work well. Example time series are chosen where the error value is low and where it is approximately equal to 1—indicating the error in the extrapolate is comparable to the variability in the data. Figure 2-16a shows the time series for 3 spatial locations where the error value for the extrapolation is low. Superimposed on each time series are 3 realizations of the 10-day extrapolation (the results of extrapolation from 3 randomly chosen points in the time series). Figure 2-16b shows a time series where the error value is approximately equal to 1. The final two points from which to do the extrapolation were chosen to show what occurs if that point is chosen just before or on a dramatic

change in the data (a data “spike”). The AR model attempts to fit the spike when it is included in the 100-point series used to find the extrapolates.

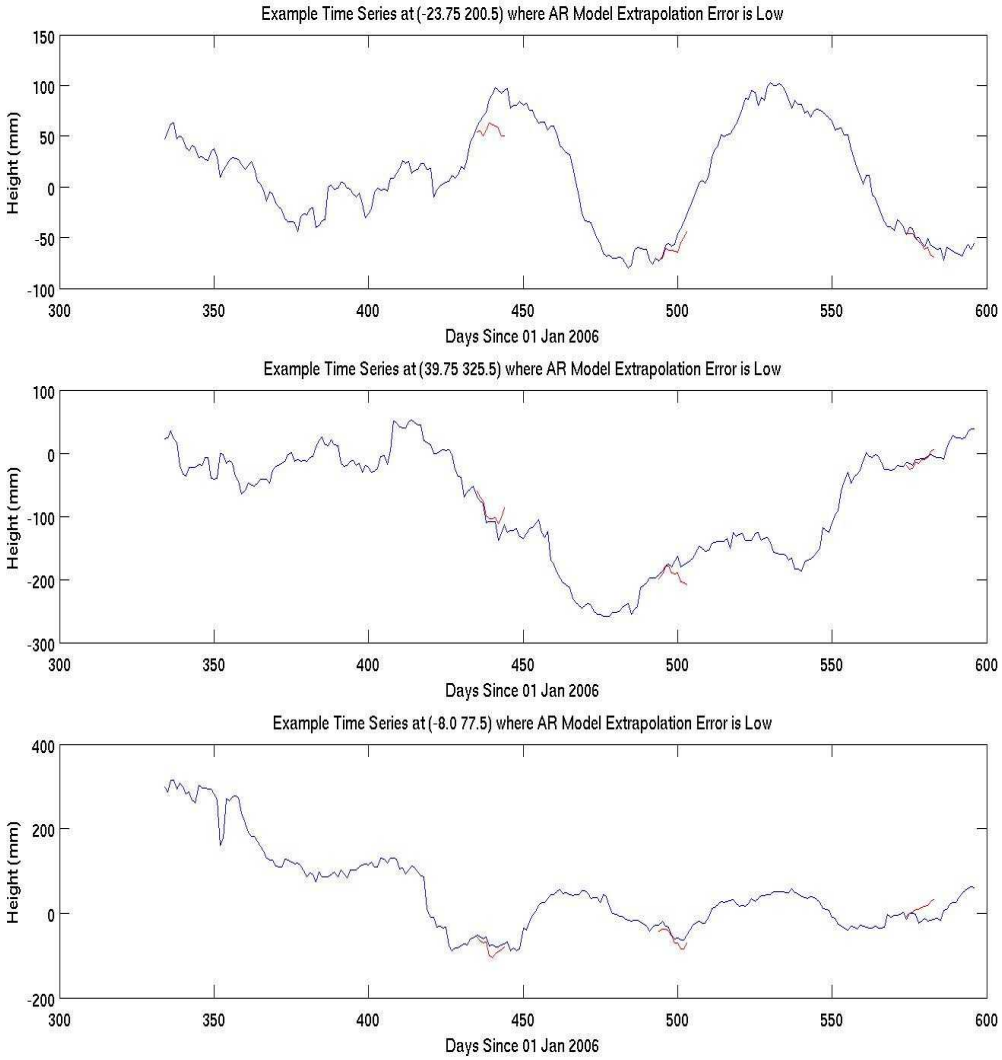


Figure 2-16a. Time series of 3 locations: (-23.75 200.5), (39.75 325.5) and (-8.0 77.5), where AR model extrapolation error is low. 10-point extrapolation from 3 different points in the data—blue line is the time series, red lines are the extrapolated values.

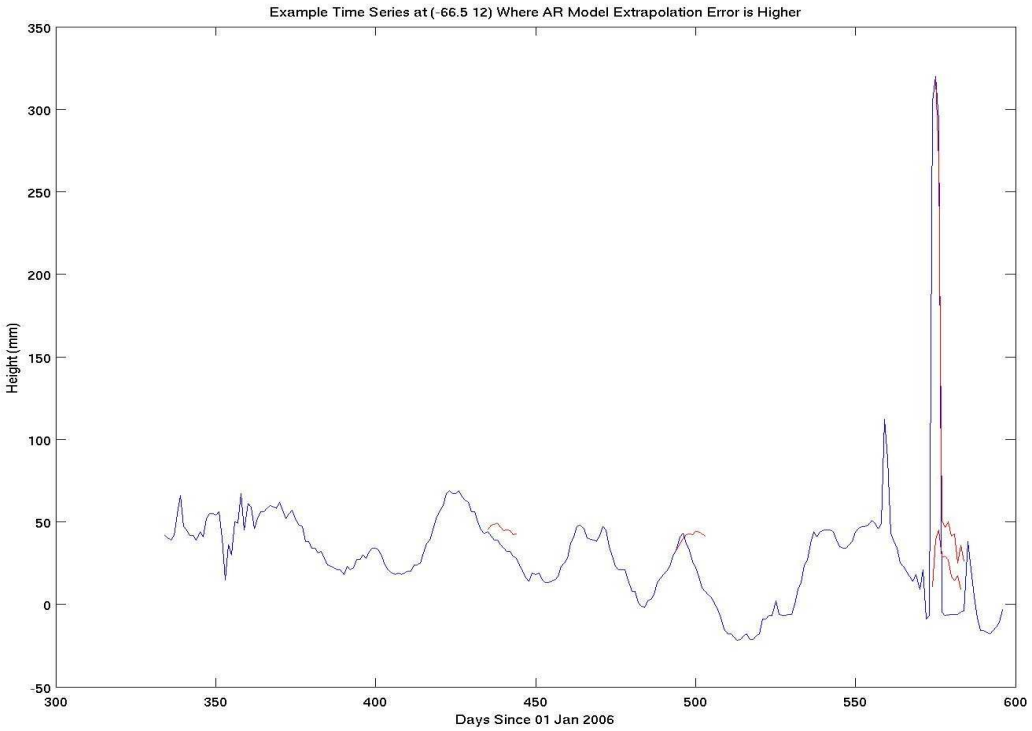


Figure 2-16b. Time series of the location (-66.5 12) where AR model extrapolation error is comparable to variability in the data. Note the large spike in the height value around day number 575—these points from which to extrapolate were chosen to show what happens to the extrapolation if there is a data spike.

Given the shape of the time series in Figure 2-16b, places where these tend to occur on the global map are identified. Figure 2-16c shows the points are located in the Arctic, Antarctic and areas of high eddy energy. These are areas that are generally more difficult to reproduce with the calculations involved in this process. A collection of all the time series for the points in the Arctic and Antarctic are shown in Figure 2-16d. The Antarctic time series are more uniform than those from the Arctic region. Since the extreme points for the Antarctic occur near the end of the time series (Southern Hemisphere Winter), this issue could possibly be related to the location of the edge of the ice sheet. The Arctic region appears to be problematic at all times of year. Based on information provided by National Snow and Ice Data Center on sea ice, this may be because (in relative terms) more Arctic sea ice tends to remain through the Northern

Hemisphere Summer (compared to the Antarctic where most of the sea ice melts in Southern Hemisphere Summer), or it also might be because of the asymmetry of Arctic ice cover (relative to the pole) with longitude due to uneven exposure to warm and cold water.

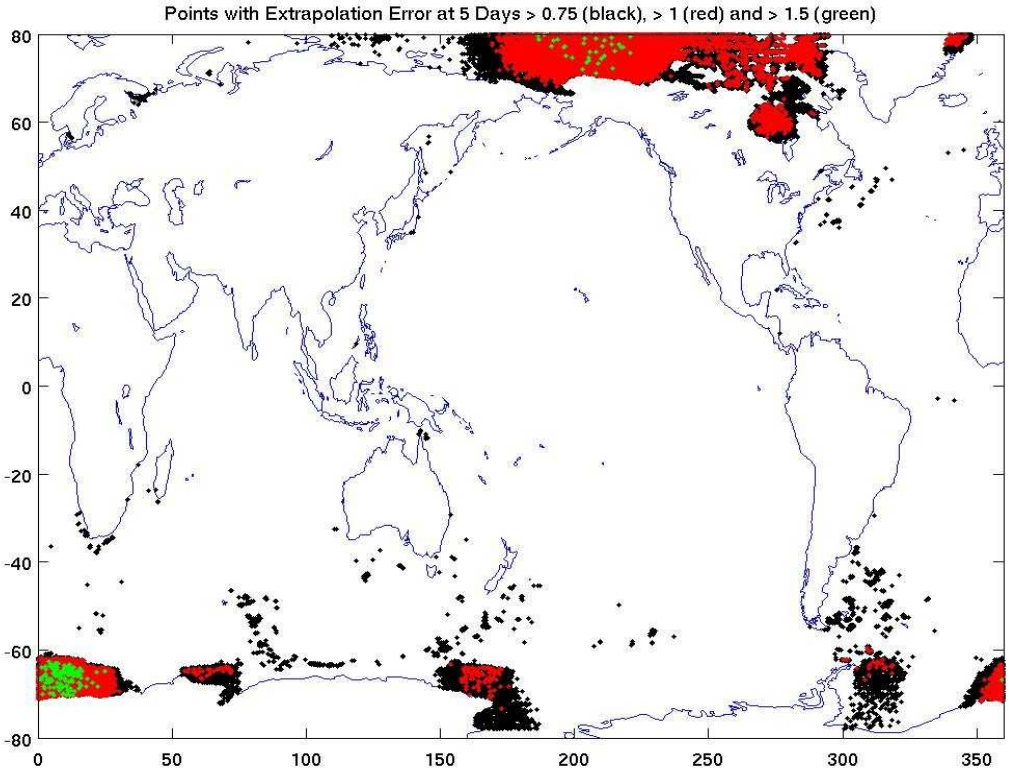


Figure 2-16c. Map of points where extrapolation error at 5 days is >0.75(black), >1(red) and >1.5(green). Note points are in the Antarctic, Arctic and areas of high eddy energy

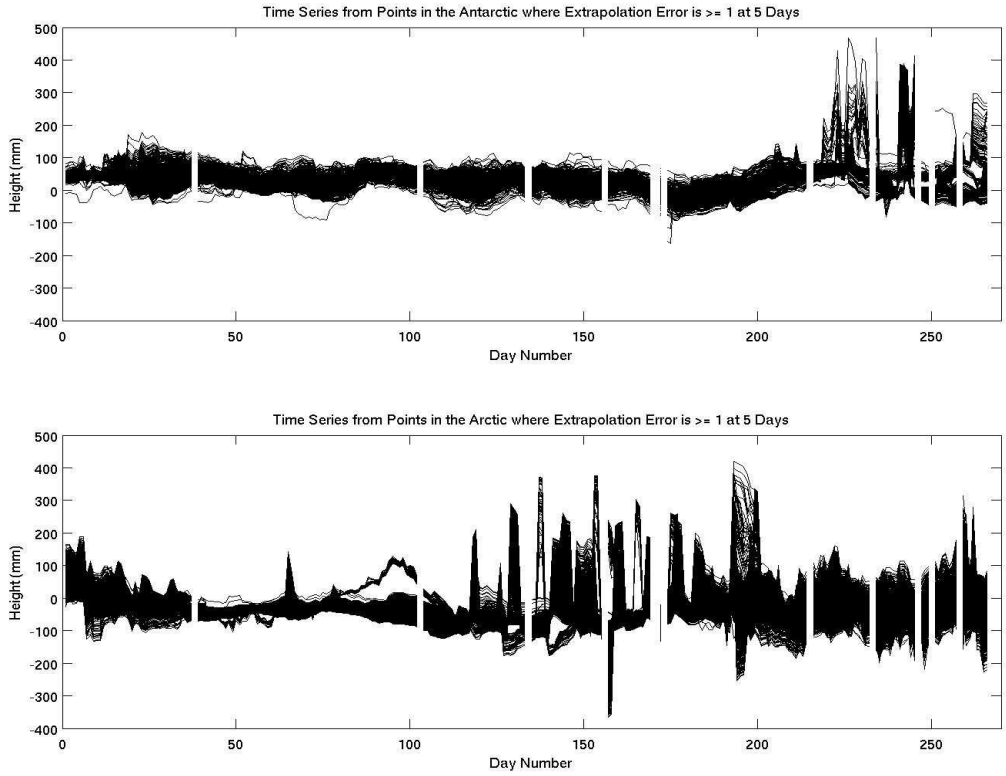


Figure 2-16d. Plot of all time series where extrapolation error at 5 days is >=1 for the Southern Hemisphere (top) and the Northern Hemisphere (bottom). Note the S. H. time series become more variable towards their end, while the N. H. time series are variable throughout.

2.4.2 Checking the AR Model Extrapolation Method with a 10-Day Hindcast

Now that much of the AR model extrapolation method has been characterized and it has been seen that it works acceptably well so far, it is necessary to set up a formal test using a 10-day hindcast. The following is the general step-by-step set-up for this test:

1. Choose the most recent AVISO NRT data available and call it “today”, making it the last day in a 10-day stretch of data.

2. Take “today's” date and the appropriate NRL file and shift it back the necessary 5 days so that the height grids match up in time.

3. Use the corresponding 5 days of NRL → NRT data to estimate the first 5 days of NRT data by interpolating it to the NRT grid and adding the long-term time mean and the 25-

mm global offset.

4. Use the NRL→NRT time series to estimate the next 5 days of NRT by extrapolating 5 days beyond "today".

Evaluating the performance of the NRL→NRT product based on the actual NRT is done over this 10-day hindcast with differences, average global area weighted standard deviation and global area weighted correlation in height anomaly between the two products. Figures 2-17a and 2-17b show the maps of actual NRT and NRL→NRT (time means removed), along with the difference map on days 1 and 10 of the hindcast. A comparison of the differences on days 1 and 10 show the maximum increase in difference on this fifth day of the extrapolation and gives an idea of how accurate the NRL→NRT product will be when using it to estimate NRT over the data gap. Looking at the average spatially weighted standard deviation between NRT and NRL→NRT gives an idea of the increase in differences over the entire 10-day period. Figure 2-18 shows there is an approximate 25% increase in the average standard deviation from day 1 to day 10. The spatially weighted correlation in height anomaly shown in Figure 2-19 is another perspective on how well the NRL→NRT product can do in estimating NRT. The correlation value remains relatively high (>0.7), even on the last day of the extrapolation, with an approximate decrease of 15% over the time period. All of the results from this 10-day hindcast test indicate the NRL→NRT product is sufficiently accurate to be turned over to ESR for incorporation of this method in producing height fields for OSCAR.

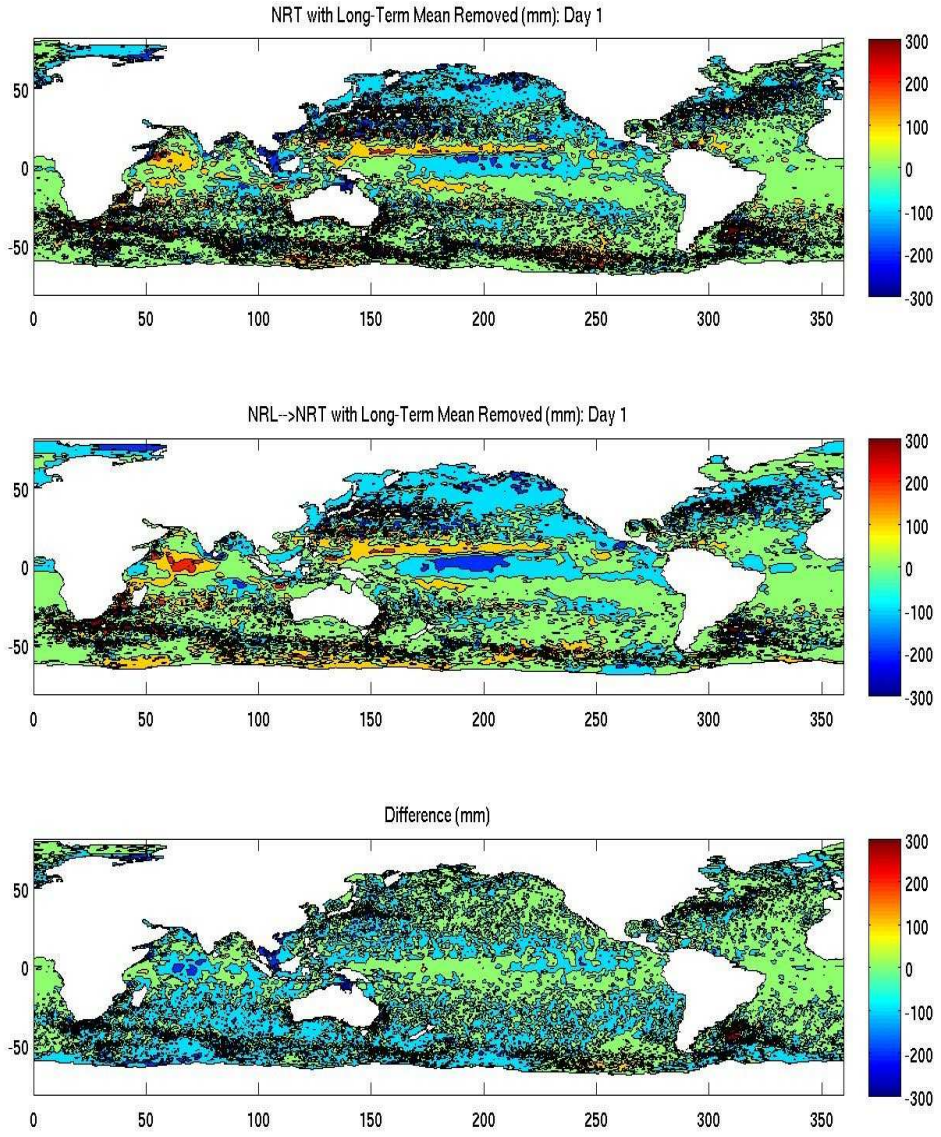


Figure 2-17a. Map of NRT (actual), NRL → NRT (with long-term time mean removed from both) and difference between the two on day 1 of the hindcast with no extrapolation.

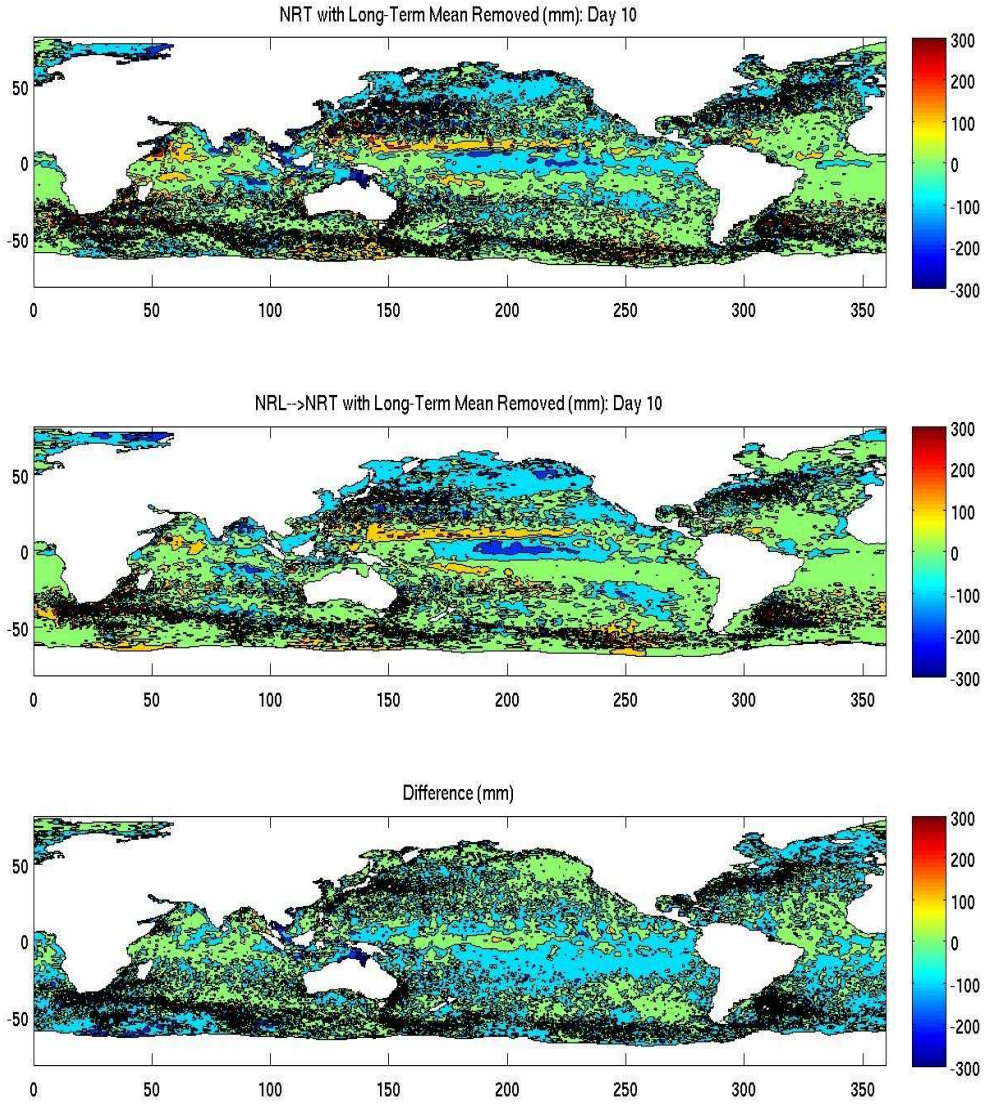


Figure 2-17b. Map of NRT(actual), NRL→NRT (with long-term time mean removed from both) and difference between the two on day 10 of the hindcast on day 5 of extrapolation. Note the increase in the difference values.

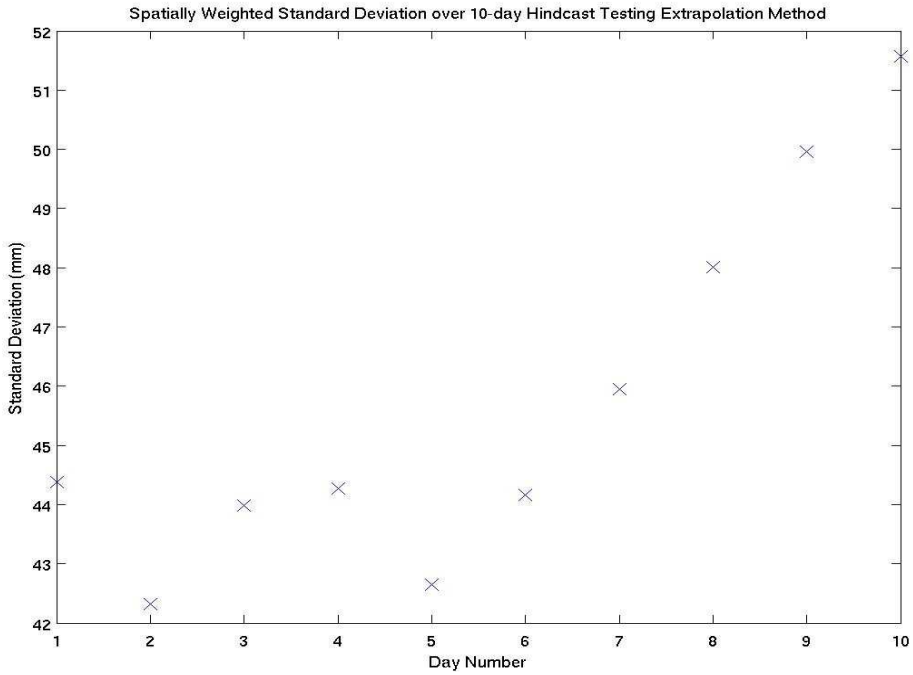


Figure 2-18. Average spatially weighted standard deviation between NRT (actual) and NRL→NRT for the 10-day hindcast. Note the roughly constant variability for the first 5 days and the increase of approximately 25% over the 10-day extrapolation period.

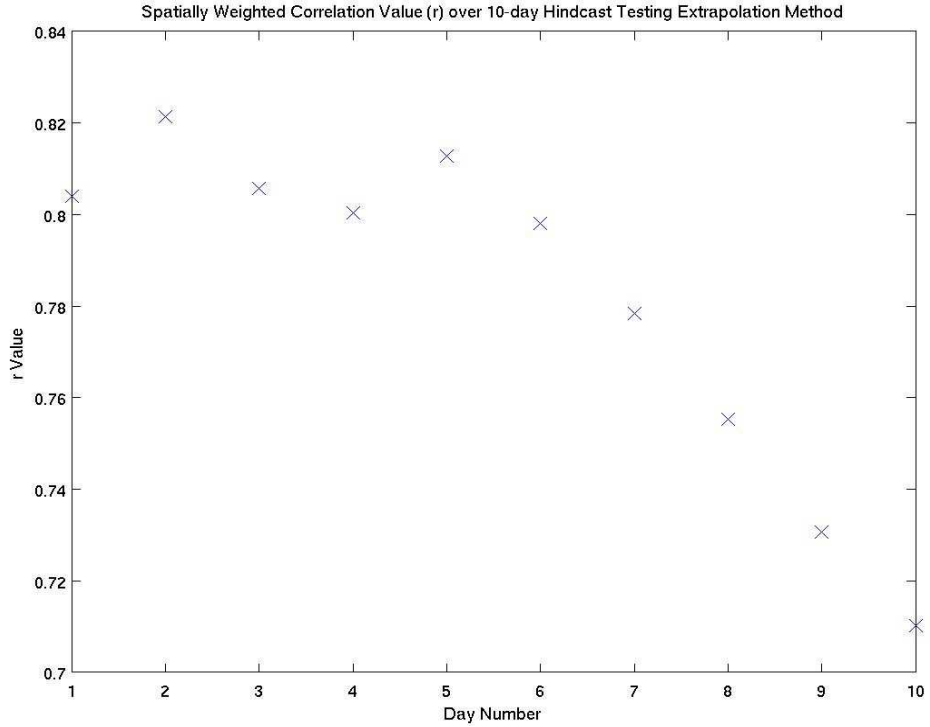


Figure 2-19. Spatially weighted correlation in height anomaly of NRT (actual) and NRL→NRT for the 10-day hindcast. Note the roughly constant correlation for the first 5 days and the decrease of approximately 15% over the 10-day extrapolation period.

2.5 Test Production of the NRL→NRT Product

2.5.1 Particulars of Output Files

The heights product output files were produced daily via automated system at the University of South Florida College of Marine Science and made available for download by the OSCAR team at ESR. After approximately 2 years, the software was transferred to ESR and integrated into the OSCAR system. The properties of these daily files and the automated checks used are detailed in this section. The file name for the NRL→NRT product is of the form NRL*NNN*.mat, where *NNN* is the 3-digit number representing the last day in the extrapolation given as the number of days since 01 January 2006. The .mat file extension indicates the file is in MATLAB data format. The variables in the output file are:

- ht(30,915,1080) = 30 days of height values generated by the NRL→NRT product
- time(30,1) = 30-day time vector corresponding to the height grids, given in days since 01 January 2006 (where 01 January 2006 = day 0)
- extrap(30,1) = 30-day logical variable that = 0 when the height grid is NRL interpolated to the NRT grid only (no extrapolation) and = 1 when the height grid is the result of NRL extrapolated and interpolated to the NRT grid
- latgr(915,1) and longr(1,1080) = latitude and longitude vectors, respectively
- readme = a string variable containing descriptive information on variables “time” and “extrap”

Typically, the number of extrapolated days will be 6. This is because in reality, the NRL daily height file is not available until the afternoon. So, when the NRL→NRT height output is produced, the routine must run that night. This makes it necessary to add one more day of extrapolation so that last day of the resulting file that is ready in the morning matches up with the day the file is made available for download. If the most recent NRL

file is missing from the NRL database, the product will have one more day of extrapolated data. The number of extrapolated days will change only occasionally, but at unpredictable times, so this is the purpose of providing the “extrap” variable.

With the working NRL→NRT output, some sort of quality check must be implemented to ensure something in the processing does not go awry from one day to the next. First though, a note on daily changes in extrapolated height grids—any quality check will reveal that a particular height grid that is the result of extrapolation will change values from one day's dataset to the next, but it should be kept in mind that each new daily dataset for that particular extrapolated day provides a better estimation for that day than the previous dataset. The parameter used for comparing today's output to yesterday's is global standard deviation between days that are common in the datasets. Days that are NRL interpolated to the NRT grid only (no extrapolation) should be exactly the same, (standard deviation equal to 0) and for the extrapolated days, the standard deviation should increase from the first day of extrapolation to the last. A safety threshold is built in to flag a dataset if this quality check parameter becomes too large, say, greater than 0.5. In the time the product has been tested, values for the global standard deviation generally have been less than 0.5, so the choice of 0.5 seems reasonable. Looking at a distribution of 1 year of past standard deviation values (Figure 2-20) reveals that 0.5 is unlikely, while most values are around 0.25 to 0.3.

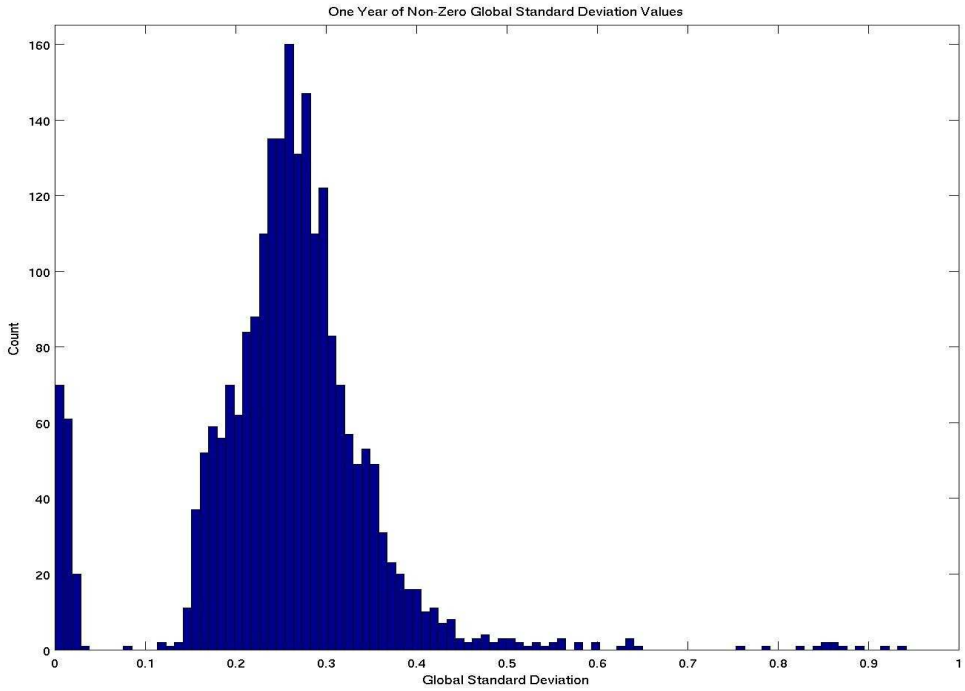


Figure 2-20. Distribution of 1 year of past global standard deviation values used in the daily quality check of the heights product.

2.5.2 A Note on the Behavior of the Standard Deviation Value used as a Quality Check

Looking back on 1 year of past standard deviation values, a couple of points are worth noting. First, when there is a gap in the series due to an NRL data outage, the values on days during the gap are very small (approximately 0.01), while the values corresponding to the day the data outage ends are very large. The day corresponding to the day the data outage ends is likely to fail the quality check. The longer the gap has been, the larger the standard deviation value will be. Figure 2-21 shows 1 year of past standard deviation values along with number of extrapolated days (scaled down for convenience) , and illustrates this point.

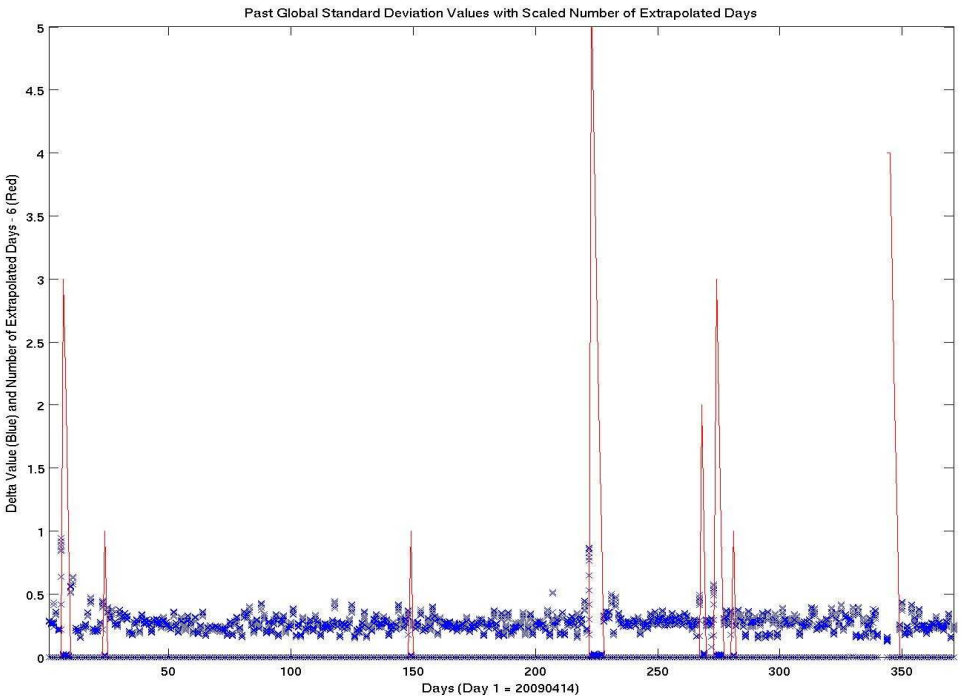


Figure 2-21. One year of past global standard deviation values used in the daily quality check of the heights product along with the number of extrapolated days, scaled down for convenience, to show how NRL data outages affect the quality check parameter. Note large values result after the data gap ends.

2.5.3 Expanding to the Full NRL Grid

During the first several days of producing the NRL → NRT product, the NRL height files were reduced from one-eighth degree resolution to one-quarter degree resolution in order to reduce computation time during the testing phase. After that time, the full one-eighth degree resolution NRL grid is used. In this section, differences between the test version and the final NRL → NRT product are documented. First, a simple difference computed over all 30 days of data for a particular file is examined. Figure 2-22a shows the distribution of the difference over all 30 days and Figure 2-22b shows locations where, for any of the 30 days, a difference value was greater than three thresholds: 10, 30 and 50 mm. Percent coverage over all valid points for all 30 days of data are 13%, 1.9% and 0.5%, respectively.

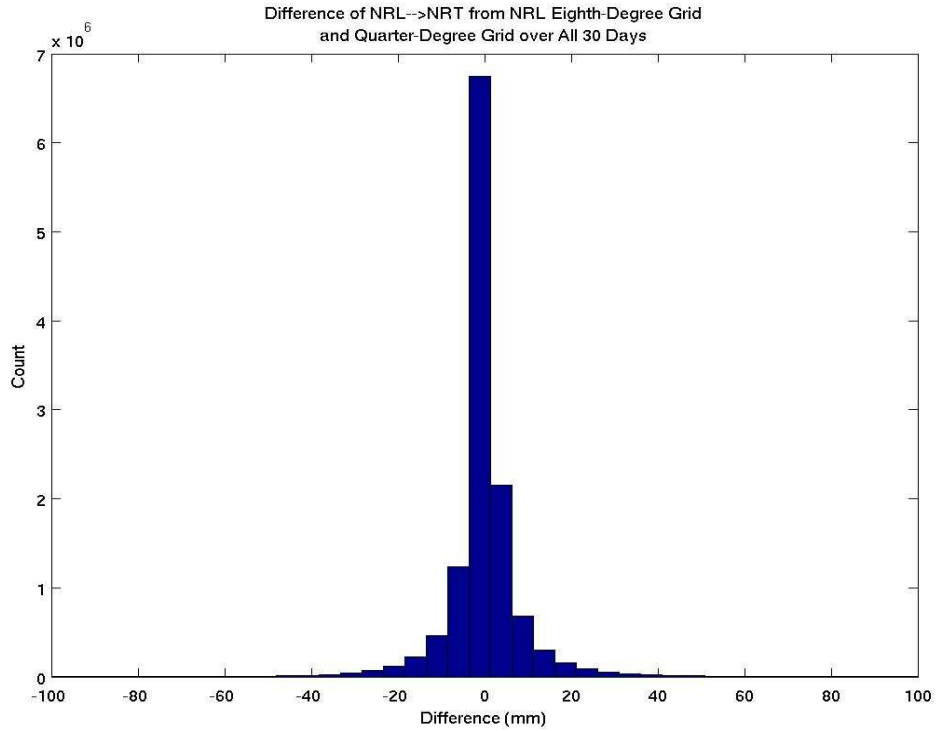


Figure 2-22a. Difference comparison of 30 days of product data computed using the 1/4 degree resolution NRL grid and the 1/8 degree resolution NRL grid.

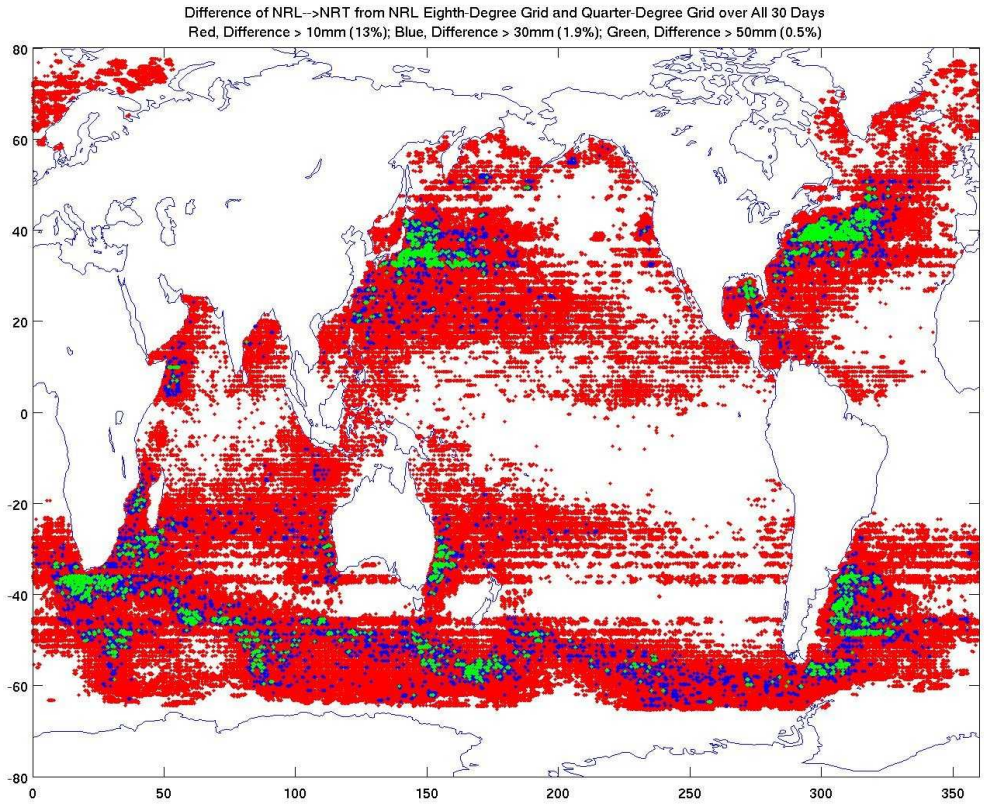


Figure 2-22b. Map of differences between product data computed using 1/4 degree NRL grid and 1/8 degree NRL grid over all 30 days with 3 thresholds: >10mm, >30mm and >50mm. Percentages are from the ratio of the number of points meeting the threshold over any of the 30 days and the total number of valid points over the 30 days.

Earlier, the standard deviation of the differences was used as a diagnostic check comparing NRL→NRT and actual NRT. Here, the difference in this parameter is utilized for the one-quarter and one-eighth degree NRL grids used in generating the NRL→NRT product. Computing this difference shows where the ability of this method to reproduce variability in SSH is the most affected by the change. Figure 2-23a shows the distribution of the difference, while Figure 2-23b indicates the locations of values greater than three thresholds: 10, 30 and 50 mm. Percent coverage values given (all <1.5%) are over all valid points on the standard deviation of differences grid. The larger-valued points occur in high-energy regions and along certain portions of coastline. Differences are large in these areas because there can be steep gradients over relatively small features

that are washed out when the full-resolution grid is not used. There are some valid data points gained in moving to the one-eighth degree NRL grid, and they are mapped out in Figure 2-24. These points are in coastal areas both along continents and around islands.

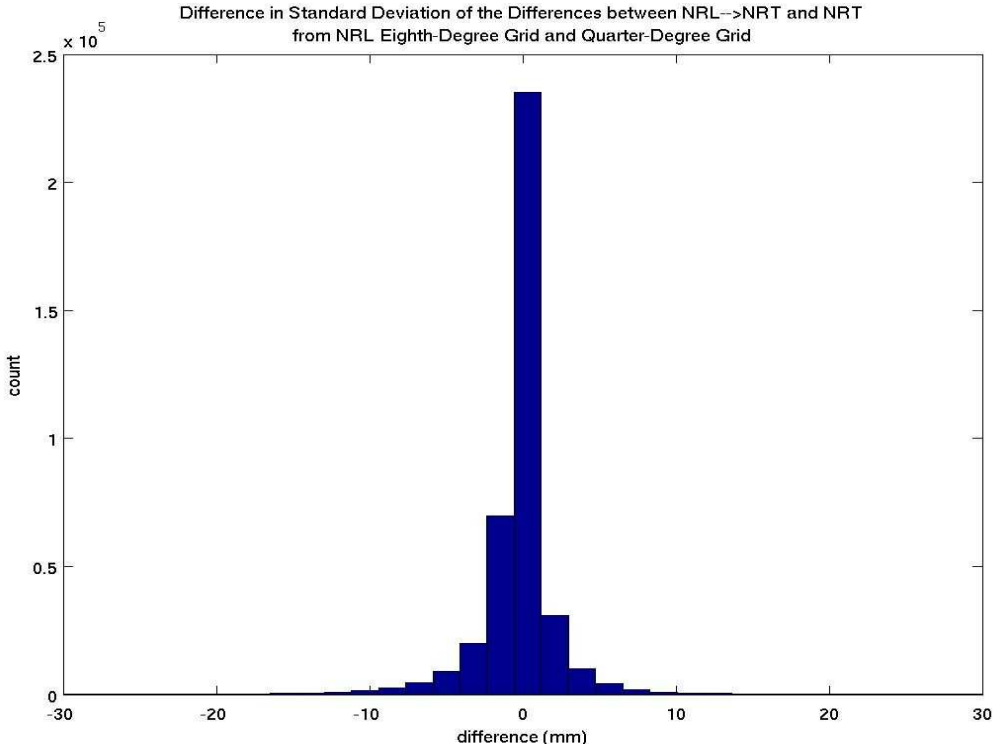


Figure 2-23a. Differences between standard deviation of the differences for NRL→NRT and actual NRT using 1/4 degree NRL grid and 1/8 degree NRL grid.

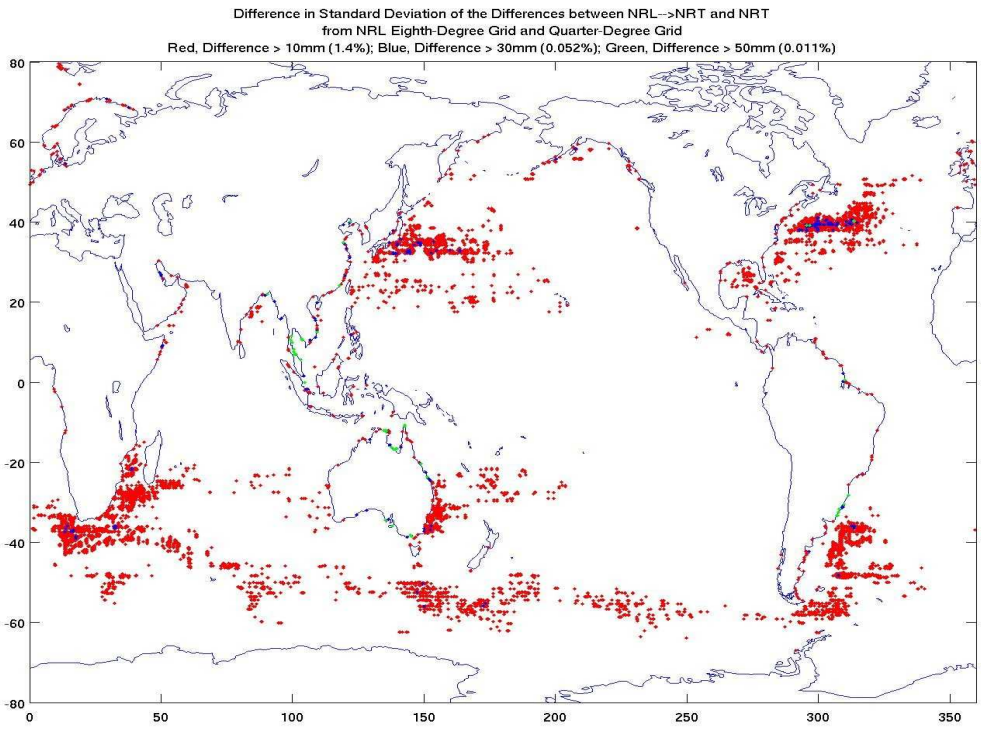


Figure 2-23b. Map of differences between standard deviation of the differences for NRL→NRT and actual NRT using 1/4 degree NRL grid and 1/8 degree NRL grid with 3 thresholds: >10mm, >30mm and >50mm. Percentages are given over the total number of valid points on the grid.

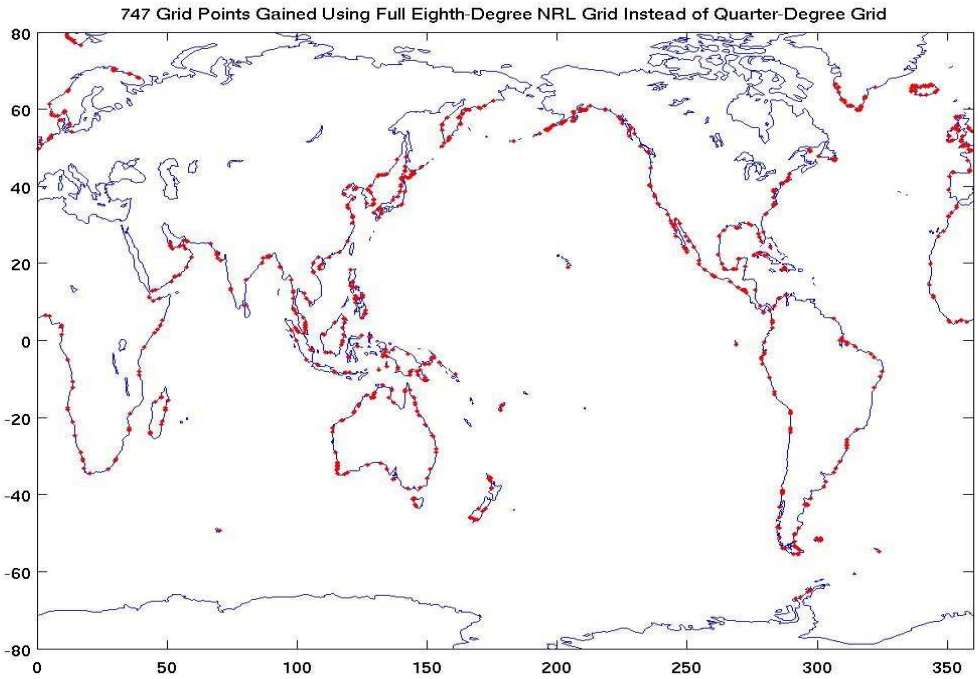


Figure 2-24. Map of valid data points gained in the final NRL→NRT product grid when changing over to the full 1/8 degree NRL grid from the lower-resolution 1/4 degree NRL grid.

2.6 Concluding Remarks for **NRL→NRT**

With the tests and comparison parameters provided in this chapter, the strengths and pitfalls of the NRL→NRT product dataset are shown. Most notably problematic are those regions where the difference in the amount of smoothing between the raw NRL and NRT products most affects the results (areas of high eddy energy and high wave number features), and those coastal regions where there is presumably some difference in the tidal model or correction being used in the raw NRL and NRT products. In a global sense, however, the NRL→NRT product does well in filling the data gap from the last day of AVISO NRT to today, with acceptable errors in extrapolation given the methods that have been presented here.

3 Evaluation of OSCAR in the IAS

3.1 Chapter Overview

OSCAR is evaluated in the IAS in three ways. First, AVISO heights (on which the geostrophic component for OSCAR is largely based) are compared to sea level series measured by tide gauges. Second, OSCAR velocities are compared to moored acoustic Doppler current profiler (ADCP) velocity data. Third, OSCAR velocities are compared to shipboard ADCP velocity data. The in-situ velocity data are spatially distributed in different areas of the IAS. A variety of comparison techniques in the time and frequency domains are utilized.

A network of tide gauges measures local fluctuations in sea level in many coastal locations around the globe. These often long, reliable and well-maintained records are used for evaluation purposes with many other types of sea level measurements such as altimetry. The AVISO DM product is chosen for the comparison because the archived OSCAR product going back to the beginning of the altimeter record (October 1992) being used for the velocity comparisons detailed in later sections has been generated using the AVISO DM height fields. The AVISO DM series are compared to tide gauge records here to maintain consistency across the comparisons. Tide gauge records in the IAS are compared to corresponding time series extracted from the AVISO DM product. The AVISO series cannot capture the higher-frequency variability seen in the tide gauge records. For each tide gauge location, the “high-frequency roll-off” value is identified

from the transfer function to determine the higher frequencies over which AVISO is deficient compared to the tide gauge series. These values vary geographically and have a range of values corresponding to timescales of ~ monthly and shorter. The inability of the AVISO product to reproduce the variability in the heights at these timescales stems from the interpolation processes in both space and time that are necessary to create a gridded field from along-track data. In the coherence squared, relatively narrow bands of low coherence squared values sometimes correspond to tidal aliasing frequencies for the altimeters combined to make the AVISO product. This is attributed to the global tidal model used to remove the aliased signal from the altimetry data also removing some of the actual sea level signal recorded by the altimeter.

The long, reliable moored ADCP velocity data records in the IAS available over the OSCAR times are found in the Gulf of Mexico. Vector correlation from Crosby et al. 1993 is used to match the velocity data series with the corresponding OSCAR velocities. Vector correlation values improve with increasing water depth and distance to nearest land at the comparison location. Altimetry is more accurate in deeper water and farther from the coast, indicating the altimetry plays a dominant role in the ability of OSCAR to reproduce measured velocities. The method of computing rotary spectra to compare vector data detailed by Mooers 1973 is used to gauge the agreement of the velocities in frequency space. The high-frequency roll-off, determined from the ratio of the rotary auto spectra, is found to have a range of values similar to those found in the AVISO to tide gauge comparison and show that the amount of smoothing in OSCAR varies geographically. The rotary cross spectral quantities of inner coherence squared and phase show that the measured velocity series from the moored ADCP locations and the OSCAR

velocity series have similar phasing and a fair amount of agreement over frequencies corresponding to timescales of monthly and longer.

The shipboard ADCP velocity data used in the comparison to OSCAR velocities is distributed throughout the northern Caribbean. Velocity data along ship tracks are smoothed to have a similar spatial and temporal resolution so that they are easily matched point-by-point with OSCAR. Because the vector velocity series vary in both space and time, vector differences and direct comparisons of speed and direction are used to determine the agreement between the measured and OSCAR velocities. The most prominent discrepancy is that OSCAR underestimates the larger-magnitude measured velocities by a factor of ~ 2. Since the largest speeds in the IAS occur along a narrow jet of highly energetic flow from the Yucatan through the Gulf Stream, it stands to reason that additional spatial and temporal smoothing done in the processing of OSCAR hinders its ability to reproduce these larger speeds.

3.2 Tide Gauge Sea Level Comparison to Altimetry: Data and Altimetry Series Chosen for Comparison

3.2.1 Data Sources

Tide gauges are placed in coastal areas to measure changes in water level over time. Sea level data from tide gauges has many practical applications and is used to calibrate altimeter SSH data. The Joint Archive for Sea Level (JASL) at the University of Hawaii Sea Level Center houses a collection of tide gauge data from around the world. This database is supported by NOAA's National Coastal Data Development Center and is maintained in conjunction the the National Oceanographic Data Center and the World Data Center for Oceanography, Silver Spring. Tide gauge data series in the IAS were

provided through a direct request to JASL. The gridded altimetry product used in this comparison is the delayed mode (DM) AVISO as discussed in Section 1.2.3. Figure 3-1 maps all of the tide gauge records in the IAS available for comparison to OSCAR.

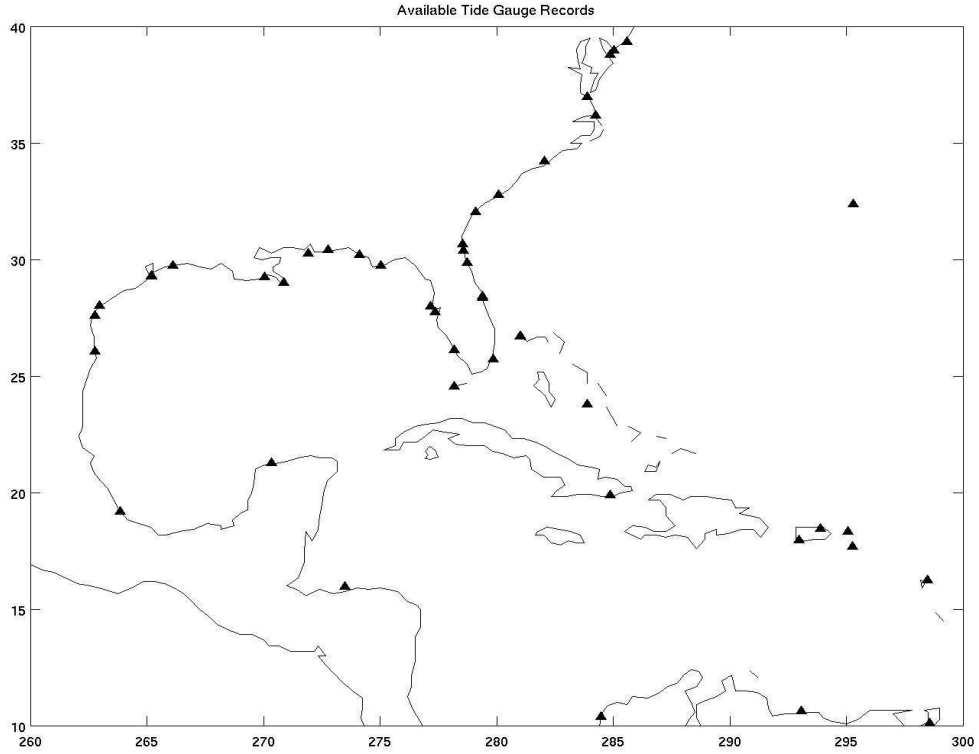


Figure 3-1. Available Tide Gauge Records in the Intra-American Sea.

3.2.2 Matching Tide Gauge Data to Altimetry Series

The tide gauge records from JASL have been smoothed to daily temporal resolution for inclusion into the database. AVISO's temporal resolution is weekly, so it is interpolated to daily using a combination of linear and folding interpolation where data points on either side of the gap are used to estimate the values between. The folding interpolation method attempts to preserve the spectral characteristics of the original series. A box, centered on the tide gauge location, approximately one square degree in size of AVISO DM SSH series is selected and each series is compared to the tide gauge

sea level series. The AVISO series with the maximum correlation squared value between the records is chosen as the best series for the comparison. This search is performed because the spatial smoothing done in the creation of a gridded product such as AVISO does not necessarily allow for the assumption that any particular latitude and longitude grid location is the same between datasets. The tide gauge series and the "best match" AVISO series are plotted. In all cases, visual inspection of the series shows a fair bit of agreement overall except for high-frequency information in the tide gauge data that the AVISO or any other altimeter-based product cannot capture.

3.3 Tide Gauge Sea Level Comparison to Altimetry: Methods Used for Comparison

To better characterize the differences between the tide gauge and altimetry series, a comparison is made in frequency space. Two quantities related to spectra, coherence squared and the transfer function, are used in the comparison. The coherence squared is defined as

$$\gamma^2 = \frac{|C|^2}{(A_A)(A_{TG})}$$

where C is the cross spectrum of the AVISO and tide gauge series and A_A and A_{TG} are the auto spectra of the AVISO and tide gauge series respectively. The coherence squared is like a correlation in that it is a measure of the agreement between the series at each frequency. Variable-length smoothing with longer smoothing windows at higher frequencies having lengths of 3, 5, 7 and 9 frequency bands is used in the spectral calculation. This allows for the features of the spectrum in the noisier higher frequencies to be better discerned. The 95% confidence level is computed for the amount of smoothing used at each spectral frequency. While determining significant coherence

squared values at frequencies where the series agree may be useful, deficiencies (low coherence squared) are of greater interest. Selection criteria are set to identify what are to be considered "low" coherence squared values. A band in frequency of coherence squared must be below the 80% confidence level and at least half a smoothing width wide. Another spectral quantity, the transfer function—a ratio of the auto spectra—is computed for a frequency-by-frequency comparison of the spectra (as described by Bloomfield, 2nd edition). In this case the ratio is constructed with the auto spectrum for the AVISO series over the auto spectrum for the tide gauge series. If the series are equal, the value of the transfer function will be unity. The 95% confidence interval is computed for the auto spectra ratio. Transfer function values that fall outside of the confidence envelope are likely to be significantly different from a value of 1. The confidence intervals depend on the coherence squared—intervals decrease (increase) in width as coherence squared increases (decreases). Where the coherence squared value is low, the value of the transfer function is not well-determined. Where the series are coherent, comparisons of the power in the series can be made. When a low coherence squared value is identified, it may also show up as a transfer function value that is less than 1; however, it is also possible for the two series to have a similar amount of spectral energy at a given frequency (transfer function is 1) and still be completely out of phase (indicated by a low coherence squared value). The confidence envelope on the transfer function is also used to determine the high-frequency roll-off (HFRO) value for the comparison. The HFRO is defined empirically as the first point along the transfer function that is at least a 10% decrease in the median value from the first frequency to the Nyquist frequency and falls outside the 95% confidence envelope.

3.4 Tide Gauge Sea Level Comparison to Altimetry: Comparison Results

3.4.1 Comparison Plots

A series of plots are generated for each comparison location. Figure 3-2a shows this series of plots for St. Petersburg, FL. The map in the upper left indicates station name, location number and location. The map in the upper right gives the tide gauge location (black circle), the correlation squared values with the AVISO series in the surrounding area (dark blue with white squares are AVISO land points), a representation of the topography (white line) and the location corresponding to the maximum correlation squared value in the box (black 'X'). Notice that the highest-correlated point is not co-located with the tide gauge location. This is to be expected due to the spatial smoothing (interpolation) done during processing of a gridded product like AVISO. The center plot shows the two series for the comparison. Overall, the two series appear to be in relatively fair agreement except for the higher-frequency information in the tide gauge series that AVISO cannot reproduce. The bottom pair of plots are the coherence squared (top) and transfer function (bottom) as described in Section 3.3. The frequencies with coherence squared values meeting the selection criteria as being “low” are indicated by the vertical cyan lines. The black line is the 95% confidence level for variable-width smoothing. The transfer function plot has the 95% confidence envelope (gray shading) centered on 1 (black line). The HRFO value is indicated in both plots by the vertical green line. The red line on the far right is the Nyquist frequency for AVISO. Since the “sampling” period of AVISO is every 7 days, the frequency corresponding to a period of 14 days is the highest frequency where a fair comparison is possible. Figure 3-2b shows the AVISO and tide gauge time series for Saint Petersburg, FL broken into 2-year intervals so the higher

frequency detail can be easily seen. The tide gauge series data (red) is daily while the AVISO (blue) has been interpolated to daily from the weekly AVISO DM product. AVISO tracks along with the tide gauge on seasonal timescales rather well, in fact, the AVISO product appears to agree with the tide gauge height measurements down to monthly timescales. Note that the range in heights is 400 cm on either side of zero for each section of the time series and time is marked off in 4-week intervals. Later examples will be presented using this same style plot with the same height and time increments.

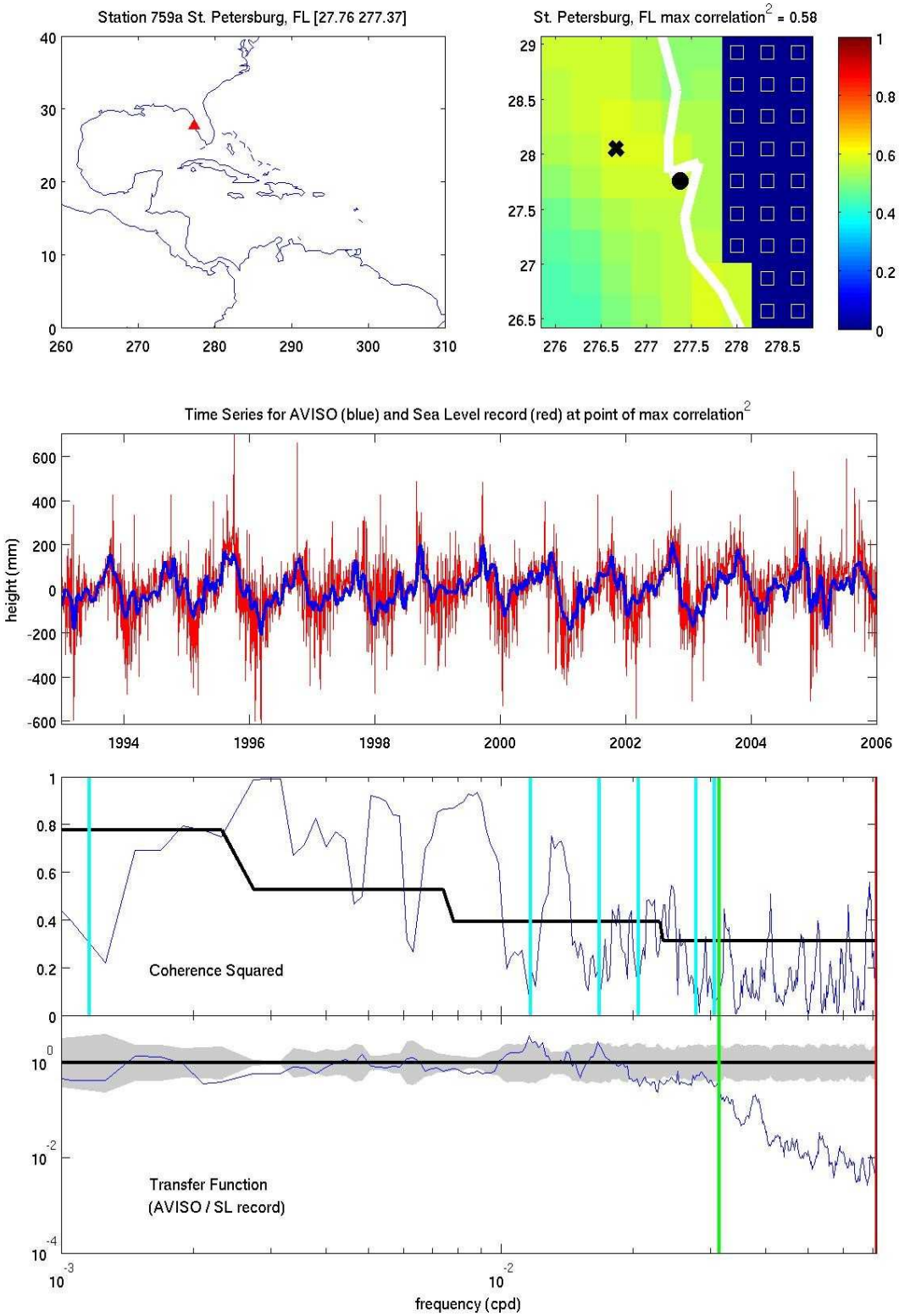


Figure 3-2a. Example AVISO and tide gauge comparison plots for St. Petersburg, FL.

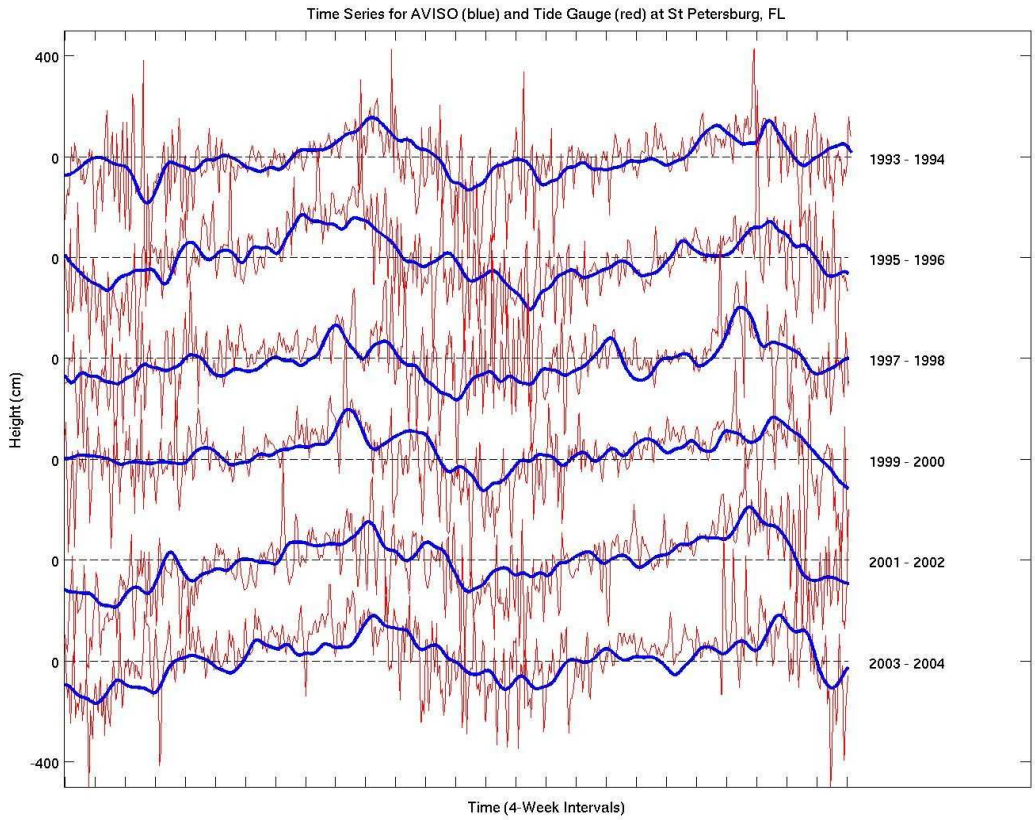


Figure 3-2b. Time series for AVISO and the tide gauge at St Petersburg, FL, broken up to show the higher frequency variability in the series.

3.4.2 Results from the Coherence Squared

The distribution of the central frequency band value corresponding to low coherence squared values is shown in Figure 3-3. The most common values range from 0.001 to 0.00681 cycles per day (cpd) and includes the annual and semi-annual frequency bands. The distribution of these values is broken up into lower, mid and higher thirds of the frequency values and plotted on the map in Figure 3-4. Note that filled symbols indicate a coherence squared value with a corresponding transfer function value significantly less than 1 at the 95% confidence level. Locations with these low transfer function values tend to be locations where it is likely that the tide gauge is placed in a bay, inlet or other topographically sheltered area. These types of locations are places

where a global tide model is not necessarily expected to reproduce local tides accurately. Because both the coherence squared and transfer function values are low, this means the series are not at the same phase at this frequency nor do they contain a similar amount of variance (energy). A pair of series can have low coherence squared and still have a transfer function value of 1 because while the series may have similar variance at a particular frequency, they can still be completely out of phase at that same frequency. Several locations have a full range of coherence squared values while others do not. There appears to be some geographical dependence in that the higher-frequency values are only located in peninsular Florida and the Caribbean.

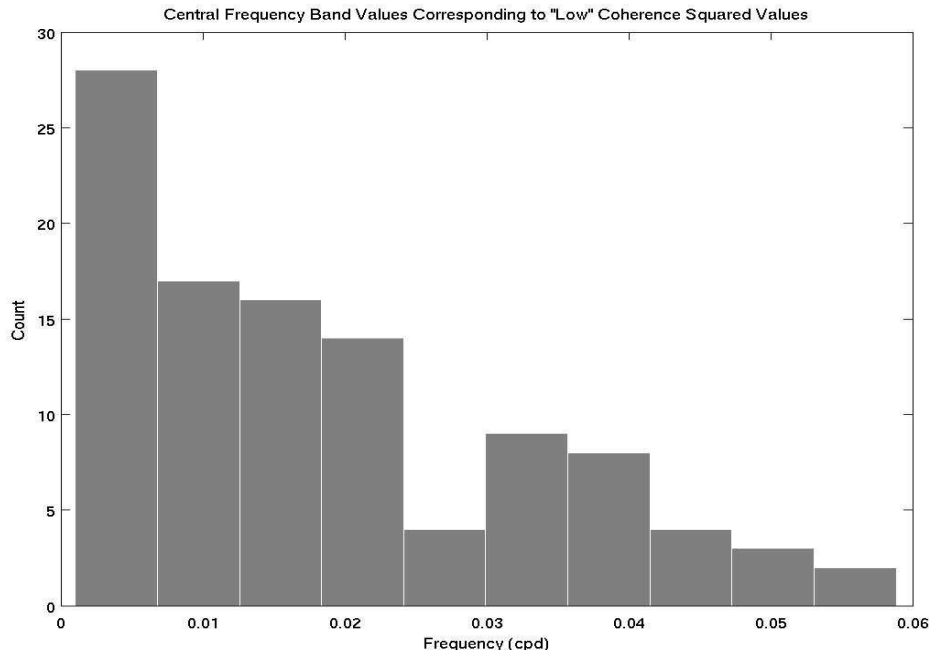


Figure 3-3. Distribution of the central frequency band value corresponding to low coherence squared values. The range of the most commonly identified frequencies include the annual and semi-annual frequencies.

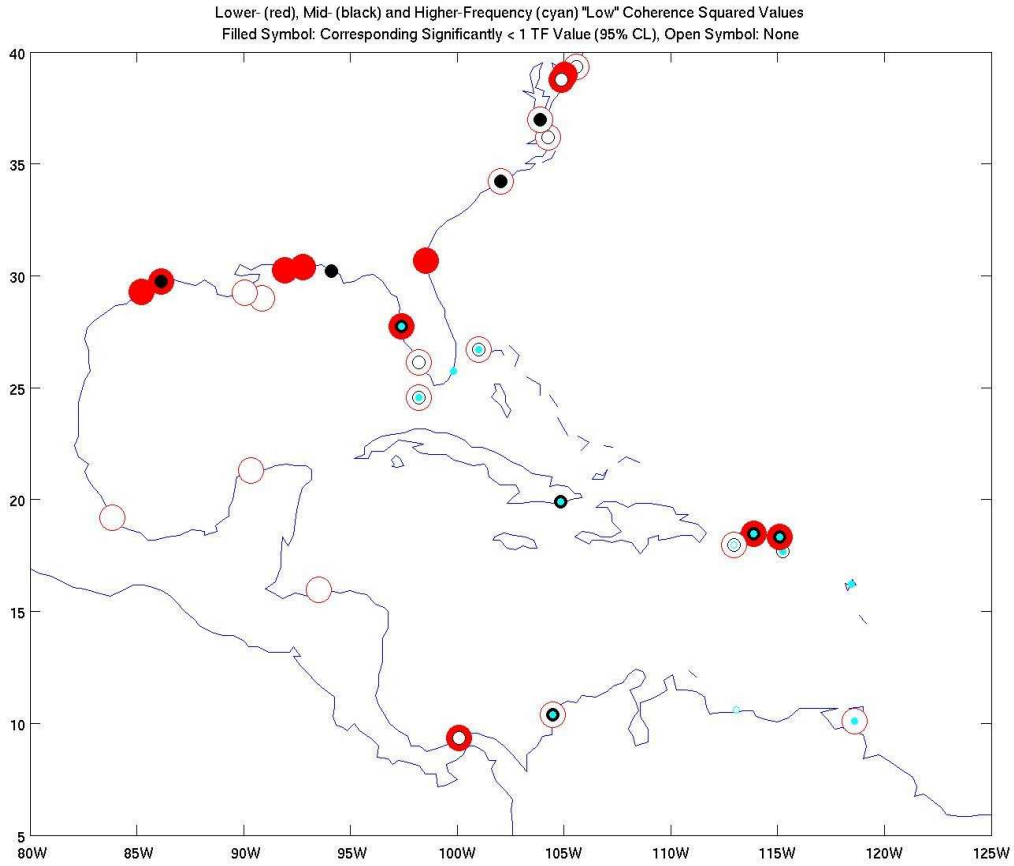


Figure 3-4. Map of relative low coherence squared value at comparison locations. Filled symbols indicate a corresponding transfer function value significantly less than 1 at the 95% confidence level was identified.

Since the AVISO product is derived from altimetry and altimeters have repeat periods too long to explicitly capture the ocean tides, a tide model is used to account for height fluctuations due to tides. The tidal energy will be aliased into the height signal from the altimeter. Table 3-1 gives 5 major tidal constituents and their frequencies along with the aliasing period for the altimeters used in the generation of AVISO. Their repeat periods are also noted for reference.

Table 3-1. Tidal Aliasing of 5 Major Constituents and Repeat Periods for Different Altimeters.

Tidal Constituent	Frequency (cpd)	Topex/Poseidon (days)	Geosat (days)	ERS-1 (days)
M2	1.93227	62	317	95
S2	2.00000	59	169	∞
N2	1.89598	50	52	97
O1	0.92954	46	113	75
K1	1.00274	173	175	365
Repeat Period (days)		9.91564	17.05050	35.00000

Figure 3-5 shows all of the frequencies associated with the low coherence squared values identified by the selection criteria of being below the 80% confidence level and at least half a smoothing width wide at all of the comparison locations. Because the 80% confidence level was used, many of the values will have been that low by random chance and are not meaningful. However, it is interesting to note that a sizable fraction (approximately 30%) of the frequencies correspond to the tidal aliasing frequencies for different altimeters. Figure 3-5 shows these values with the tidal aliasing frequencies in red. Considering all frequencies, most of the low coherence squared values are at relatively high frequencies with corresponding periods of less than 90 days. This means that most of the disagreement between AVISO and and the tide gauge sea level series is in these higher frequencies. Those locations where low coherence squared values can be accounted for with a tidal aliasing frequency are mapped out in Figure 3-6. The T/P aliasing frequencies are the most commonly recognized. ERS is largely confined to the Gulf of Mexico. Considering the spatial density of the tracks for each altimeter, ERS has a much higher concentration in the Gulf of Mexico than the other two missions. It is sensible to conclude that AVISO in the Gulf of Mexico is biased towards ERS data.

The number of times a spectral frequency is identified as having a low coherence squared value and a transfer function value significantly less than 1 (this time at the 80% confidence level) is plotted in Figure 3-7. It is suspected that when a model is applied to remove the tidal aliasing energy, some other energy captured by the altimeter at that frequency that is not related to the tidal aliasing is also being removed. The overestimation in the model used by AVISO is probably due to local differences in tidal fluctuations compared to the global model used, therefore it follows that neither the phase (from coherence squared) or variance (from the transfer function) information will agree with the tide gauge at that frequency. The tidal aliases for each altimeter are marked with color-coded dashed lines. In the range of the tidal aliases, many points fall on or near (within the band-smoothing width) those frequencies. The higher frequencies range over corresponding periods of 14 to 45 days. This monthly to sub-monthly domain is difficult for gridded height products like AVISO to reproduce. Even though the AVISO DM product "sampling" period is every 7 days, the time window used in processing is often longer, smoothing out features at frequencies lower than the 1/14 cpd Nyquist. There are also many physical phenomena that can occur in this frequency range that the tide gauge could record that the altimetry would never see. It is possible for local winds to set up over these timescales, changing water level recorded by the tide gauge. AVISO cannot represent this due to the spatial and temporal interpolation done in processing. The episodic passage of gravity waves produced by winds distant the tide gauge can also occur on these timescales and would not be captured by AVISO at any one particular location. As can be seen in the next section, this frequency range is mostly above the high-frequency roll-off for a majority of the comparison series.

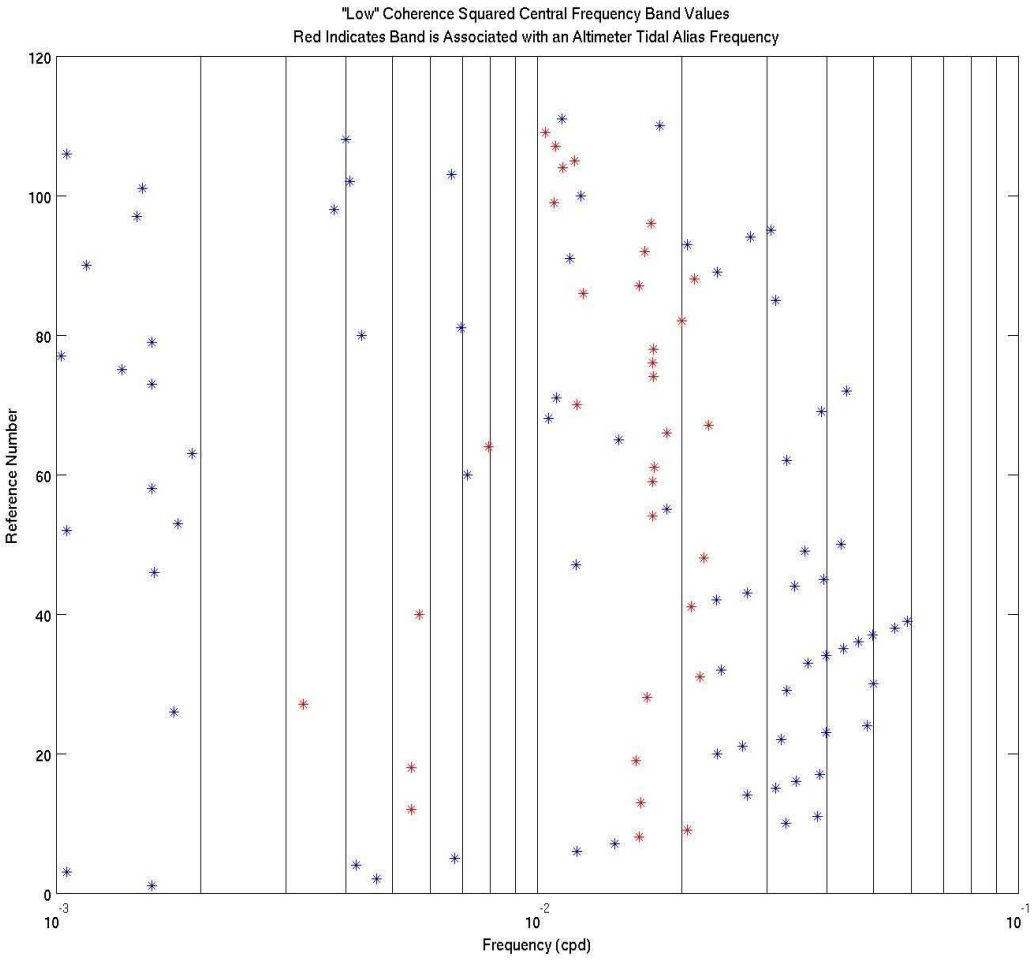


Figure 3-5. Central frequency band values for low coherence squared values. Red points indicate a direct match to an altimeter tidal aliasing frequency.

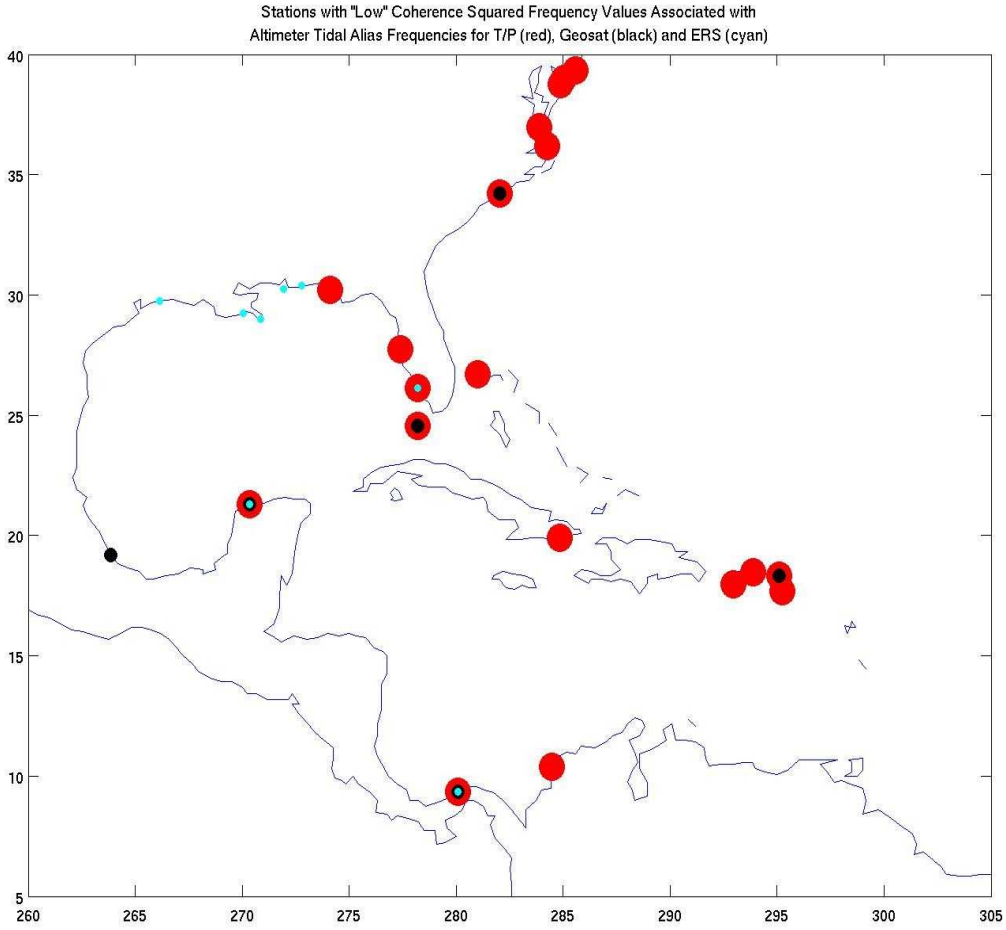


Figure 3-6. Comparison locations where low coherence squared value frequencies correspond to tidal aliasing frequencies for different altimeters. T/P (red), Geosat (black) and ERS (cyan).

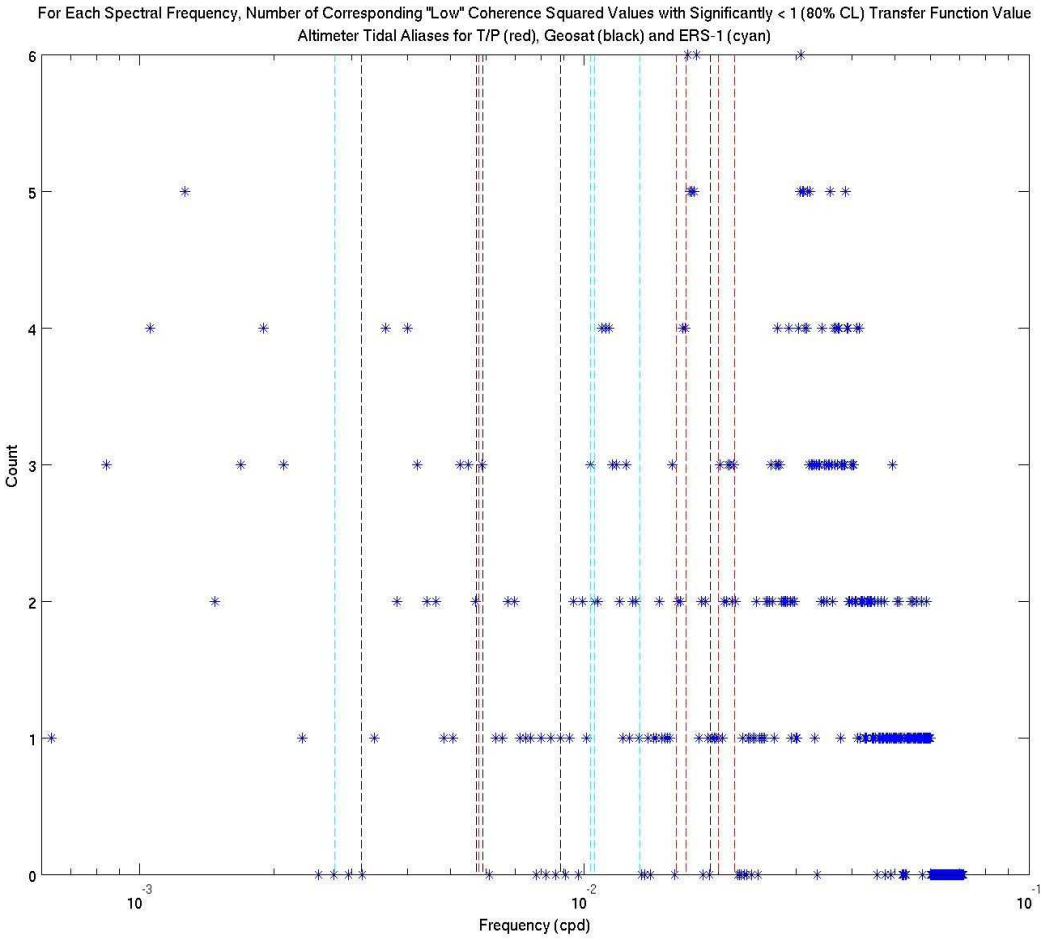


Figure 3-7. The number of times a spectral frequency is identified as having a low coherence squared value and a transfer function value significantly less than 1 (80% confidence level). Dashed lines indicate tidal alias frequencies for T/P (red), Geosat (black) and ERS (cyan).

3.4.3 Results from the Transfer Function

The HFRO values are identified as discussed in Section 3.3. Their distribution is shown in Figure 3-8a. This distribution is divided into thirds and mapped in Figure 3-8b. There appears to be a clear geographical dependence on HFRO value. This indicates that AVISO becomes deficient in energy (variance) at different frequencies depending on location, meaning the amount of effective smoothing in the AVISO product is not constant. This difference is an important result because height information at certain frequencies may or may not be able to be discerned from the AVISO product depending

on location. While it is not known what is causing these differences, it is speculated that the length scales set for smoothing in the creation of the AVISO product vary or perhaps the relative size of higher to lower frequency signal varies with location. This partly explains the pattern from the low coherence squared values where the higher-frequency values were only located in the Caribbean and the Florida peninsula. Because the roll-off point occurs at a higher frequency, the series can be better compared at relatively higher frequencies than at other locations.

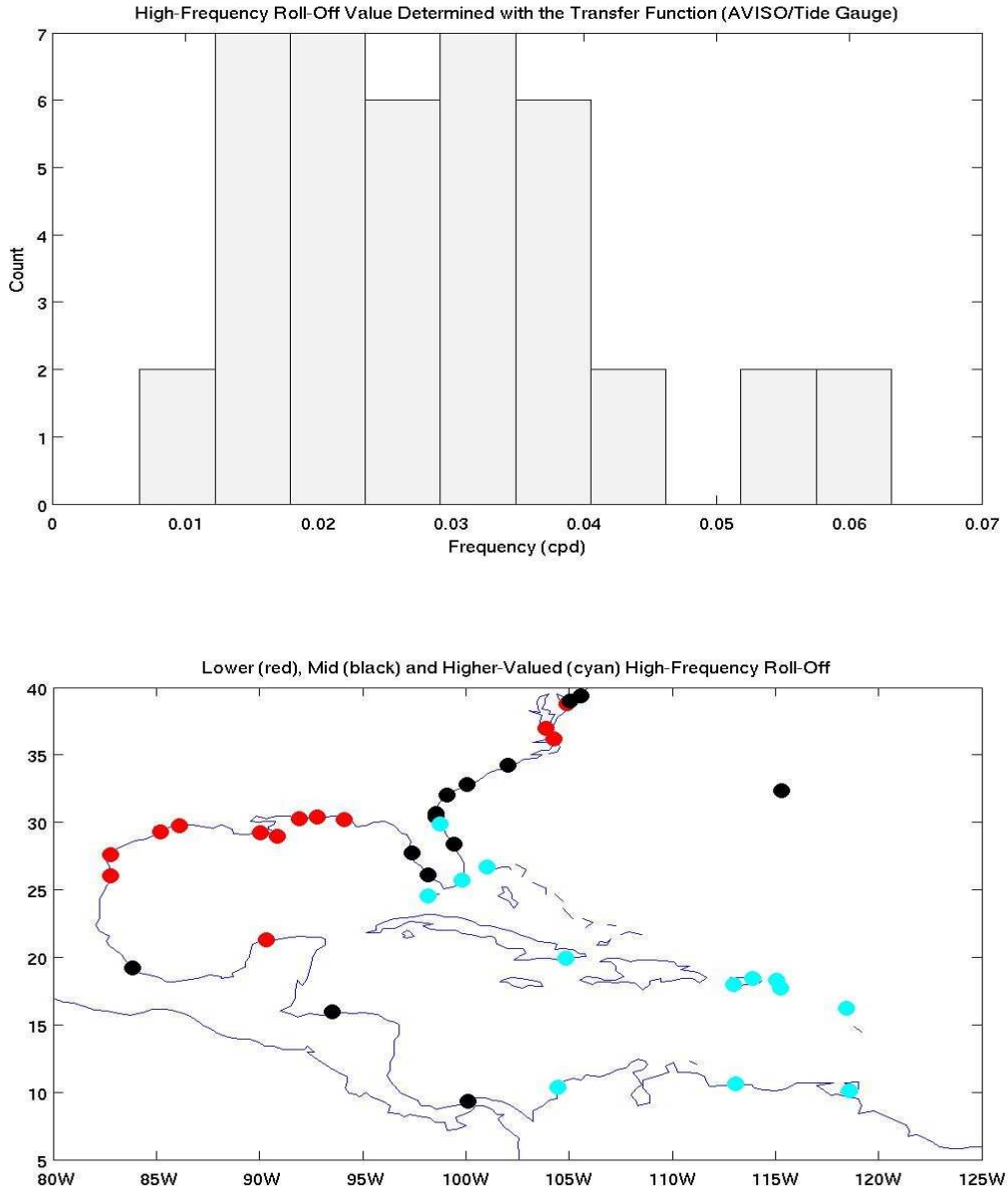


Figure 3-8. a. Distribution of high-frequency roll-off values determined from the transfer function (AVISO/tide gauge), b. Map of high-frequency roll-off values divided into thirds. Note the geographical distribution.

Series comparison plots showing examples of locations with lower, mid and higher-frequency HFRO values are shown in Figure 3-9a, b and c. Figure 3-9a is the lower-frequency HFRO value location of Sabine Pass, TX, Figure 3-9b is the mid-frequency HFRO value location of Port Canaveral, FL and Figure 3-9c is the higher-

frequency HFRO value location of San Juan, PR. Looking at the AVISO series (center plot on each figure), it is apparent that the AVISO series for Sabine Pass, TX is much smoother than the AVISO series for Port Canaveral, FL or San Juan, PR while the tide gauge series have a similar amount of high-frequency variability. This is more evidence that the amount of smoothing in AVISO varies geographically. Figure 3-9d, e and f are the close-up views of the AVISO and tide gauge time series for Sabine Pass, TX, Port Canaveral, FL and San Juan, PR, respectively. Note the height scale on the plot for San Juan is smaller than previous plots. At Sabine Pass, TX, AVISO captures variability on timescales down to about 4 weeks (about a month). For Port Canaveral, FL, the AVISO series obviously has higher frequency information in the heights, on timescales ranging from about 2 to 4 weeks. The HFRO value at San Juan is slightly higher than that at Port Canaveral and contains height variability down to about 2 weeks. For these examples, the time resolution on AVISO can be up to 2 times higher from one location to another.

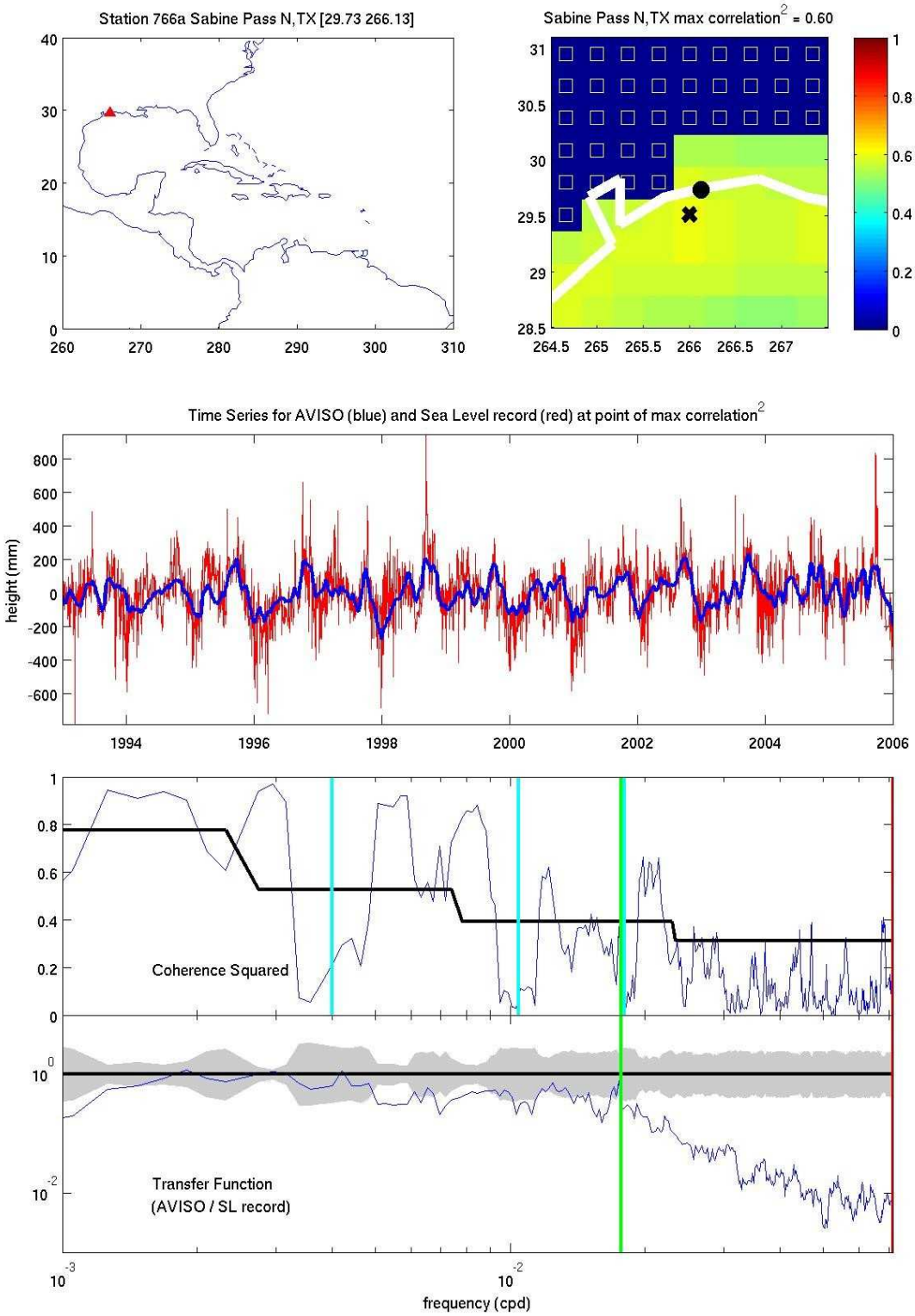


Figure 3-9a. Comparison location with lower-frequency high-frequency roll-off value. Note that the AVISO series for the lower-frequency location is much smoother than at the other series locations even though the tide gauges record a similar amount of high-frequency information at all locations.

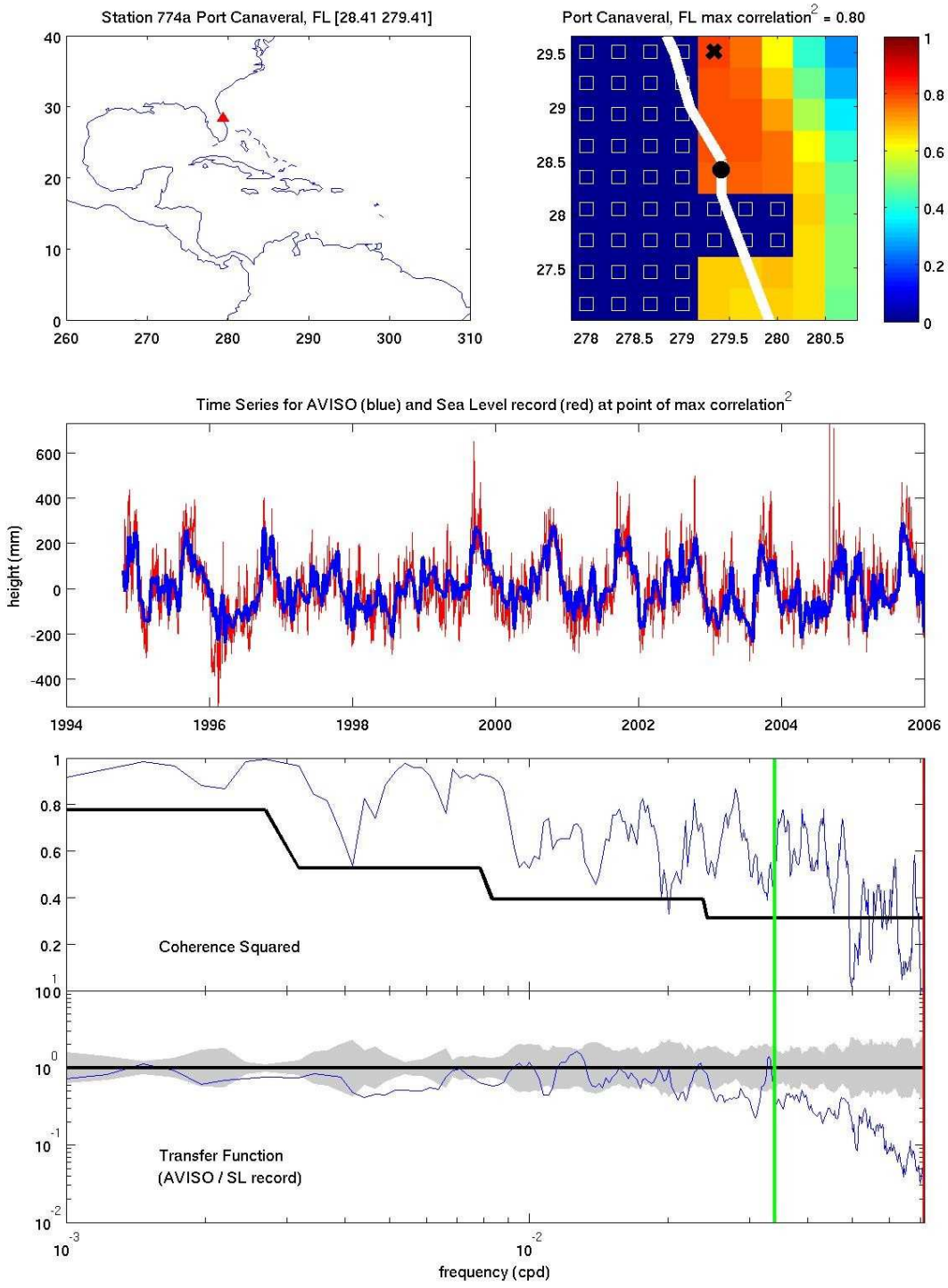


Figure 3-9b. Comparison location with mid-frequency high-frequency roll-off value. Note that the AVISO series for the lower-frequency location is much smoother than at the other series locations even though the tide gauges record a similar amount of high-frequency information at all locations.

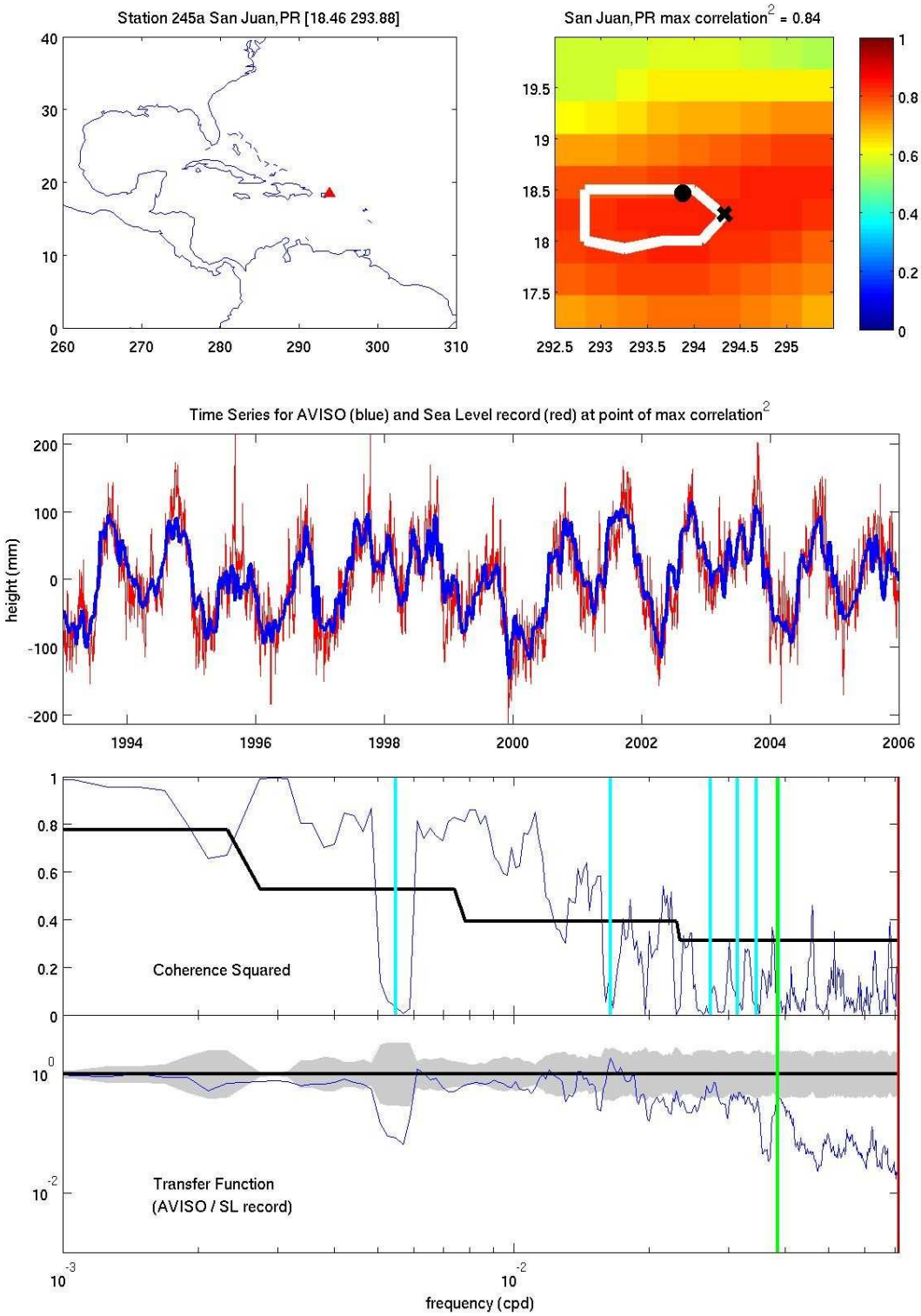


Figure 3-9c. Comparison location with higher-frequency high-frequency roll-off value. Note that the AVISO series for the lower-frequency location is much smoother than at the other series locations even though the tide gauges record a similar amount of high-frequency information at all locations.

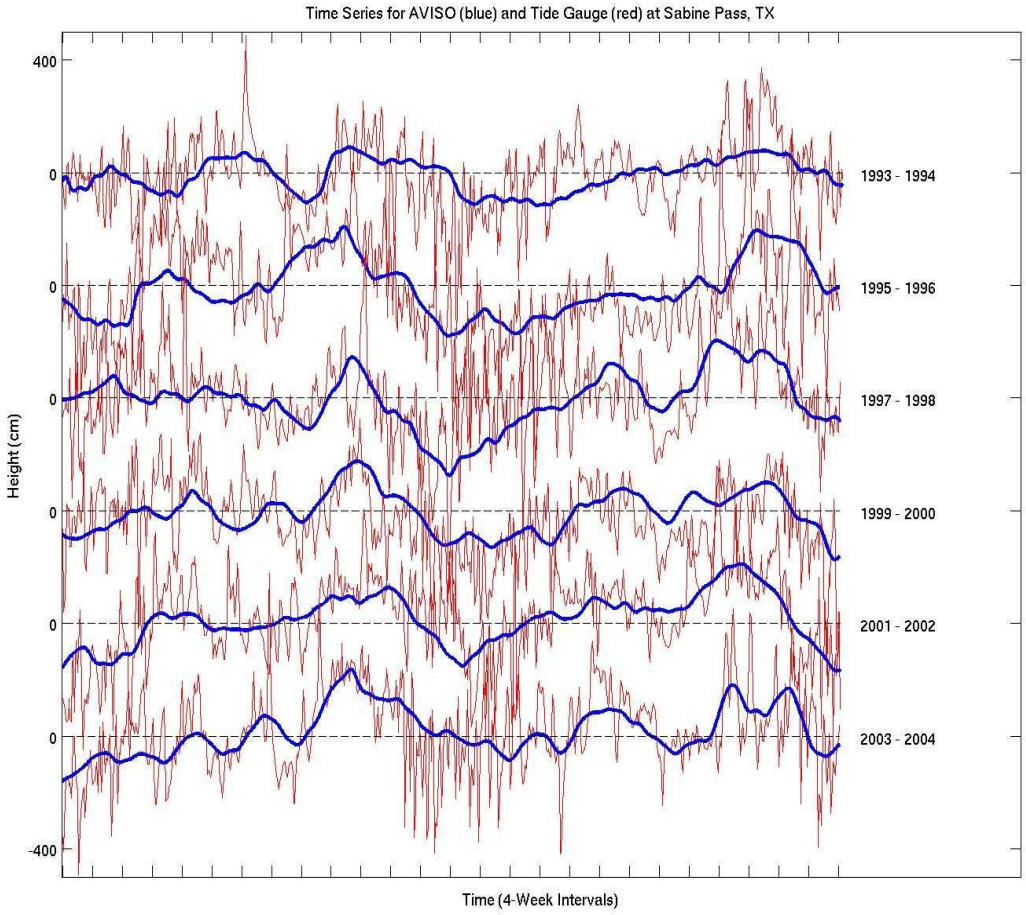


Figure 3-9d. Detailed view of heights series from AVISO and the tide gauge at Sabine Pass, TX. Variability is visible down to timescales on the order of about 4 weeks.

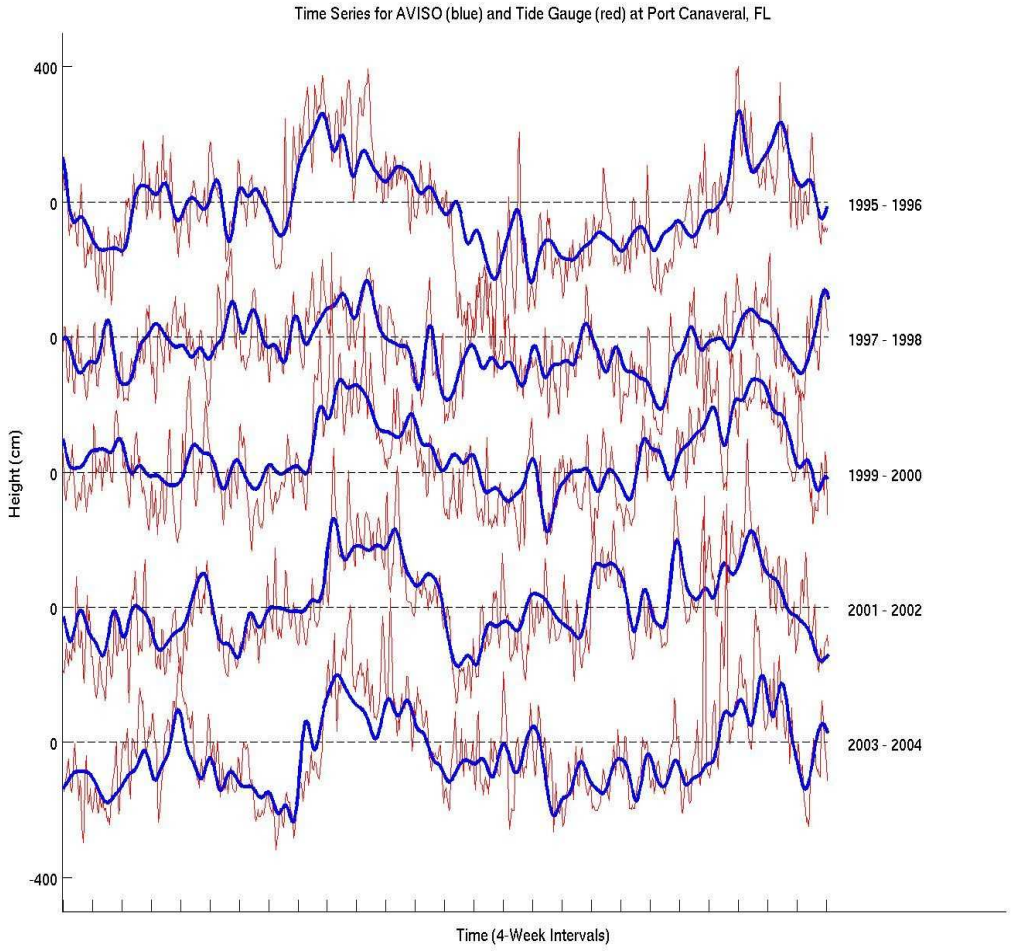


Figure 3-9e. Detailed view of heights series from AVISO and the tide gauge at Port Canaveral, FL. Notice the higher frequency variability in the AVISO series than at Sabine Pass, TX on the order of 2 to 4 weeks.

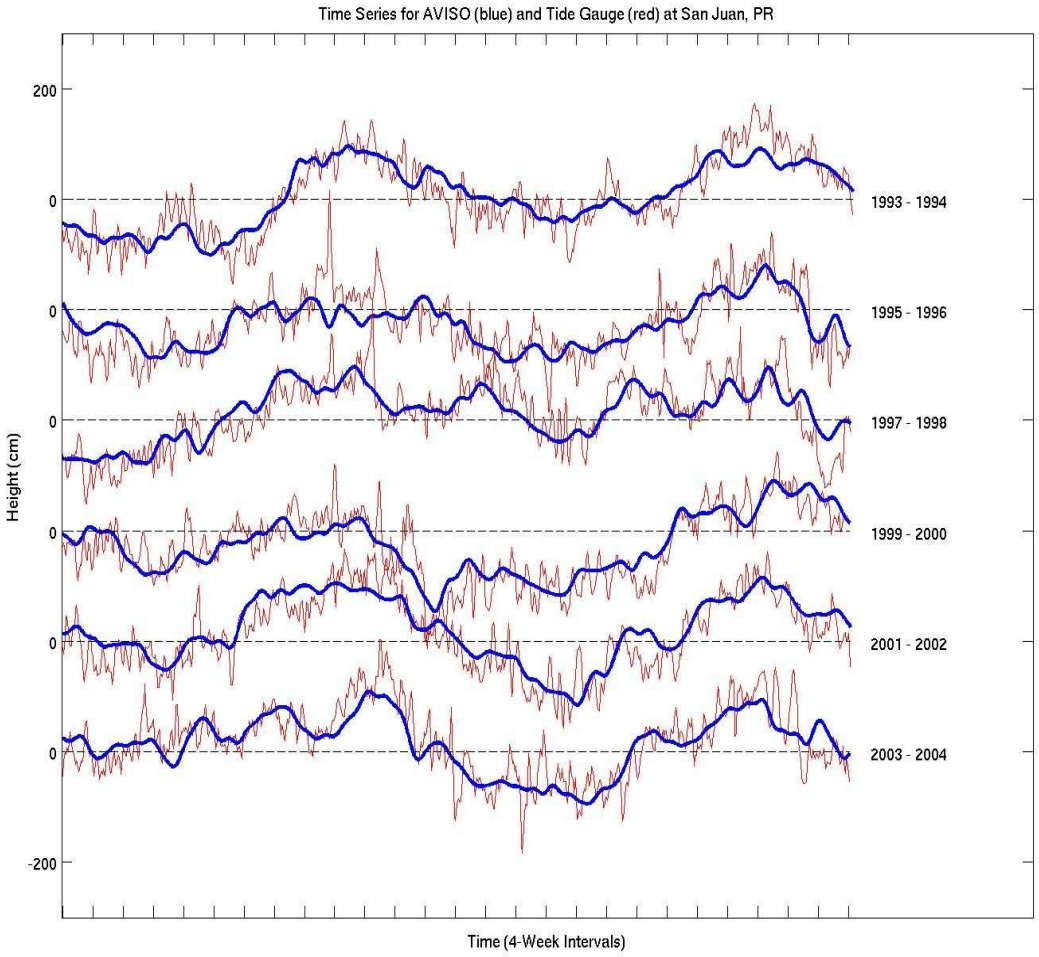


Figure 3-9f. Detailed view of the heights series from AVISO and the tide gauge at San Juan, PR. Note the change in the scale for the heights compared to previous figures. Shorter timescale features on the order of 2 weeks are visible.

3.5 Moored Velocity Comparison to OSCAR: Data and OSCAR Series Chosen for the Comparison

3.5.1 Data Sources

In-situ velocity data from moored Acoustic Doppler Current Profilers (ADCPs) used to compare to OSCAR velocity output are provided courtesy of the National Oceanic and Atmospheric Administration's National Data Buoy Center (NDBC) and the University of South Florida's (USF) Ocean Circulation Group (OCG). ADCP measurements are provided in “u” (zonal) and “v” (meridional) components of ocean current velocities at various water depths. The NDBC data used comes from the formerly-named Minerals Management Service's (MMS) program to collect ADCP data on oil-drilling rigs in the Gulf of Mexico. Any available data falling within the OSCAR time period October 1992 through 2008 were identified and downloaded from the MMS ADCP website at http://www.ndbc.noaa.gov/. Each station is assigned a Station ID—a five-character World Meteorological Organization station identifier that is assigned to all weather observation stations, including ADCPs. The OCG data comes from buoys operated through the USF College of Marine Science and was accessed through the OCG website, http://ocgweb.marine.usf.edu/. Length of the time series and temporal resolution varies greatly from station to station. It should be noted that not all ADCP instruments are the same. Any stations with less than one year of concurrent data with OSCAR were excluded from the comparison with OSCAR. Figure 3-10 maps the locations of the stations used in the comparison.

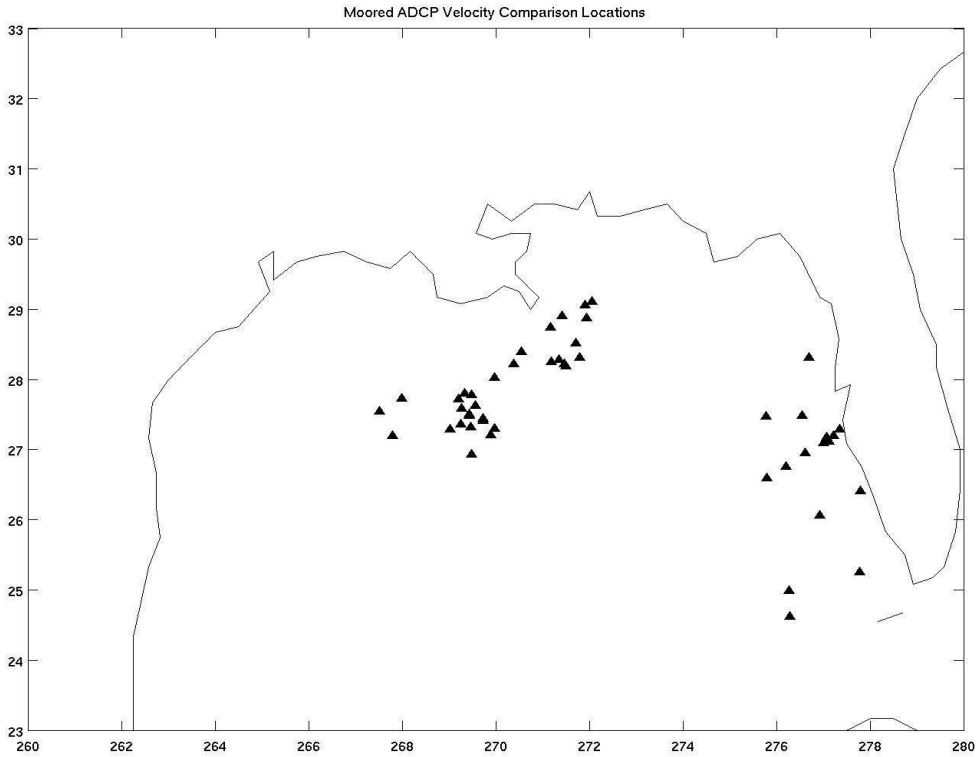


Figure 3-10. All moored ADCP data locations used in the comparison calculations with OSCAR. Data sources are the formerly-named Minerals Management Service and the Ocean Circulation Group at the University of South Florida College of Marine Science. To be included in the comparison, at least one year of concurrent data with OSCAR at any location is required.

3.5.2 Matching Data and OSCAR Series

In order to perform the comparison calculations, it is necessary to standardize the data and OSCAR to the same temporal resolution and to determine which OSCAR location best matches the data location. The latter is not straightforward, as there is some spatial smoothing done in creating a gridded product like OSCAR that could distort the OSCAR grid location compared to the latitude and longitude location of the data. To make the times match, the data series are checked for regular time intervals and smoothed to daily using box averaging. OSCAR is interpolated from every 5 days to daily using linear 3-dimensional (latitude, longitude and time) interpolation. To locate the best OSCAR/data series pair, the vector correlation squared of Crosby et al. 1993 (explained

in Section 3.6) is computed for the four OSCAR series nearest to the data location. The series leading to the highest correlation value is chosen for the comparison.

3.6 Moored Velocity Comparison to OSCAR: Methods Used for Comparison

3.6.1 Vector Correlation Parameter

Crosby et al. 1993 gives a definition for a vector correlation parameter that is of particular use in meteorology and oceanography. Vector correlation is not straightforward because vectors have both magnitude and direction as opposed to a scalar quantity. Correlation of scalar vector components can be computed using the traditional Pearson correlation coefficient commonly represented by "r", but the correlation is not unique since it is dependent on the choice of coordinate system. At the time of the publication of Crosby et al. 1993, there was no single definition in use for vector correlation and this provided the motivation for the derivation of this particular vector correlation parameter.

The basic properties of the vector are that it is in a Cartesian coordinate system with the positive u-component in the positive x-direction (east) and positive v-component in the positive y-direction (north). The sample covariance matrix is comprised of the auto and cross covariances of all combinations of u and v in the usual way such that the trace of that matrix is the sum of the variances of the u and v components. The correlation coefficient is invariant under linear transformations and is the ratio of the sample covariance and the u and v component standard deviations.

To compare two two-dimensional random vectors $\mathbf{W}_1$ and $\mathbf{W}_2$ with u and v components, define

$$\boldsymbol{X}=\begin{pmatrix}\boldsymbol{W_1}\\ \boldsymbol{W_2}\end{pmatrix}=\begin{pmatrix}u_1\\ v_1\\ u_2\\ v_2\end{pmatrix}$$

and

$$\Sigma_X=\begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix}$$

where each $\Sigma_{jk}(j,k=1,2)$ is the centered covariance matrix for each pair of the four vector components.

Early definitions of vector correlation include that of Sverdrup (1917) with

$$\rho_S^2=\frac{Tr(\Sigma_{12})^2+[\sigma(u_1,v_2)-\sigma(u_2,v_1)]^2}{Tr(\Sigma_{11})Tr(\Sigma_{22})}$$

and Kundu (1976) with

$$\rho_T^2=\frac{Tr(\Sigma_{12})+i[\sigma(u_1,v_2)-\sigma(u_2,v_1)]}{[Tr(\Sigma_{11})Tr(\Sigma_{22})]^{1/2}},$$

the complex vector correlation coefficient. For both formulas, the Σ_{jk} are as defined above, Tr indicates the trace of the matrix, $\sigma(u_j,v_k)$ are the cross covariances and i is $(-1)^{1/2}$. Sverdrup's definition was used by meteorologists in the mid-20th century and the two are related to each other by $|\rho_T|^2=\rho_S$. Crosby et al. 1993 defines the vector correlation parameter as

$$\rho_v^2=Tr[(\Sigma_{11})^{-1}\Sigma_{12}(\Sigma_{22})^{-1}\Sigma_{21}]$$

and is based on earlier work by authors including Hotelling (1936), Hooper (1959) and Jupp and Mardia (1980). Generally for the two-dimensional case, ρ_v^2 ranges between 0 and 2, and must be normalized (by dividing by a constant value, 2 in this case) to have

the vector correlation parameter range from 0 to 1 in the usual way. It is a generalization of the standard Pearson correlation coefficient for a pair of one-dimensional variables and includes both magnitude and direction. This vector correlation parameter has many desirable properties such as being invariant under transformations of the coordinate axis including rotations and changes in scale and it is equal to zero when the vectors are independent and is a maximum if and only if the vectors are linearly dependent such that one vector can be expressed as a linear combination of the other. Perfect correlation results in the following cases for two two-dimensional vectors:

- The vectors are identical
- The magnitude of one vector is equal to the magnitude of the other vector times a constant
- The direction of one vector is is equal to the direction of the other vector rotated by a constant angle
- The magnitude and direction of one vector are equal to the magnitude and direction of the other vector times a constant and rotated by a constant angle, respectively

It should be noted the authors demonstrate that small sample sizes (of order 10) can result in relatively high (>0.6) vector correlation values. For the moored velocity to OSCAR output comparisons, no series shorter than 365 days were used, so this problem has been avoided.

3.6.2 Rotary Spectra

Since the vector correlation parameter can be ambiguous in cases where there are constant differences is magnitude or constant rotation between two series, a rotary spectral analysis is performed. The method used is that laid out by Mooers 1973,

rehashed in Emery and Thomson's *Data Analysis Methods in Physical Oceanography*. Rotary spectral analysis is applied to vector time series like ocean currents and winds with zonal (u) and meridional (v) components, expressed as a complex quantity, for example, $w = u + iv$. It is assumed that u and v are continuous and stationary and have zero-mean. It is also assumed that u and v have Fourier integral representations. The variability of the two-dimensional velocity in time can be completely described by the auto-spectra and the cross spectrum. It provides a statistical measure of the coherence and phase as a function of the frequency. For each frequency, the spectrum for the velocity can be decomposed into two counter-rotating, circular components, each with amplitude and phase. The spectrum is two-sided and so is defined for positive and negative frequencies. The anti-clockwise (ACW) component is defined for positive frequencies and the clockwise (CW) for negative frequencies. The orientations are paired with these particular signs of frequency to match with what is considered positive and negative rotation in physics. When the components are added together, the resulting vector traces an ellipse that has some angle of orientation with respect to the x-y plane. The rotation sense of the vector around the ellipse is determined by which component (ACW or CW) is larger in magnitude. This ellipse represents the behavior of the vector over time. The Mooers 1973 method makes improvements over previously presented methods in that the coherence is independent of the coordinate system in which the observations are made (because the circular components have circular symmetry)—very helpful in oceanography—and the coherence and phase for polarized components is more easily interpreted with some physical meaning. The spectra are computed by starting with the discrete Fourier transform of the complex velocity vector w, of the form

$$W(f_k) = \Delta t \sum_{n=0}^{N-1} w(t) e^{-2\pi i k_n / N} = U(f_k) + V(f_k)$$

where U and V are the Fourier transforms of u and v, $f_k = k/N\Delta t$ is the kth Fourier frequency and Δt is the time interval of w. There are four coherence and phase functions (from cross-spectra) corresponding to all combinations of the circular components: 1^+2^+, 1^-2^-, 1^+2^- and 1^-2^+ where 1 and 2 are the first and second of two series being compared and + and – indicate the sign of the circular component.

Two types of spectra—the "inner" and "outer"—are computed. The terminology of inner and outer spectra was chosen by Mooers 1973 because the rotary autocovariance functions derived from the autospectra resemble the general mathematical definitions of inner (dot) and outer (cross) products. The spectral formula is written as

$$S_{w_j w_k} = \langle W_j^* W_k \rangle$$

where $j,k = 1,2$, $W_{j,k}$ are the Fourier transforms of series 1 and 2 and $*$ indicates and complex conjugate of the complex series. All are functions of frequency. Inner spectral quantities are those computed for the co-rotating (+ + and - -) pairs of series 1 and 2 and the outer are for the counter-rotating (- + and + -) pairs of series 1 and 2. The angled brackets indicate that the ensemble average (like N-band smoothing) is taken to compute the spectrum from the periodogram. When j=k, it is the inner or outer auto spectrum for series 1 or series 2 and when $j \neq k$, it is the inner or outer cross spectrum. The coherence squared is given by

$$\gamma_{12}^2 = \frac{\left| S_{w_1 w_2} \right|^2}{\left(S_{w_1 w_1} \right)\left(S_{w_2 w_2} \right)}$$

and the phase is

$$\chi_{12} = \arctan\left(\frac{-\Im(S_{w_1 w_2})}{\Re(S_{w_1 w_2})}\right)$$

where the script I and R indicate the imaginary and real parts of the cross spectrum, respectively. Both the coherence squared and phase have values at each frequency. For the all spectral quantities, the + + and - + pairs are defined for positive frequencies and - - and + - pairs are defined for negative frequencies.

The inner auto-spectra are an indication of the amount of energy contained in the series at each frequency for the ACW and CW components. When the inner auto-spectra for two series are considered, it is possible to see any differences in frequency space such as which rotational component is dominant and how the energy recorded by each series compares. The outer auto-spectra are explained by Mooers 1973 to be related to the uv-Reynolds stress, an analysis that is beyond the scope of this comparison. The coherence squared must have a confidence level placed on the values to indicate how significant the values may be. If the coherence squared is near zero, the vectors are not related and if the coherence squared approaches 1, they are in agreement. The phase indicates how much the given component of series 1 leads series 2. If the two series being compared are expected to (or should be) the same, then the phases corresponding to frequencies with significant coherence values should be equal to zero.

3.7 Moored Velocity Comparison to OSCAR: Comparison Results

3.7.1 Preliminary Evaluation of Available Data with Vector Correlation

The data is evaluated to ascertain some of its general properties in comparison with OSCAR before a more in-depth method is used. The two data sources have time series that are measured at different depths. The OCG data are mostly in the shallower

depth bins, while the MMS data is mostly in the deeper depth bins. The distributions of all available individual data points are shown in Figure 3-11. Since OSCAR output velocities represent near-surface currents at 15-meters depth, it must be determined whether it is appropriate to compare velocity taken at different depths. Figure 3-12 shows the vector correlation squared (Crosby et al. 1993, detailed in Section 3.6) for each comparison location over all available depth bins. With the exception of one location, it is apparent that depth does not greatly change the vector correlation between the data at various depths and OSCAR series. This is not surprising, since this area of the Gulf of Mexico tends not to have strong horizontal density gradients resulting in little vertical shear over the upper 100 meters. The lack of dependence on the depth of the measurement also implies that in a time-mean sense, the Ekman contribution is relatively small in this area. It is found that this is indeed true and will be discussed in a later section (Section 4.3.2). Each series is compared with OSCAR, regardless of depth.

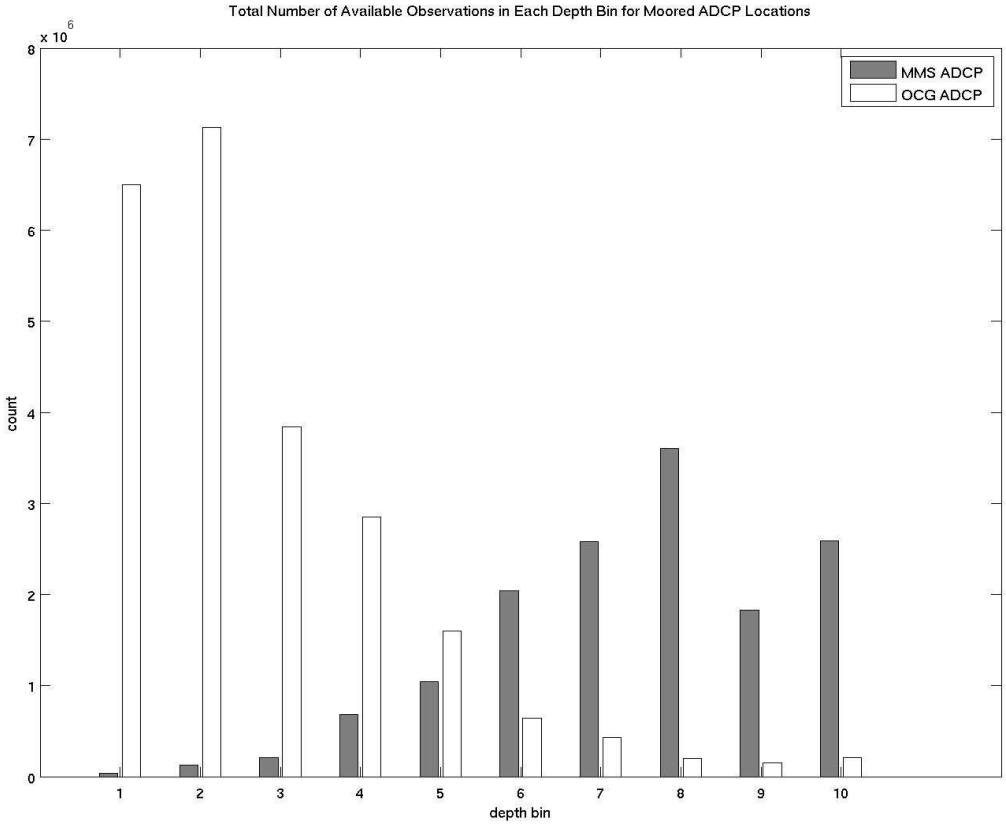


Figure 3-11. All available time and space data points from the MMS and OCG ADCP. Note that not all points were used in the comparison to OSCAR.

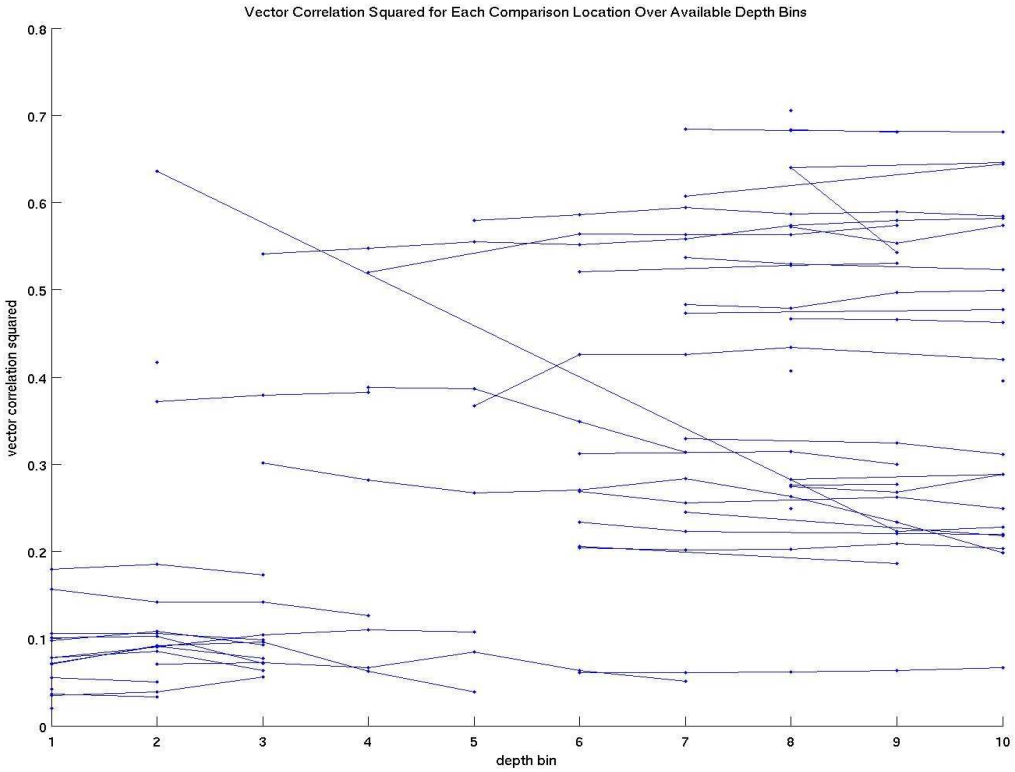


Figure 3-12. Vector correlation squared between data and OSCAR and depth bin of the data measurement. Most locations show little dependence on depth of the measurement with correlation due to the absence of strong shear in the upper 100 meters in this area of the Gulf of Mexico.

The data are distributed in areas with a wide range of water depths and distance to the nearest land. Looking at a map of the distribution of vector correlation squared values for the moored ADCP data locations (not shown) suggested there may be some dependence on water depth such that lower correlation values tended to be in shallow water. Since one of the goals of this analysis is to identify potential problems with OSCAR in the IAS, the lower-correlated locations are of particular interest. The vector correlation squared is plotted with water depth in Figure 3-13. Note this is not the depth of the velocity measurement, but the total water depth in which the ADCP is located. The plot shows that virtually all of the low values for the vector correlation squared clustered at less than ~ 0.15 occur in shallow water and not in deeper depths. Shallow water areas

have many physical phenomena (like tides and shelf dynamics) varying over frequencies much too high for altimetry to capture and therefore are not expected to be well-represented in OSCAR. In addition, altimetry does not do well near to land due to inaccuracies introduced by the radiometer used to make the water vapor correction to the height data. Figure 3-14 shows the same vector correlation squared values with distance to the nearest land. Near to land, within approximately half a degree or about 50 kilometers (consistent with the scale of the radiometer footprint), all of the vector correlation squared values are low. However, not all of the low correlation values from Figure 3-13 are accounted for by being in shallow water, suggesting the lack of correlation between the ADCP data and OSCAR at those locations may be attributable to higher-frequency signals on the shelf not captured by altimetry.

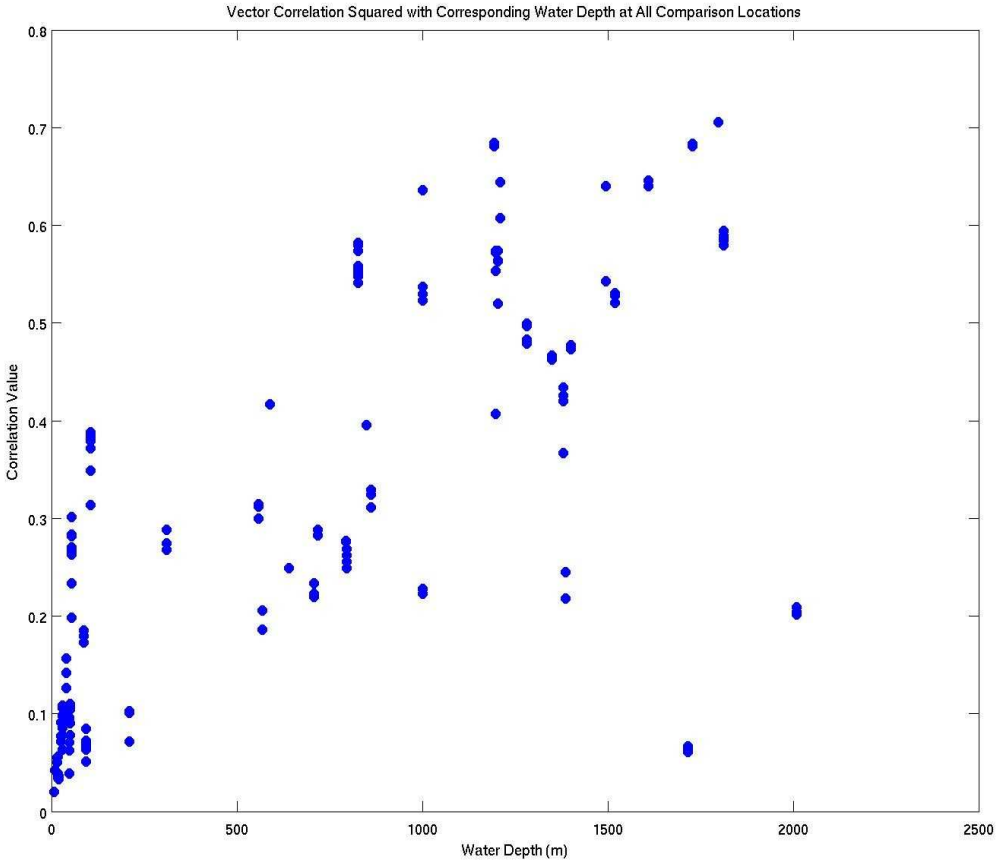


Figure 3-13. Vector correlation squared between data and OSCAR and water depth in which the instrument is located. There is a positive relationship—deeper water depth gives higher correlation.

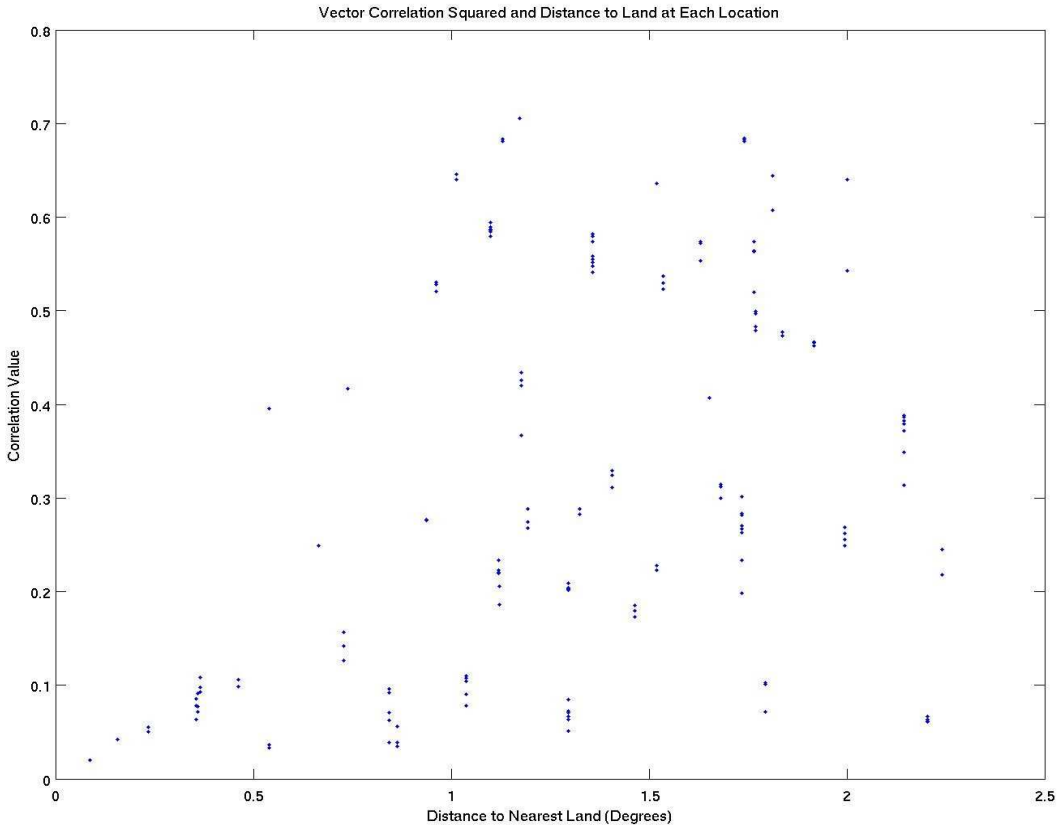


Figure 3-14. Vector correlation squared between data and OSCAR and distance to the nearest land from each comparison location. There is a positive relationship though it is widely scattered for distances greater than 1 degree.

3.7.2 Rotary Spectral Analysis

Three parameters from the rotary spectral calculations are utilized in evaluating OSCAR velocity output against moored ADCP velocity measurements. The first, the high-frequency roll-off (HFRO) value, comes from the ACW and CW inner auto-spectra. The HFRO is the frequency at which the amount of energy (the spectral value) in the OSCAR output velocities drops off compared to the previous spectral values at lower frequencies. This is expected in OSCAR because there is a certain amount of smoothing inherent in the process of creating a gridded product like OSCAR and it is not necessarily uniform. The HFRO is empirically defined as the frequency at which the ratio of the OSCAR and data auto-spectra (OSCAR/data) drops 10% from the median value at all

lower frequencies. This percentage may not seem very large, but in logarithmic space it is an appreciable drop. Figures 3-15a and 3-15b show the geographical distribution of HFRO values for the ACW and CW spectral ratios. Values are divided into three categories—low, mid and high. Low values are frequency values less or equal to 1/35 cycles per day (cpd), mid are between 1/35 and 1/25 cpd and high are greater than 1/25 cpd. These threshold values are chosen from the distribution of all of the HFRO values. Bins of 1/45 to 1/35 cpd, 1/35 to 1/25 cpd and 1/45 to 1/35 cpd include most of the values. The geographical distribution of HFRO values for ACW and CW spectral ratios appear roughly similar upon visual inspection. For the available depth bins, the locations in the eastern Gulf of Mexico have mostly higher-frequency values while the northern Gulf of Mexico locations have most of the lower-frequency values (though all values are present in the western locations). The distributions of the HFRO values divided into separate geographical regions of the eastern and northern Gulf of Mexico are shown in Figure 3-16. While the locations in the northern Gulf of Mexico have a range of values, it is apparent that the eastern Gulf of Mexico only contains the higher-frequency values. This could mean that the amount of smoothing in OSCAR is more constant over the West Florida Shelf. It would make sense for the gridding algorithm for altimeter data (on which OSCAR is largely based) to treat this area in a special way due to it being both a shallow water region and near to land, where altimetry has the most difficulty reproducing height. HFRO values have also been computed for the altimetry to tide gauge comparison. Figure 3-8b shows the geographical distribution with the same low, mid and high designations and Figure 3-8a shows the distribution of HFRO values. The values themselves have a range and shape of distribution similar to those in the OSCAR

to moored velocity comparison and also have a geographical dependence. Though the values do not necessarily agree exactly, there is a similar pattern of lower-frequency HFRO values in the northern Gulf of Mexico, higher ones around Florida and the highest-frequency values farthest to the West. The important point here is that both comparisons suggest there is some spatial difference in the amount of smoothing in OSCAR for the IAS.

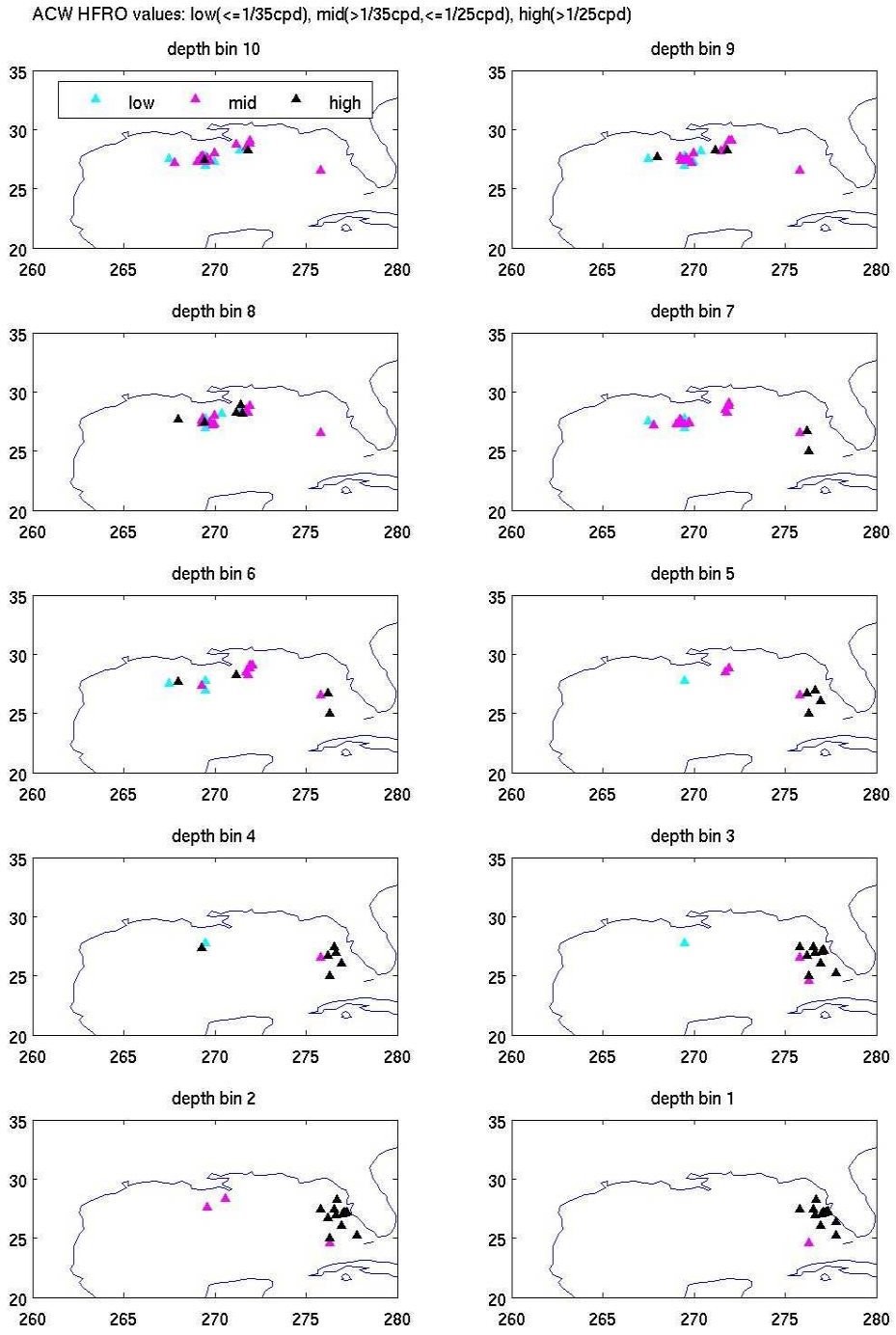


Figure 3-15a. Geographical distribution of high-frequency roll-off (HFRO) values for the anti-clockwise (ACW) auto-spectral ratios. Values are divided into three categories—low, mid and high. Low values are frequency values less or equal to 1/35 cycles per day (cpd), mid are between 1/35 and 1/25 cpd and high are greater than 1/25 cpd. There appears to be an east-west difference in the distribution of values.

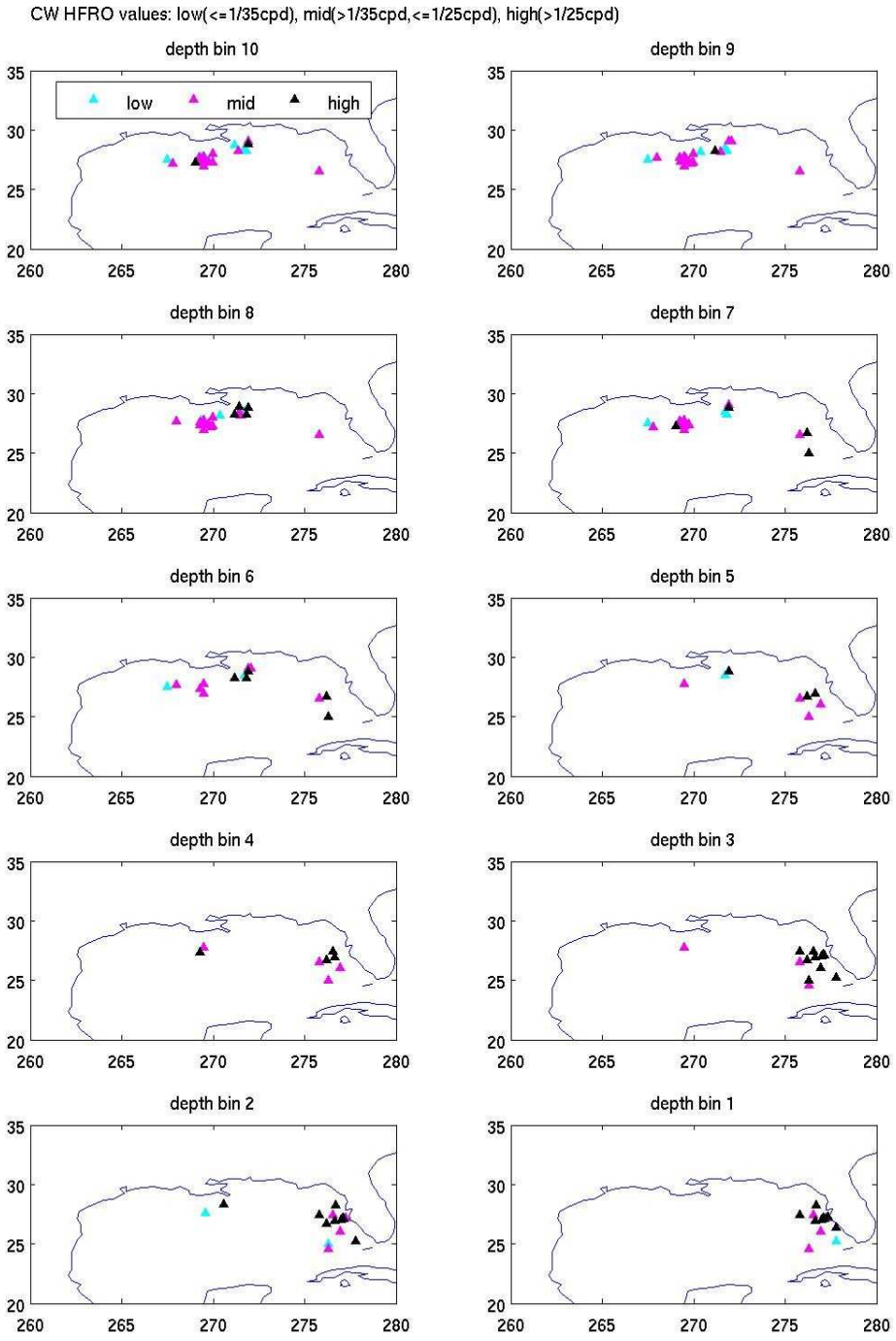


Figure 3-15b. Geographical distribution of high-frequency roll-off (HFRO) values for the clockwise (CW) auto-spectral ratios. Values are divided into three categories—low, mid and high. Low values are frequency values less or equal to 1/35 cycles per day (cpd), mid are between 1/35 and 1/25 cpd and high are greater than 1/25 cpd. There appears to be an east-west difference in the distribution of values.

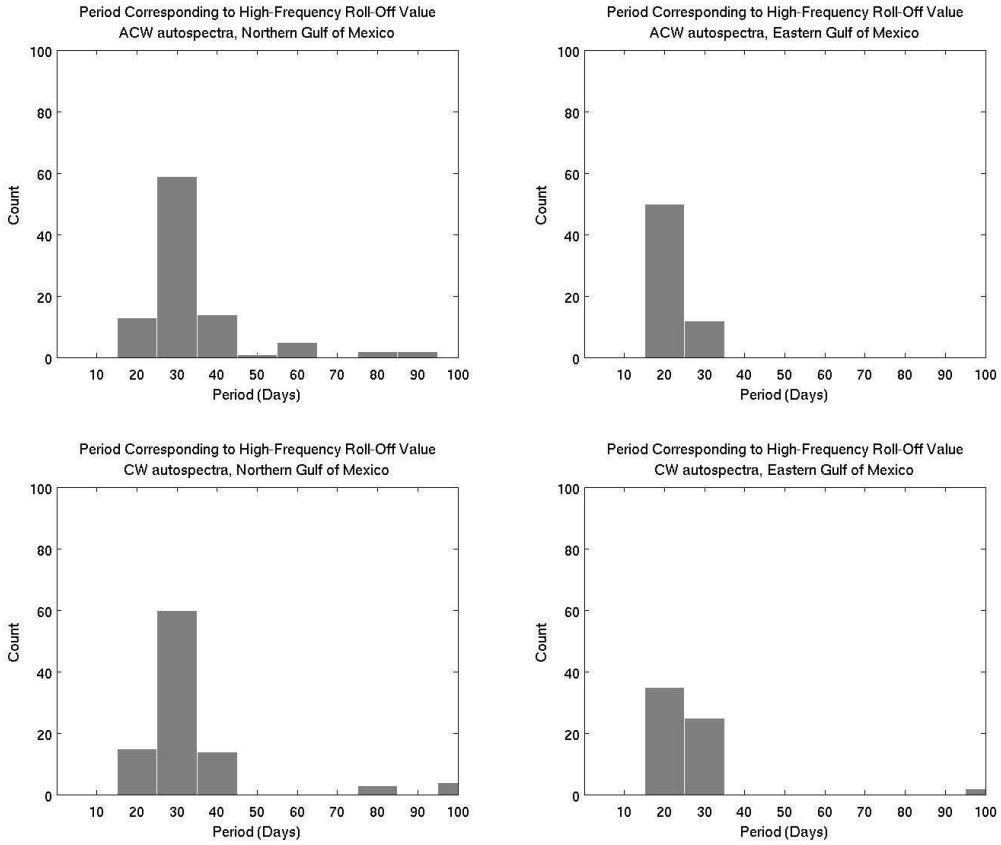


Figure 3-16. Distributions of the high-frequency roll-off (HFRO) with the period (in days) corresponding to the HFRO value frequency. Values are divided into separate geographical regions of the eastern and northern Gulf of Mexico. Locations in the northern Gulf have a range of HFRO values, but the eastern Gulf is restricted to the higher-frequency values.

Next, the coherence squared and phase values from the comparison are considered. A confidence level of 90% is chosen for the evaluation. All coherence squared values above the 90% confidence level and their corresponding phase values are selected from all four pairs of the inner and outer coherence squared and phase (cross-spectral quantities, + +, - -, - + and + - pairs). Maps of significant coherence squared and phase values for all available depth bins (not included) show no geographical dependence for any of the four spectral pairs. Plots of significant coherence squared, phase and frequency values versus depth bin of the velocity measurement (not included) show no dependence on the depth of the measurement when compared to the OSCAR velocities.

The distributions of the coherence squared, phase and frequency values for ++ and -- pairs are shown in Figure 3-17a (+ +) and 3-17b (- -). All values included in the distributions are the coherence squared significant at the 90% confidence level and the corresponding phases and frequencies. The + + and - - pairs show similar distributions for all three values shown. Coherence squared values range from 0.45 to 0.95, all relatively high for this type of data. The phase values are centered on 0 degrees. This is the desired result for series that are supposed to be the same. Most often for the + + and - - pairs, the OSCAR and in-situ velocity series are coherent over lower frequencies that are monthly and longer. For the - + and + - pairs (not shown), the distributions are also similar to each other. The coherence squared values have a similar range and shape of distribution to that of the + + and - - pairs, but there are no preferred phase or frequency values corresponding to them (i.e. their distributions are flat). This could indicate that the coherence squared values identified at the 90% confidence level as significant occurred randomly and have no physical meaning. These results make sense since the + + and - - pairs are for co-rotating components and the - + and + - pairs are for counter-rotating components. It is not expected that there would be coherence between the oppositely-oriented components for two time series representing the same data.

An example of the OSCAR and moored ADCP velocity comparison for a particular location is shown in Figure 3-17c and Figure 3-17d. Figure 3-17c shows the speed and direction for the data series (red) and OSCAR (blue). Current speed in cm/s is shown in the top panel. OSCAR tracks the patterns in the measured speed well on timescales that are seasonal and longer though it is consistently underestimating the speed when the measured speed is larger than about 0.5 cm/s. Current direction is shown in the

bottom panel. Though there is a lot more scatter in the data since it records more high frequency information than OSCAR, the general patterns agree fairly well. Figure 3-17d shows the set of plots used for the comparisons with rotary spectra. A map of the comparison location is in the top left, the CW and ACW autospectra in the top right and coherence squared and phase for each co-rotating and counter-rotating pair in the bottom panels. You can see that the co-rotating components (++ and - - pairs) have significant coherence squared values over a wide range of longer frequencies and agreement in the monthly band for the - - pair around zero phase. As described above, the combination of significantly high coherence squared and zero phase between the series means that they agree well.

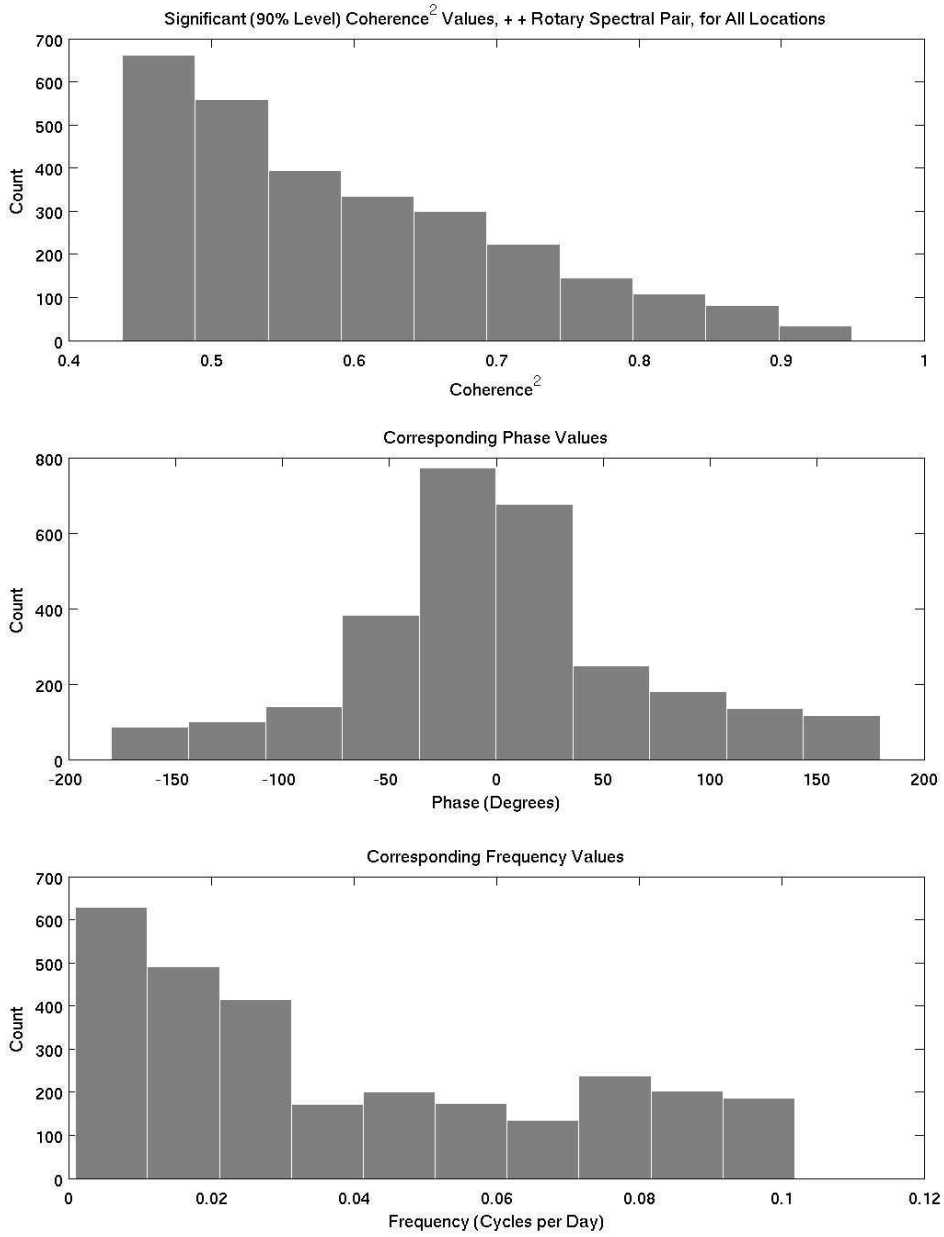


Figure 3-17a. Distributions of the significant (at the 90% level) coherence squared, phase and frequency values for ++ pair considered in the comparison. + + and - - pairs show similar distributions—the co-rotating pairs are coherent on timescales monthly and longer. - + and + - pairs show similar distributions—the counter-rotating pairs are not coherent on any preferred timescales.

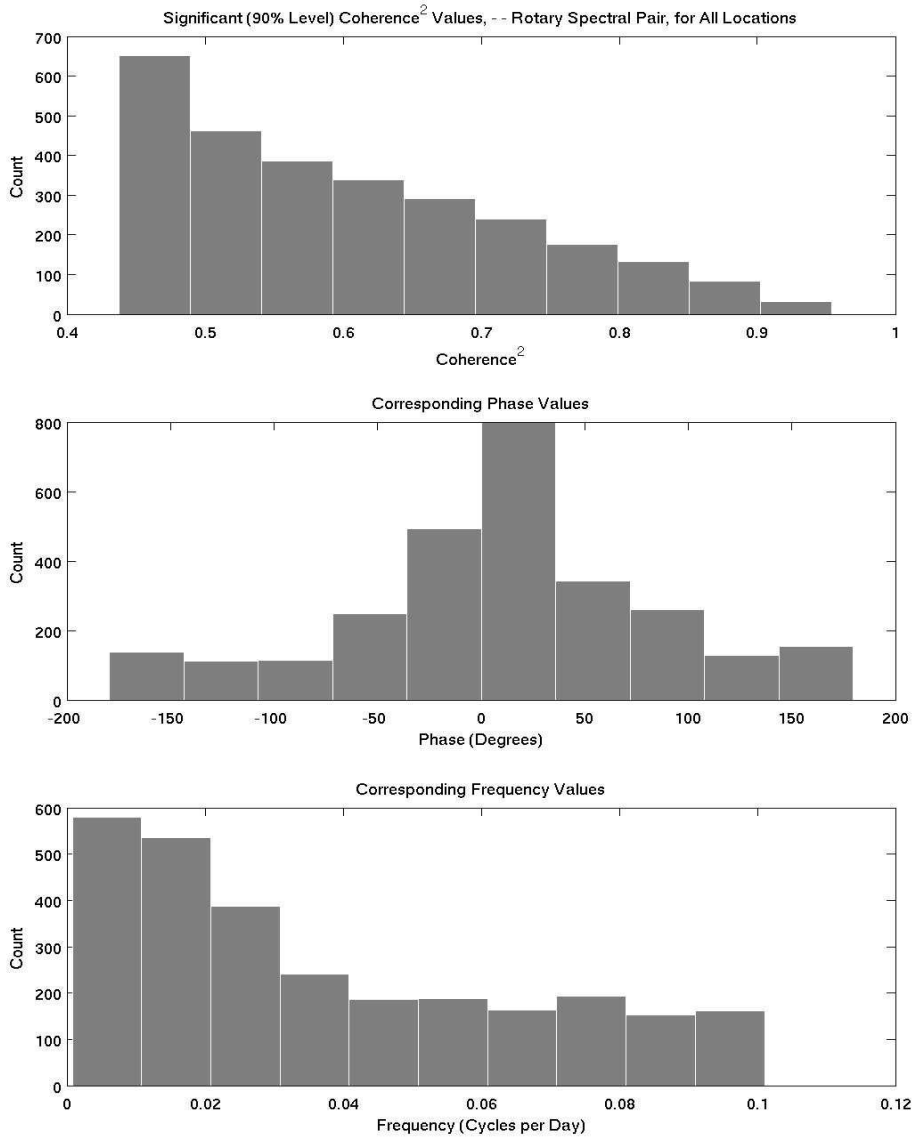


Figure 3-17b. Distributions of the significant (at the 90% level) coherence squared, phase and frequency values for -- pair considered in the comparison. + + and - - pairs show similar distributions—the co-rotating pairs are coherent on timescales monthly and longer. - + and + - pairs show similar distributions—the counter-rotating pairs are not coherent on any preferred timescales.

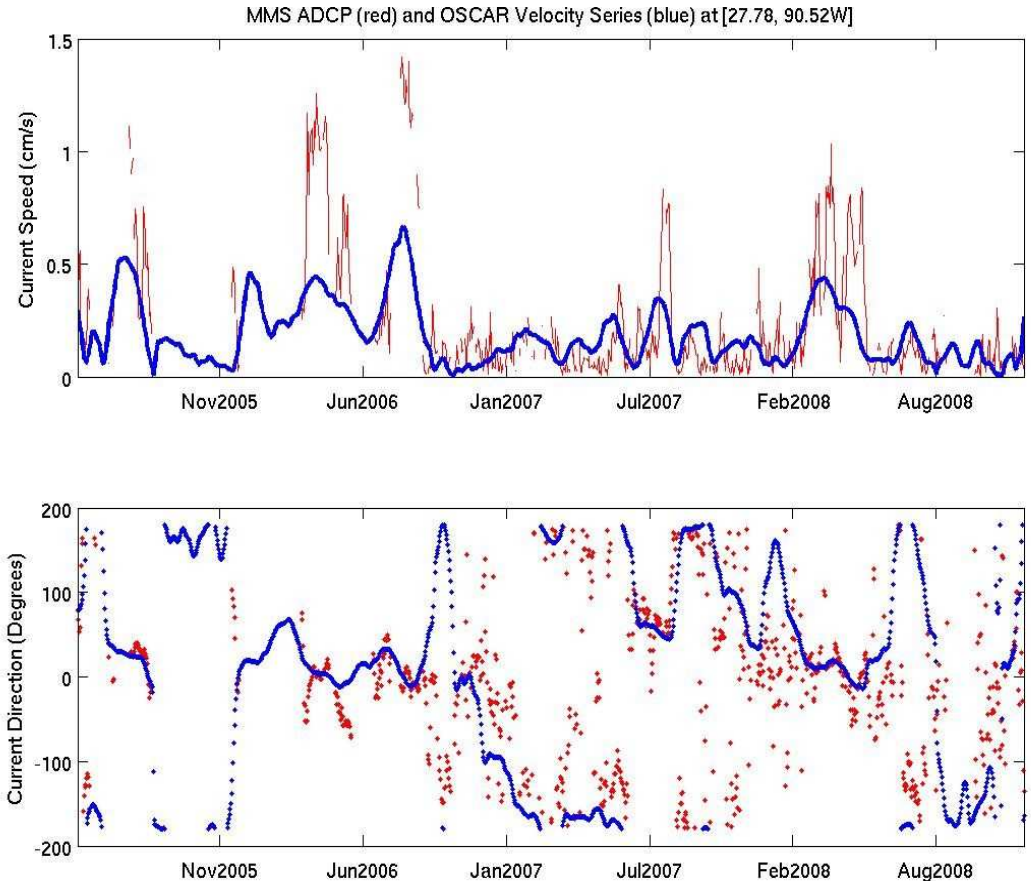


Figure 3-17c. The velocity series from a particular MMS ADCP location and the corresponding OSCAR series used in the comparison. Top panel: current speed in cm/s. Bottom panel: current direction in degrees. OSCAR tracks with the general pattern of the data but underestimates higher current speeds.

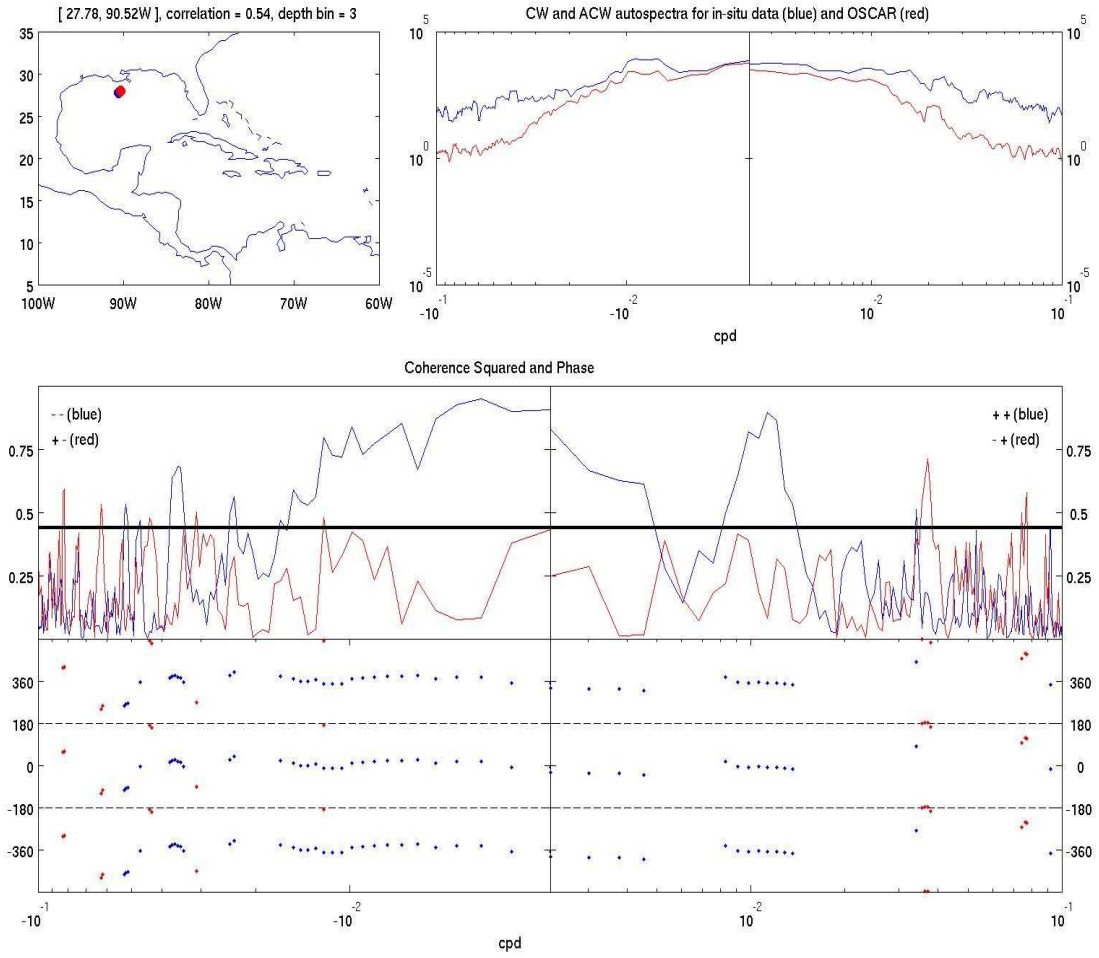


Figure 3-17d. An example set of comparison plots for the rotary spectral analysis. This location shows agreement in the co-rotating components over a wide range of lower frequencies, including the monthly band for the - - pair.

To take a closer look at the agreement between the co-rotating components, distributions of frequencies corresponding to phase values inside and outside the range of -45 degrees to +45 degrees are shown in Figures 3-18a (+ +) and 3-18b (- -). For the phase values nearer to 0 degrees (the majority of the values), there are the same preferred frequencies where the two series are coherent corresponding to periods of monthly and longer. For phase values outside the given range, there are no preferred frequencies. Many of these coherence squared values identified as significant above the 90% level are

there by chance and are not meaningful. The OSCAR output velocities and moored ADCP measured velocities are coherent on timescales of monthly and longer.

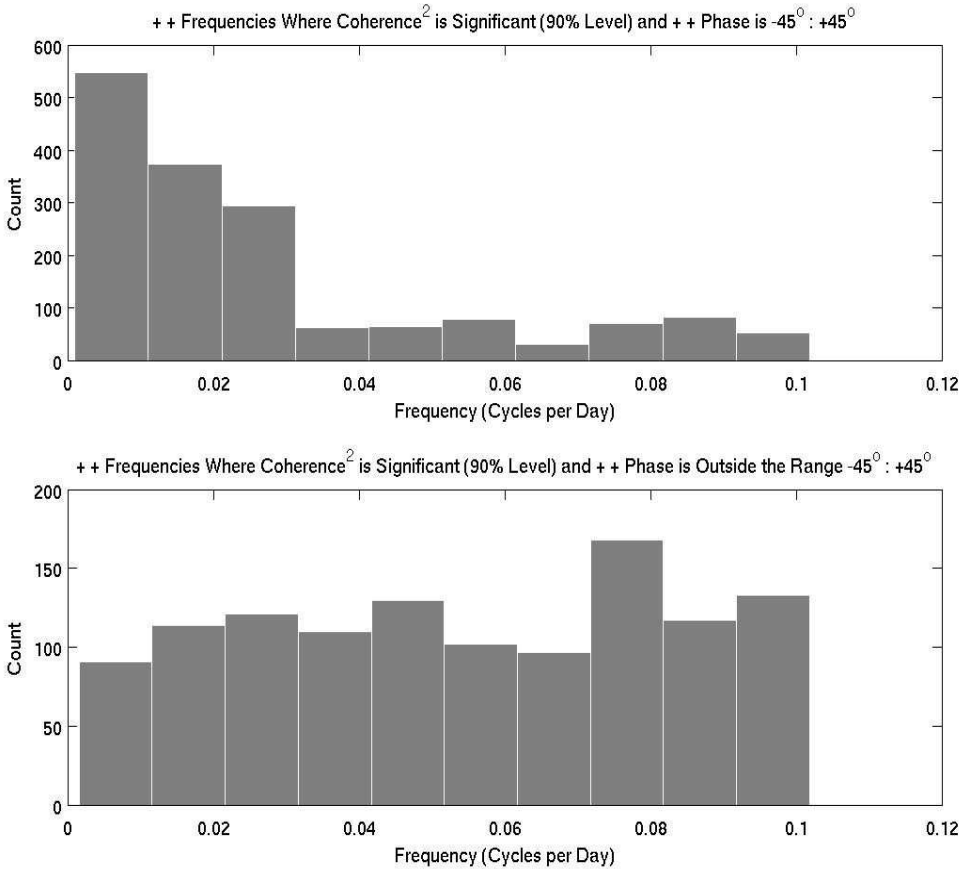


Figure 3-18a. Frequencies for the co-rotating components with phase values nearer (top) and farther (bottom) from 0 degrees. Those nearer to 0 degrees show the preferred timescales of monthly and longer. Those farther from 0 degrees show no preferred timescales and indicate many of these values may have appeared significant by chance.

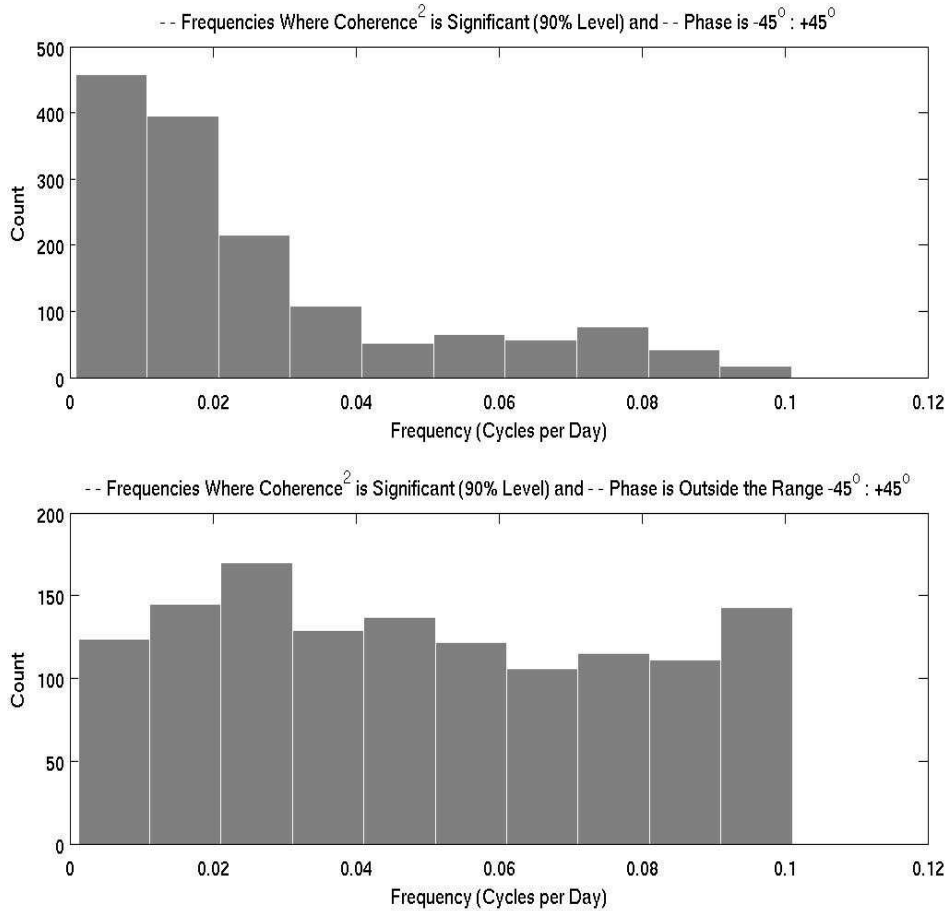


Figure 3-18b. Frequencies for the co-rotating components with phase values nearer (top) and farther (bottom) from 0 degrees. Those nearer to 0 degrees show the preferred timescales of monthly and longer. Those farther from 0 degrees show no preferred timescales and indicate many of these values may have appeared significant by chance.

3.8 Shipboard Velocity Comparison to OSCAR: Data and OSCAR Series Chosen for the Comparison

3.8.1 Data Sources

In-situ velocity data from ADCPs aboard ships of opportunity or scientific research vessels used in the comparison to OSCAR velocity output come from several different sources. Some are provided by the Joint Archive for Shipboard ACDP (JASADCP) at the University of Hawaii in conjunction with the National Oceanographic Data Center (NODC), housing data from many different cruises of both domestically and internationally based programs. Data was accessed through JASACDP at http://ilikai.soest.hawaii.edu/sadcp/. Other data is from the *Explorer of the Seas* research program at the University of Miami's Rosenstiel School of Marine and Atmospheric Science (RSMAS) with Royal Caribbean International and NOAA's Atlantic Oceanographic and Meteorological Laboratory. The *Explorer of the Seas* is a Royal Caribbean cruise ship fitted with an ADCP, recording data as it crosses the Caribbean on its usual routes. Collection of the data used for this comparison began in October 2000. Access to the data was provided directly from RSMAS. Data can now be found on the program website at http://oceanlab.rsmas.miami.edu/.

Estimated errors in the *Explorer of the Seas* ADCP data are documented by Beal et al. 2008. Mounting an ADCP on a cruise ship as opposed to a research vessel (or having a moored ADCP) presents unique challenges with the data collected due to the relatively inaccurate heading estimates of the ship and bubble contamination. Another source of error at any particular ocean velocity data point is interference between the two

ADCPs used to collect the data. When interference occurs, the ocean velocity measurements have erroneously high amplitudes.

Heading errors mainly affect the cross-track component estimates of velocity. The random error in the velocity due to heading errors was estimated by field-testing the ship's gyrocompass heading in a single Florida Current crossing in 2006. The random errors in the ocean velocities were estimated to be 4 to 6 cm/s. Bias errors on mean differences are assumed to result from heading biases that can be highly variable depending on whether or not the ship is being maneuvered. In this same field test during a crossing of the Florida Current, the bias error was estimated at 4 cm/s. More maneuvering leads to greater bias error. Conversely, with the ship under steady steam, the bias error can be much smaller—in degrees from a heading standard that was assumed to be correct during the test, a difference of 1º (maneuvering) versus 0.2º – 0.3º (steady steam). The contribution to the cross-track component velocity error is directly proportional to the sine of the heading error, so the error during maneuvering could be approximately 3 to 5 times higher than under steady steam.

Two factors leading to bubble contamination are the flat-bottomed hull of the *Explorer of the Seas* and its relatively high cruising speeds compared to a typical research vessel. The flat bottom of the ship leads to some rolling motion when the ship is being maneuvered, creating bubble contamination. Bubble contamination when underway mainly affects the along-track component estimates of the ocean velocity because bubbles are dragged along with the ship's motion, biasing velocities in the direction of forward motion. Contamination is greatest during maneuvering, acceleration and deceleration or rough sea conditions and results in unrealistically large ocean velocity measurements by

the ADCP. The authors removed the contaminated single-ping data with an automated data editing method not detailed in their documentation. Using each single-ping data point instead of averaging over some number of pings (as is standard practice for this type of data) helps to recover some data that would be otherwise unusable because of bubble contamination but the individual measurements will have higher random errors. Bubble biases in the ocean velocities due to rough sea conditions remain in the dataset. Large intermittent biases from all error sources discussed here are possible throughout the dataset.

Figure 3-19 maps the shipboard ADCP data points over all available times used in the comparison. In many cases, tracks are repeated over time. The OSCAR third-degree resolution product developed by ESR was made available through NASA's Jet Propulsion Laboratory Physical Oceanography Distributed Active Archive Center.

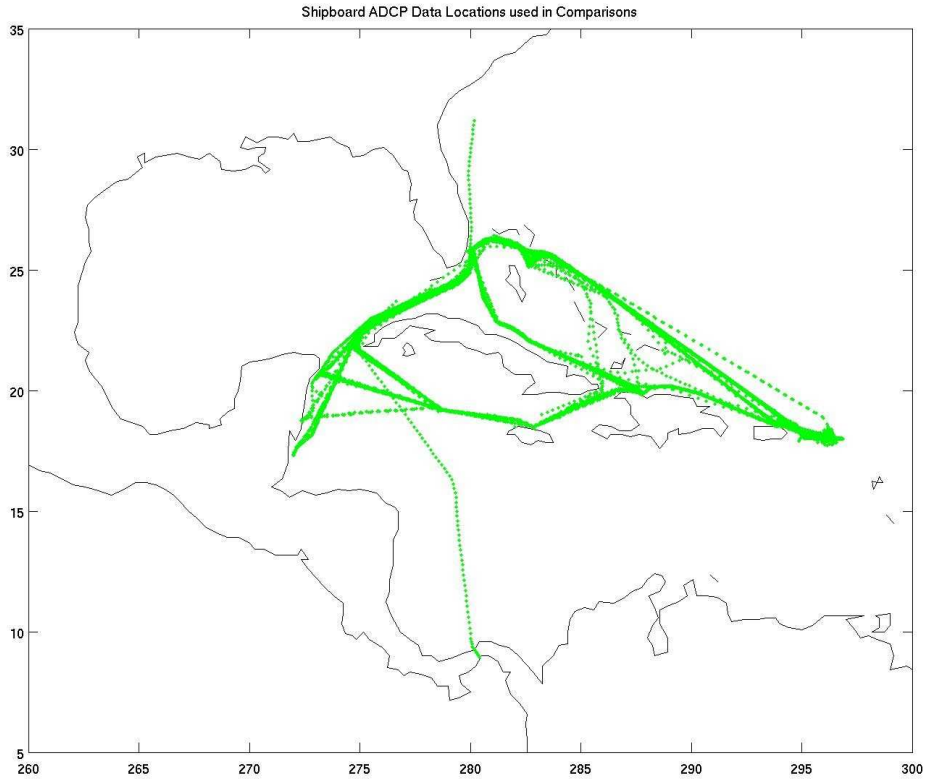


Figure 3-19. Shipboard ADCP measurement locations in the IAS used in the comparisons with OSCAR output velocities.

3.8.2 Matching Data and OSCAR Series

All data from both sources collected during the OSCAR time period October 1992 through 2008 in the IAS was identified and used in the comparisons where possible. Because some of the data were collected very near to land, a comparison with OSCAR is not possible because of the extent of the land mask of OSCAR. The selection process led to the creation of a database consisting of 469 ship tracks smoothed to an hourly time interval using box averaging. OSCAR is interpolated from every 5 days to hourly using linear three-dimensional (latitude, longitude and time) interpolation. Since each location is one measurement in time, each latitude, longitude, time point is matched to the nearest corresponding 3-D OSCAR point.

3.8.3 Preliminary Evaluation of the Series

Before doing any calculations involving the data series, the shipboard data points should be spatially smoothed so that they are as close as possible to the spatial of OSCAR along each ship track but still maintain the characteristics of the data series. Vector correlation (described in Section 3.6.1) is used to do a preliminary evaluation of the agreement between the shipboard ADCP and OSCAR velocities with different amounts of smoothing. The idea is to identify where vector correlation has increased and variance has decreased, being careful not to overstep the limit where the smoothing begins to eliminate signal instead of just noise. Figure 3-20 shows the change in the value of ρ^2 and σ^2 using different filter lengths on the hourly shipboard ADCP series. The blue dots at each filter length are the values for all of the 469 individual ship tracks compared to OSCAR and the red dots connected with the red line are the average values $<\rho^2>$ for each filter length. The same scheme follows for σ^2 and $<\sigma^2>$. The maximum average vector correlation squared values occur at 3 and 5-point smoothing. A sharp drop in the average variance is observed from no smoothing to 3-point smoothing and then the values fall off more gradually thereafter. When choosing the filter length, the physical distance between the smoothed ship points should be considered because it would not be desirable to smooth so much that the spacing between points is more than one OSCAR grid cell. A plot of the average spacing between points using 5-point smoothing is shown in Figure 3-21. For all of the ship tracks, none have spacing that is larger than the third-degree OSCAR grid spacing though some do approach it, so 5-point smoothing would be the most that is advisable.

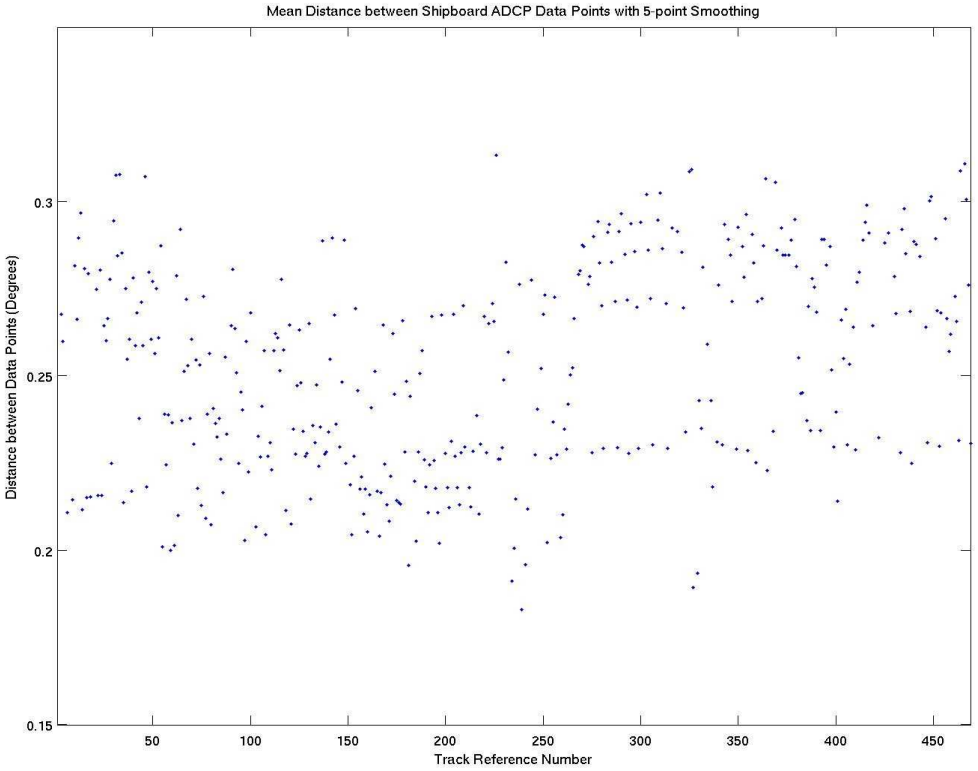


Figure 3-20. The change in values of vector correlation squared (top) and variance (cm^2/s^2) (bottom) for each ship track series at different along-track smoothing lengths (hours). Average values are indicated in red.

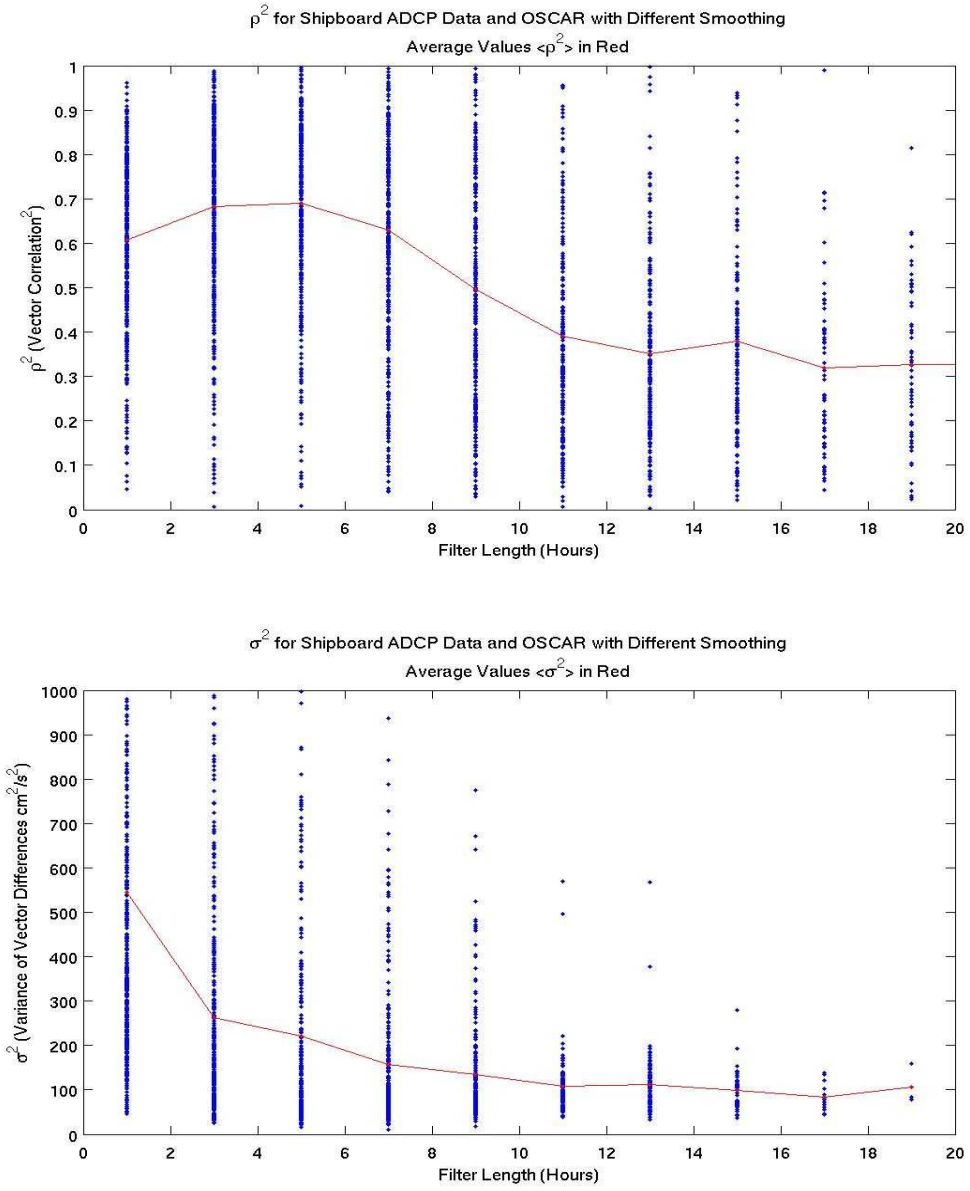


Figure 3-21. The average spacing (degrees) between smoothed ship track points using 5-point smoothing. None are larger than the grid size of third-degree OSCAR.

Before making a final decision on the amount of smoothing to be used, some other factors possibly affecting the agreement between the series should be considered. Altimetry, on which OSCAR is largely based, tends to be much more accurate away from land and in deeper water. Using 5-point smoothing and looking at the vector correlations

for water depth in Figure 3-22, there is no improvement in the agreement for increasing water depth at a given track center location. Clustering at certain depths is simply due to the fact that many tracks repeat (because the cruise ship tends to take the same route from one place to another), so their corresponding water depth will be very similar. It should be noted, however, that vector correlation value having no dependence on water depth is not a definite indicator that there is no change in the accuracy of OSCAR in different water depths. As discussed in Section 3.6.1, vector correlation values can indicate agreement between series under constant changes in magnitude and/or rotation over the length of the series. If there is some approximately constant offset between OSCAR velocities and measured shipboard velocities, vector correlation values could still be high. Further investigation with other parameters is required and is discussed in following sections.

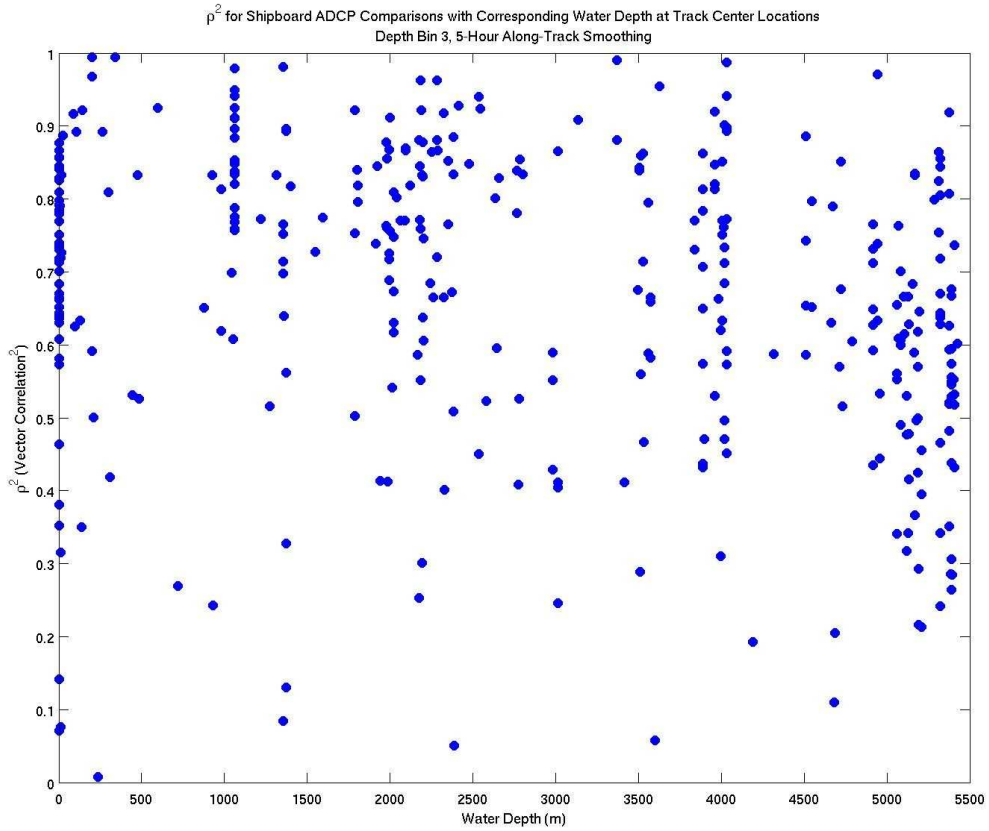


Figure 3-22. Vector correlation squared and total water depth (m) at the location of the shipboard ADCP measurement. There is no clear dependence of vector correlation squared value on water depth.

Looking at the vector correlation values at 5-point smoothing, it appears that some are much higher than is to be expected for this sort of data (very near or equal to 1). Upon further investigation, it becomes apparent that the lengths of some of the shipboard ADCP series must also be considered. Using 5-point smoothing on the shorter of the series leads to artificially high vector correlation values. For this reason, 3-point smoothing is chosen as it leads to nearly the same vector correlation values (as seen in Figure 3-20) and has the added benefit of allowing shorter series to be better represented after smoothing. A triangular (weighted) smoothing method on the hourly series is ultimately used for the comparisons.

3.9 Shipboard Velocity Comparison to OSCAR: Methods Used for Comparison

The most simple way to quantify the difference between the OSCAR output velocities and the shipboard ADCP velocities is to compute the vector difference—a vector quantity with u and v component differences having magnitude and direction. Physical factors of water depth and nearness to land at each measurement location as well as the time of the measurement are considered. Water depth and distance to nearest land can affect the ability of the OSCAR product to reproduce velocities because the accuracy of altimetry can be related to both. There should also be a check that these larger differences did not occur over a particular range in time that may be tied to either the altimetry or the shipboard measurements.

3.10 Shipboard Velocity Comparison to OSCAR: Comparison Results

3.10.1. Considering Water Depth

A plot of the magnitude of the vector difference is in Figure 3-23 (top). Each point is the magnitude of the vector difference in cm/s at each available shipboard measurement location and time. These points fall into three different groups. The first group (red) is in shallower water (<1000 meters) and above the background noise level. The second group (cyan) is in water at or deeper than 1000 meters and above the background noise level. The remaining points fall within the background noise level of the values (gray), identified by eye to be ~ 65 cm/s. A map of the locations where these points occur is at the bottom of Figure 3-23. The majority of the shallow points are geographically separated from the others. They are located in the Yucatan Strait and around southern peninsular Florida and the Florida Keys. Points in the second group are scattered along most of the ship tracks but they are most densely located along the tracks

from the Yucatan of Mexico to the Florida Keys. The water depths at these locations range from shallow to >5000 meters but are in the same geographic location. Along these tracks is the high-velocity, highly variable flow of the Yucatan Current, current at the entrance and exit of the Loop Current and the Florida Current. Both groups of larger-valued points (red and cyan) from Figure 3-23 are distributed evenly across all measurement times (not shown), so there was no effect on the vector difference in time.

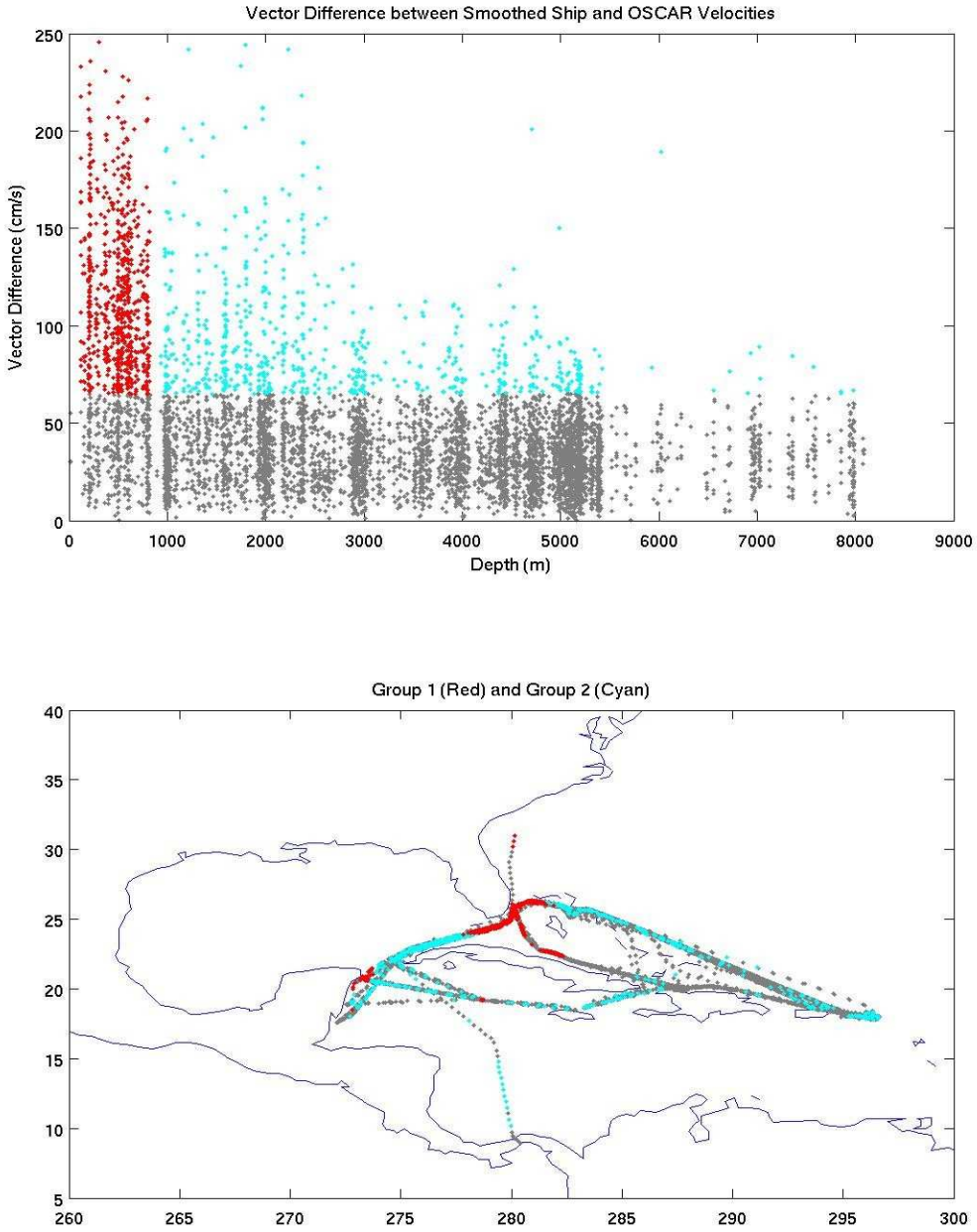


Figure 3-23. (top) Magnitude of the vector difference (cm/s) and total water depth (m) at the location of the shipboard ADCP measurement. Three groups of points: inside the background noise level (gray), above the background noise level in <1000 meters of water (red) and above the background noise level in 1000 meters of water and deeper (cyan). (bottom) Map of color-indicated shipboard ADCP measurement locations.

3.10.2. Considering Nearness to Land

The magnitude of the vector difference is plotted versus distance to nearest land in Figure 3-24 (top). The background noise level in the magnitude of the vector difference is estimated at 75 cm/s. Points above this threshold and those that are less than 1.5 degrees from the nearest land are highlighted in red. The 1.5-degree cutoff is chosen because that is the approximate value where the density of the higher-valued points falls off into the background noise. Figure 3-24 (bottom) shows the map of the locations of the highlighted points. The vast majority of these larger-valued points are along the Yucatan-to-Miami track, the same high-energy region identified with the water depth comparison. A look at the scatter plots for u and v component velocities for the highlighted near-to-land points from OSCAR and the shipboard ADCPs in Figure 3-25 reveals the higher the velocity, the more OSCAR appears to underestimate the measured value for both components. Figure 3-26 takes these same points and plots the speed and direction at those locations for OSCAR and the measured velocities. The speeds (top) from OSCAR underestimate the largest measured speeds by a factor of 2. Though there is some scatter, the direction (bottom) of OSCAR output and the shipboard measurements agree rather well. OSCAR can reproduce the direction of flow but not the magnitude, especially for the highest speeds. When the vector differences for OSCAR and the shipboard ADCP velocities are normalized by the shipboard ADCP speed (assuming the measured speeds to be the true values), the higher-valued, near-to-land locations identified in Figure 3-24 fall into the distribution with the rest of the points. Figure 3-27 shows these highlighted points and the rest of the normalized vector difference values. There is no dependence of normalized vector difference on distance to nearest land.

Notice that the highlighted points from the high-flow region between the Yucatan and Miami are clustered around a value of 0.5, again showing that OSCAR underestimates the highest velocities in the shipboard ADCP dataset by a factor of 2. The results of this analysis suggest the large vector differences between OSCAR output velocities and shipboard ADCP velocities are not predominantly caused by a dependence on water depth or nearness to land, rather, these values occur in the region of high-velocity flow from the Yucatan to Miami. OSCAR cannot represent these large velocities well because the current is narrow. In producing OSCAR, the gradient of the height field is smoothed across the spatial extent of the current. When considering a horizontal cross-section of the current for smoothing, the peak will influence the end points and vice versa, creating large point-by-point differences at the edges as well as center of the current when compared to measured velocity. It is the horizontal spatial scale of the current that is important, but this narrow current also happens to have the highest velocities and is the reason why OSCAR is consistently underestimating these values for this region.

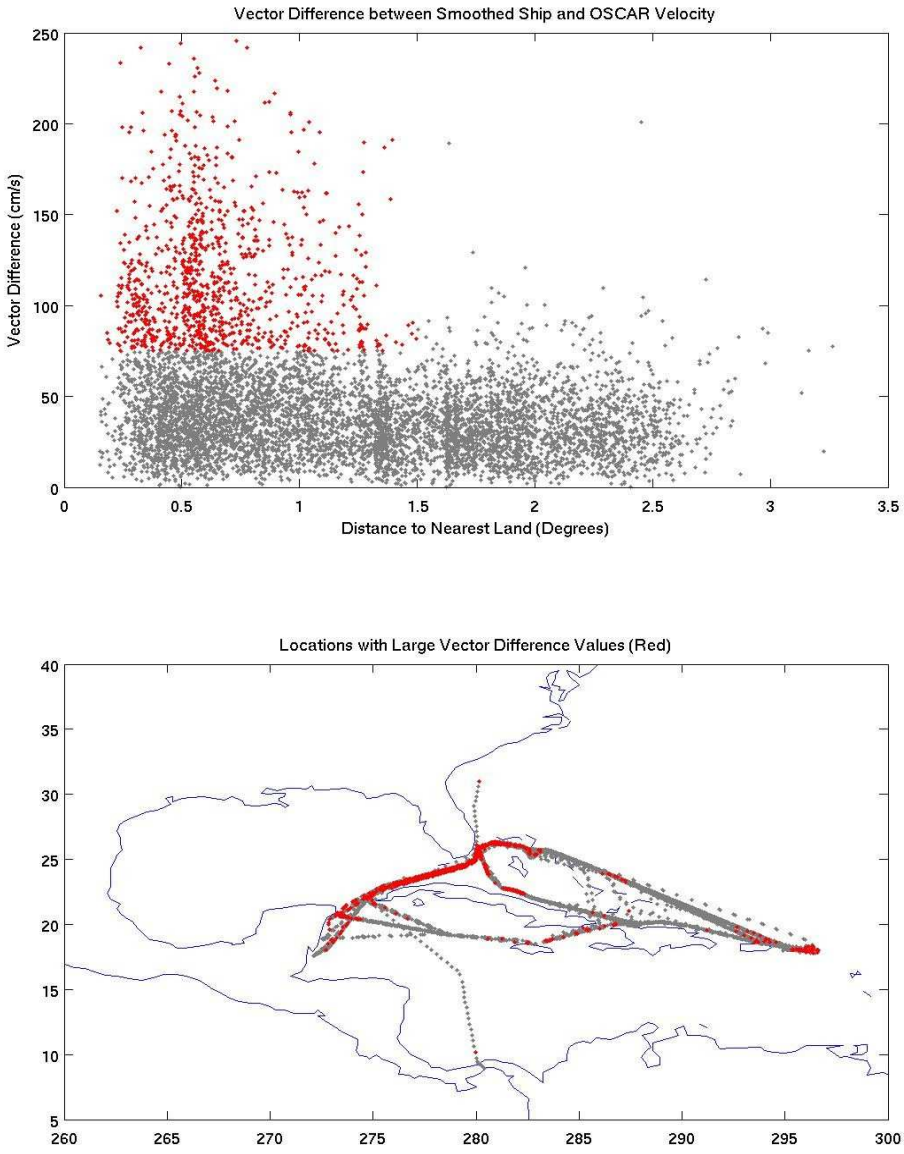


Figure 3-24. (top) Magnitude of the vector difference (cm/s) and distance to nearest land (degrees). Points above the background noise level highlighted in red. (bottom) Map of higher-magnitude vector difference values.

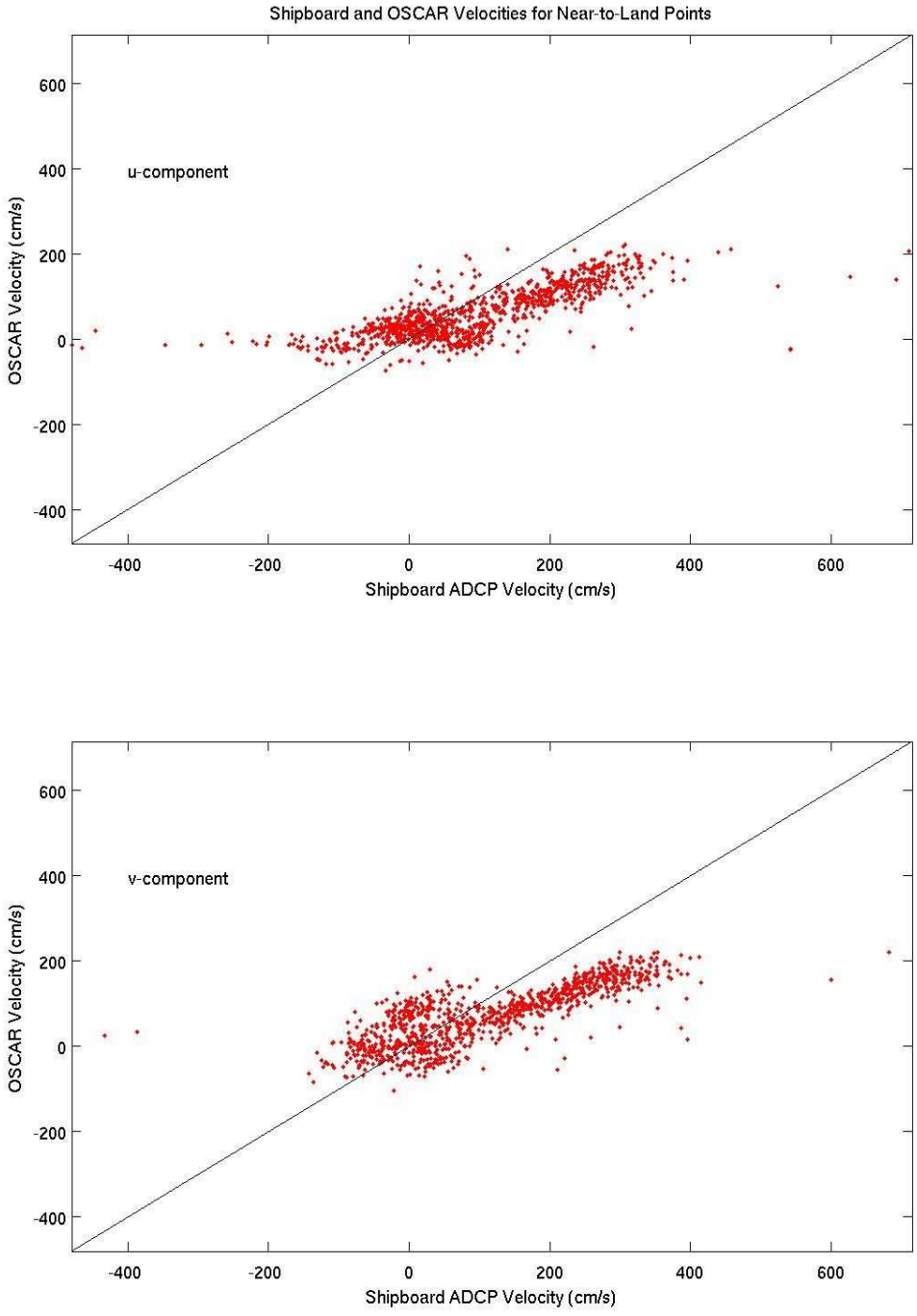
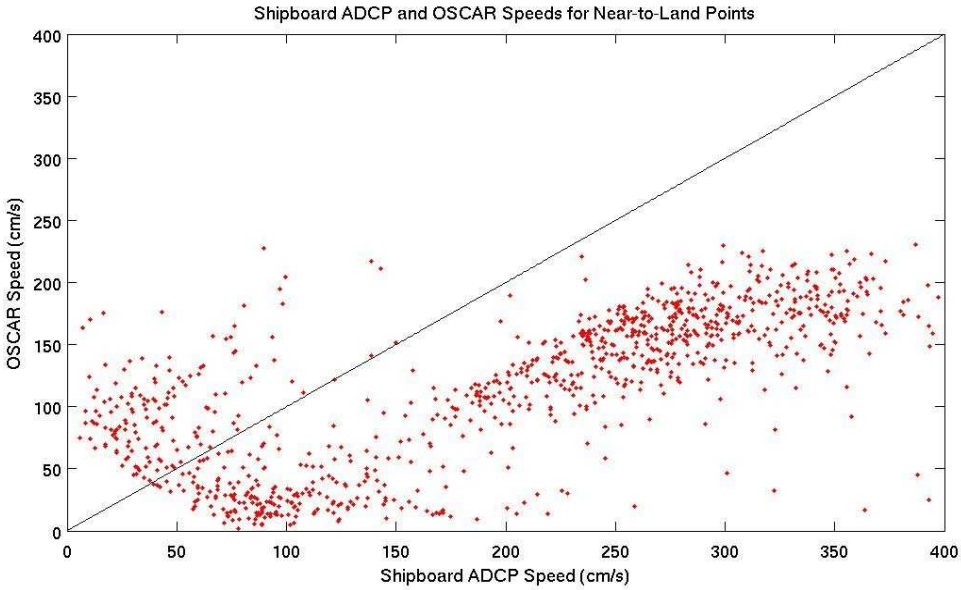


Figure 3-25. Component velocities for OSCAR and shipboard ADCP for the higher-valued, near-to-land points from Figure 3-24. OSCAR appears to increasingly underestimate the measured velocity as the magnitude increases.



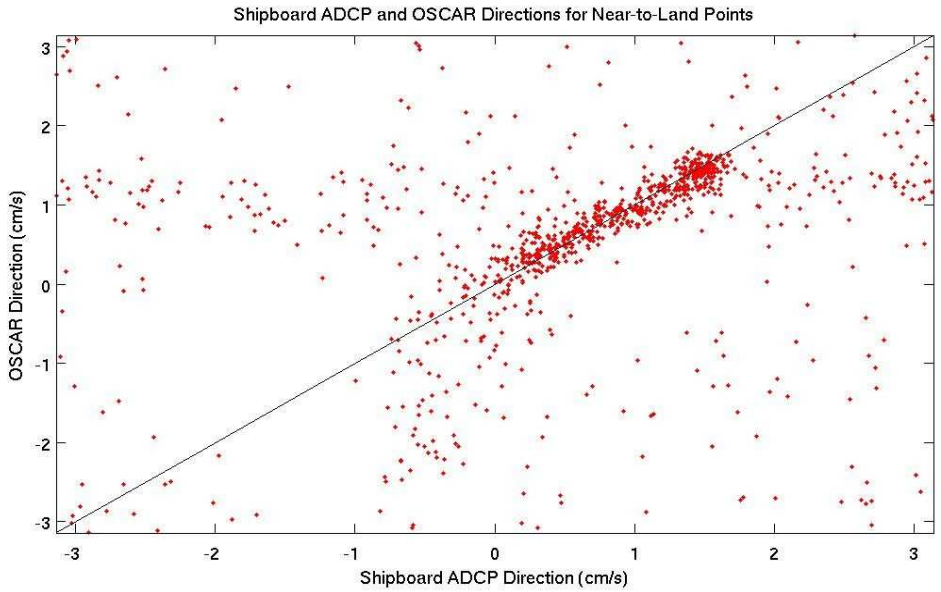


Figure 3-26. Speed (top) and direction (bottom) for locations of highlighted points from Figure 3-24. Speeds show OSCAR underestimates by a factor of 2 at highest velocities while the directions agree rather well.

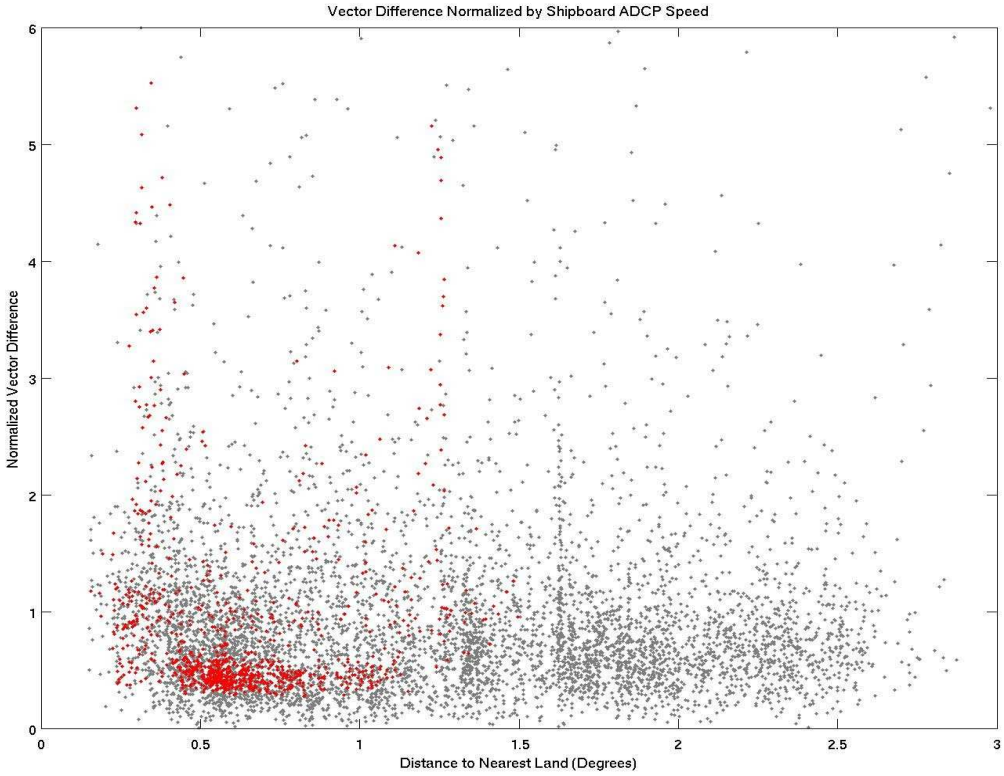


Figure 3-27. Vector difference values (cm/s) normalized by shipboard ADCP speeds. Highlighted points from Figure 3-24 are shown in red and fall into the distribution with the rest of the points. There is no clear dependence of normalized vector difference on nearness to land.

Water depth and nearness to land may still contribute to the differences, but it is impossible to separate out these factors from this data. Recall from previous discussion that this combined affect of water depth and nearness to land on the OSCAR velocities is also evident in the fixed ADCP velocities compared to OSCAR. Figures 3-13 and 3-14 show the vector correlation squared values for OSCAR output and moored ADCP velocities with water depth and distance to nearest land. Vector correlation squared versus deeper water and greater distance from land both show increasing improvement in the agreement between the series. While the northern Gulf of Mexico experiences episodes of high-energy flow with the passage of shed Loop Current eddies, there is no regular, narrow, high-speed jet as with the measurements taken along the axis of the flow

from the Yucatan to Miami. Therefore, it is easier to see the effects of water depth and distance from land on OSCAR without the inability of OSCAR to reproduce the highest velocities in those narrow currents also being a large factor.

4 Circulation in the Intra-American Sea

4.1 Chapter Overview

To be useful in real-world applications, OSCAR must be able to reproduce physical characteristics of circulation over a range of timescales. Gridded altimetry products like AVISO have been publicly available for some time and have already been used by authors (e.g., Alvera-Azcarate et al. 2009) to describe circulation features in the IAS. The geostrophic component of OSCAR derived from altimetry is the largest contributor in magnitude to the surface currents in the IAS, so the addition of the Ekman component should provide an improved description of circulation over the geostrophic alone. On the basin scale, geostrophic currents computed from AVISO and OSCAR agree well for slower current speeds, but at higher speeds, geostrophic OSCAR underestimates speeds compared to the geostrophic from AVISO. This is due to the additional smoothing done in the processing of the OSCAR product. When the total OSCAR is compared to the geostrophic currents from AVISO on basin scales, they show the same season centers over the seasonal cycle, but the OSCAR seasonal cycle has a smaller range in average speed between seasons. The differences occur in areas where large-scale, wind-driven currents should be important, so OSCAR may be able to better represent the seasonal cycle over the IAS as a whole compared to a geostrophic-only comparison.

Areas where the Ekman component makes the largest relative contribution to the variability and the magnitude of the total speed are collocated with some of the shipboard ADCP data. In these locations, the geostrophic-only OSCAR and the total OSCAR are compared in a mean sense with these measured velocity data points. To within the error on each quantity, there is a slight improvement in the agreement of OSCAR and the data when the Ekman component is included with the geostrophic component. Locations where the Ekman component is not so relatively important (collocated with the moored ADCP data), no discernible improvement in the agreement is seen to within the error.

4.2 Circulation in the IAS Based on OSCAR

4.2.1 Evaluating OSCAR with Geostrophic Currents Computed from Altimetry

The geostrophic currents used in the creation of OSCAR are computed from gridded height fields derived from altimetry combined from two different sources (see Section 1.2.3) with a predictive method applied to bring the most recent field to real time. The focus of Alvera-Azcarate et al. 2009 was to describe the circulation of the IAS based on gridded altimetry. In order to make direct comparisons to this study, OSCAR output over the same area is compared to geostrophic currents reproduced using the same data and calculation method as Alvera-Azcarate et al. 2009. Figure 4-1a shows long-term mean field for the altimeter-derived geostrophic velocity. Arrows indicate relative vector magnitude and vector velocity direction while the color bar indicates long-term mean speed. The long-term mean field over the same times for OSCAR is shown in Figure 4-1b. The fields are remarkably similar even though OSCAR is a total current that includes the Ekman component of the velocity.

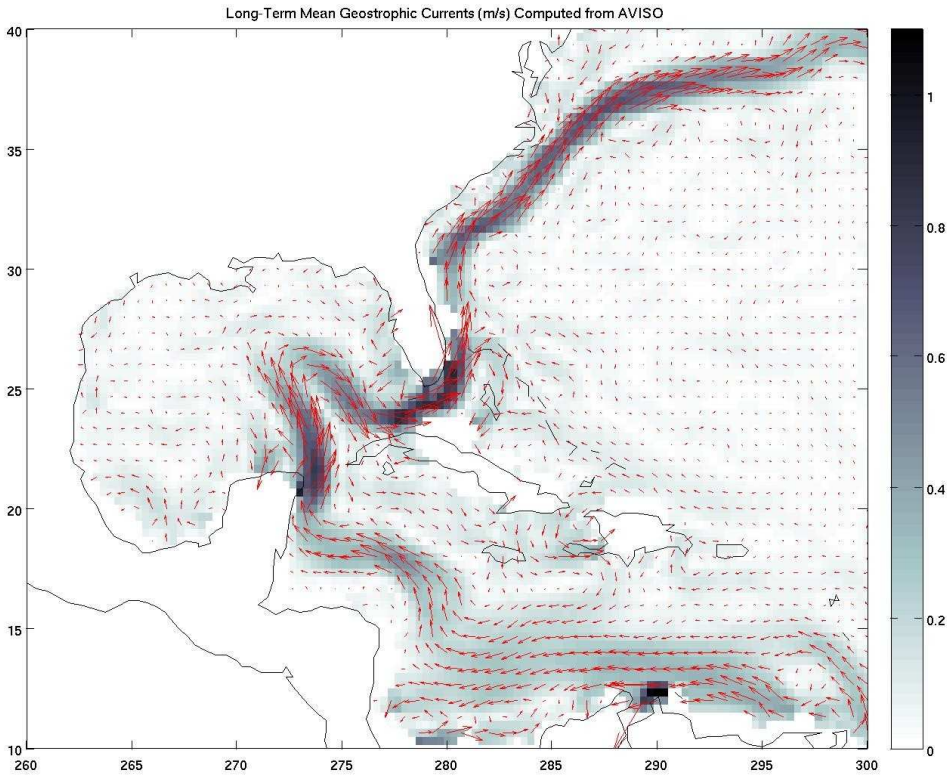


Figure 4-1a. Vector map of long-term mean geostrophic velocity (m/s) computed from AVISO height fields using the method of Alvera-Azcarate et al. 2009. Color bar indicates the long-term mean geostrophic speed (m/s).

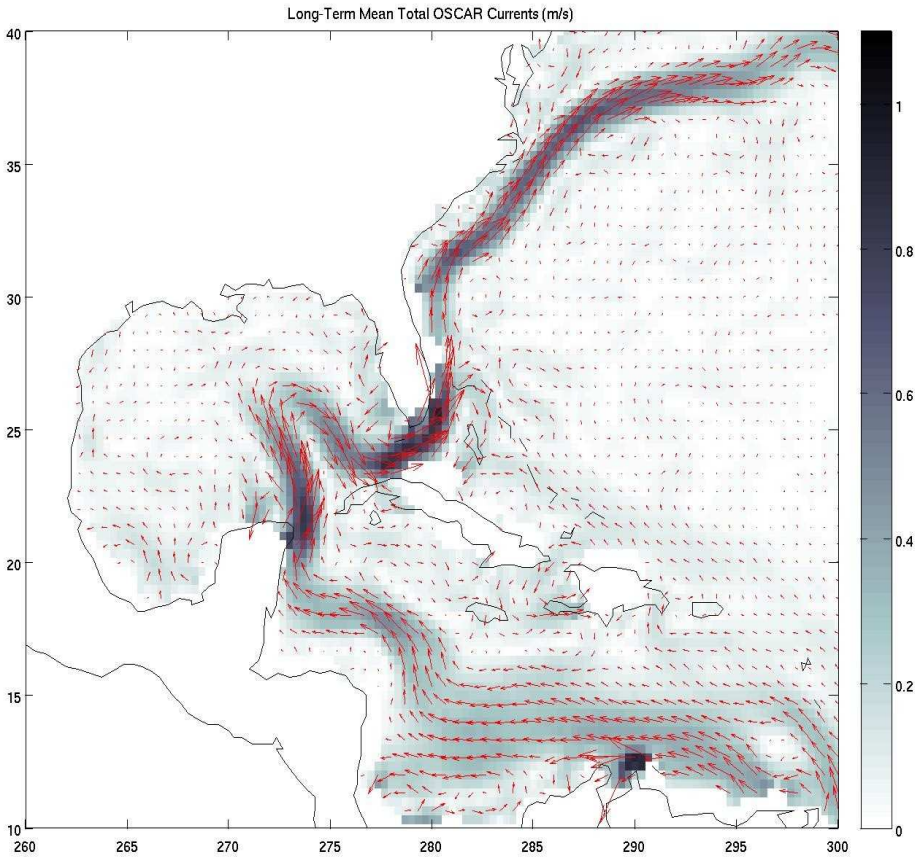


Figure 4-1b. Vector map of long-term mean total OSCAR velocity (m/s). Note the similarity of the major circulation features to Figure 4-1a even though OSCAR also includes the Ekman currents. Color bar indicates long-term mean speed (m/s).

When the instantaneous speeds for total OSCAR and the geostrophic speeds from AVISO are averaged over 100-point bins and plotted against each other (Figure 4-2), it is apparent that they agree very well at lower speeds but the OSCAR speeds begin to fall off of the equivalence line at speeds greater than about 0.8 m/s. The larger the magnitude of the geostrophic speed from AVISO, the more OSCAR falls off in comparison. This is most likely attributed to the additional smoothing done in the production of the OSCAR product. At the highest speeds (greater than 1.5 m/s), the geostrophic speeds max out while OSCAR goes up to almost 2 m/s, indicating OSCAR depicts a slightly larger range in speeds over this time period compared to the geostrophic speeds from AVISO.

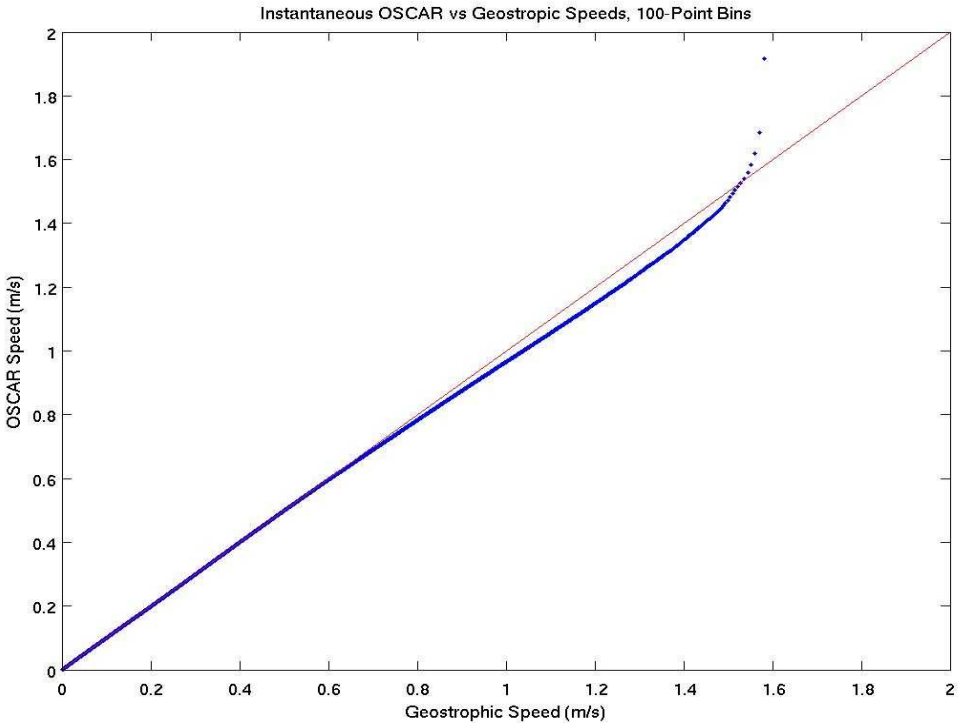


Figure 4-2. Instantaneous speeds for total OSCAR and the geostrophic speeds from AVISO averaged over 100-point bins.

Differencing the long-term mean fields results in the vector map shown in Figure 4-3. The difference between the geostrophic and total OSCAR means should be the Ekman component in the total OSCAR. The important thing to notice is that the magnitude of the differences is typically an order of magnitude smaller than the geostrophic speeds themselves or the total OSCAR speeds. This shows the scale of the Ekman component contribution. The largest non-geostrophic contribution in OSCAR is in the Caribbean Sea and generally has a NW to NNW direction. Not surprisingly, this is the area most affected by the Trade Winds. Note that even here where the Ekman appears to be largest, the magnitude of the Ekman component contribution is only ~ 10% of that of the geostrophic component.

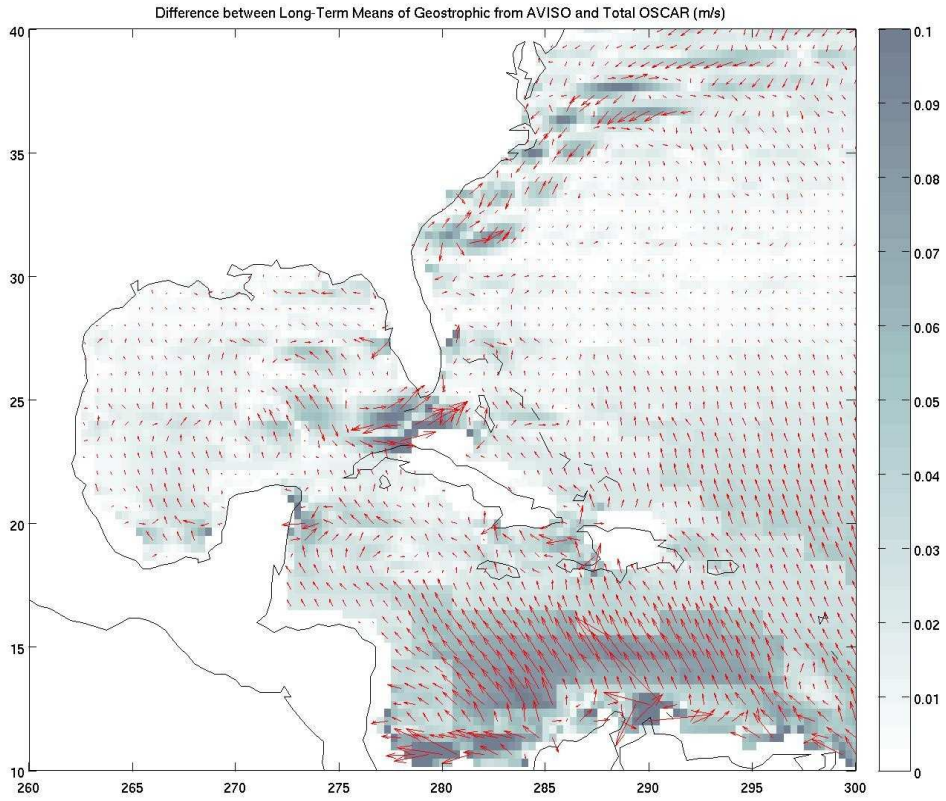


Figure 4-3. Vector map of the differences (m/s) between the long-term mean geostrophic speeds and long-term mean total OSCAR speeds. Values tend to be an order of magnitude smaller than the geostrophic alone or the total. The region that experiences the Trade Winds is prominent.

The phasing of total OSCAR and geostrophic currents from AVISO is also compared. The phases for each were computed on -π to π and then converted to a corresponding month number 1 through 12. The difference in the phasing at each point is mapped in Figure 4-4. The u (top) and v (bottom) components both show the same phasing to within 1 month at most locations. Recall from the shipboard ADCP comparisons to OSCAR in Chapter 5 that for points with large vector differences, OSCAR speeds were generally slower than the measured speeds but the directions agreed well. This is more evidence that while the smoothing in OSCAR affects its ability to

recreate the magnitude of the velocity in some cases, it is usually able to reproduce the direction and phasing of the velocity compared to velocities from other sources.

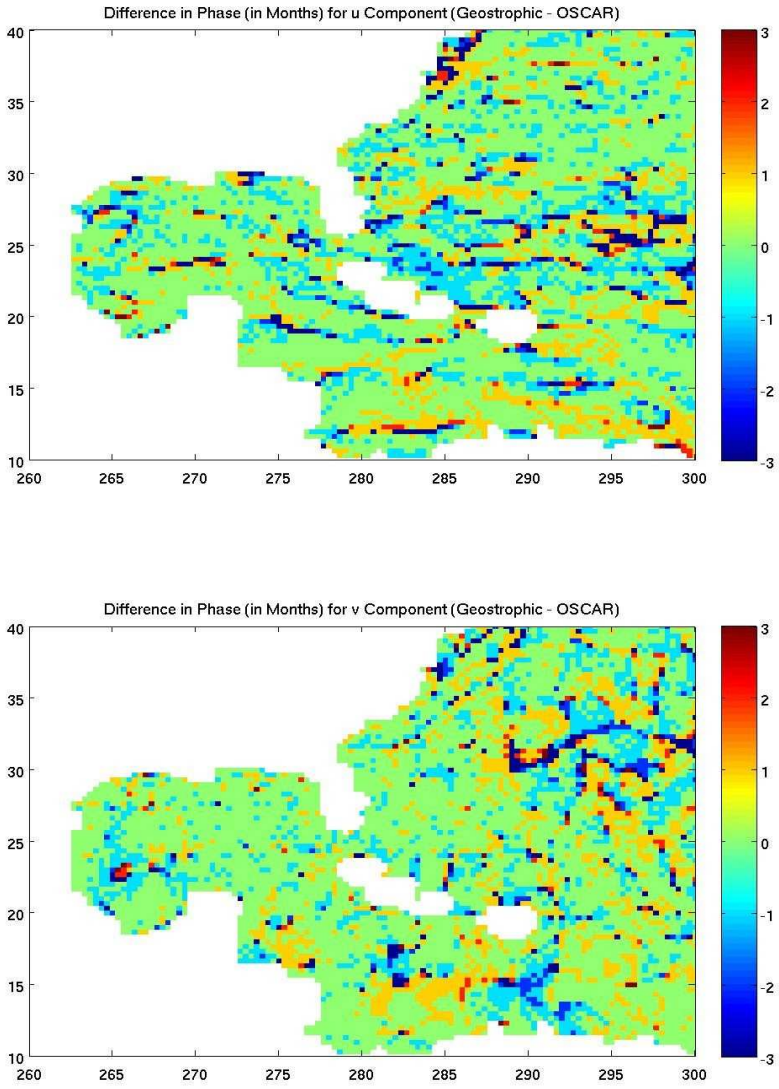


Figure 4-4. Difference (geostrophic from AVISO – total OSCAR) in the phasing at each point for the u (top) and v (bottom) components. The phases agree to within 1 month at most locations.

To further explore the variability in time of the geostrophic currents computed from AVISO and the total OSCAR output, overlapping 2-month averages are computed over the available times. This will show the intra-annual variability of each set of

velocities. A basin-wide average over the IAS for total OSCAR speed (top), u component velocity (middle) and v component velocity (bottom) is shown in Figure 4-5a. The corresponding values for the geostrophic from AVISO are in Figure 4-5b. Note that the y-axis scales for the speed or velocity are not the same either between the values for each dataset or between the datasets. Both show the usual January (winter), April (spring), July (summer) and October (fall) season centers in the overlapping 2-month averages in speed. The overall pattern of the fluctuation over the year for speed, u and v are similar and the differences in magnitude are not very large considering it is a basin-wide average and there is a lot going on that could affect the magnitude values for each dataset, especially since the total OSCAR includes the Ekman velocity contribution.

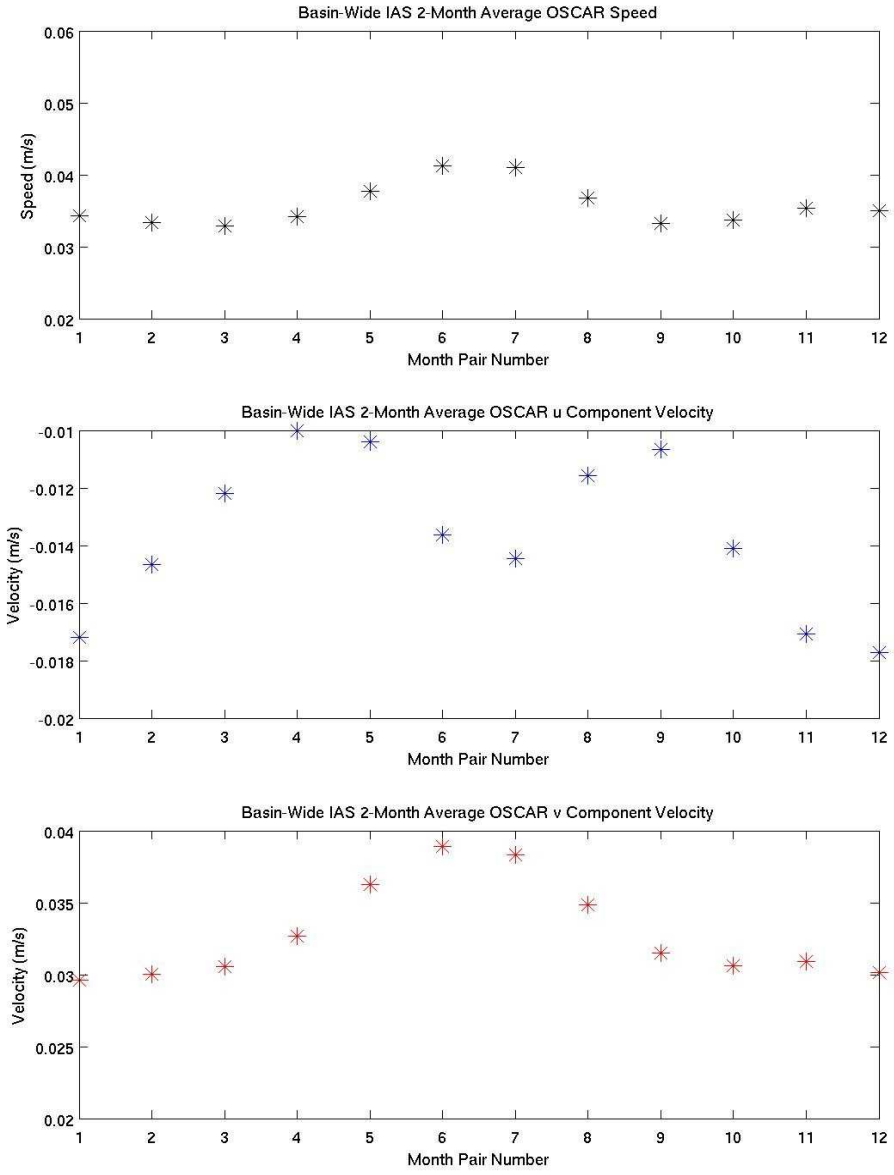


Figure 4-5a. Basin-wide overlapping 2-month averages over the IAS for total OSCAR speed (top), u component velocity (middle) and v component velocity (bottom)

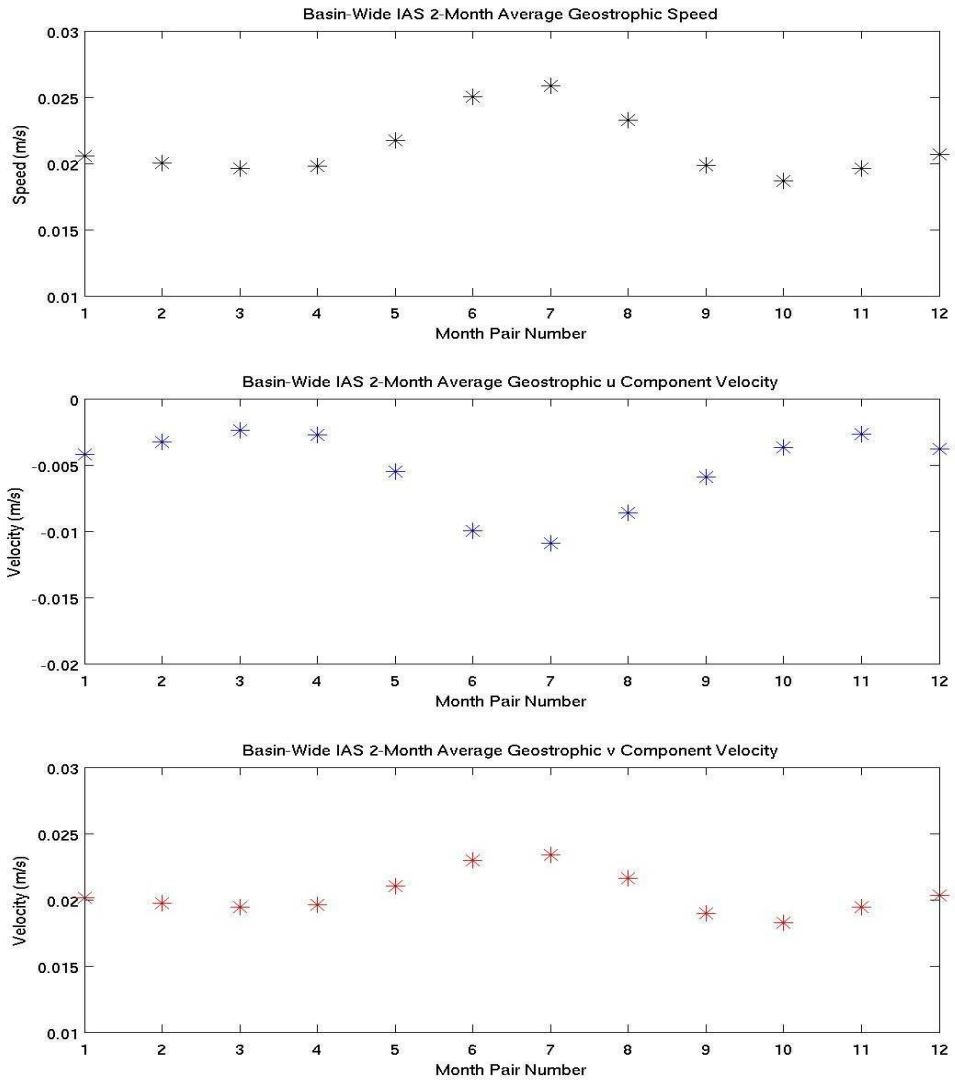


Figure 4-5b. Basin-wide overlapping 2-month averages over the IAS for geostrophic from AVISO speed (top), u component velocity (middle) and v component velocity (bottom)

To better characterize the differences in magnitude, the overlapping 2-month averages are repeated for the IAS, this time split geographically into northern and southern parts by a line chosen at 20°N latitude. The northern and southern parts of the IAS are very different as far as their flow regimes are concerned. Many authors (e.g. Richardson 2005) have shown that the southern IAS (mainly the Caribbean Sea) has a mean flow to the west in the eastern Caribbean transitioning to flow to the northwest in

the western Caribbean with its variability dominated by the passage of westward-propagating eddies. The northern IAS (the Gulf of Mexico and the Atlantic off the southeast U.S.) is dominated by the high-speed flow of the Yucatan Current, Loop Current, Florida Current and Gulf Stream. The plots for the southern IAS overlapping 2-month averages of speed, u and v component velocities for total OSCAR and geostrophic from AVISO are in Figure 4-6a and 4-6b. The plots of the speed (top) show the OSCAR speed is slightly elevated on average compared to the geostrophic speed. Looking at the components, it becomes obvious the larger magnitude is mostly contributed by the OSCAR v component, which is about 3 times larger than the v component from the geostrophic. The Ekman flow would be expected to show up mostly in the v component in this area because the dominant mean winds blow to the west and the classical Ekman model dictates the flow will be to the right of the wind. As seen in Figure 4-3, the difference in the long-term mean flow in the southern IAS has a NW to NNW direction. This is representative of the Ekman contribution but it should be noted that because the v component contribution is almost an order of magnitude smaller than the u component direction, the change in the speed with the addition of Ekman is small, even in this area where the winds are consistent in strength and direction and in a mean sense should have a relatively significant Ekman component. The northern IAS overlapping 2-month averages are in Figure 4-7a and Figure 4-7b. For speed, u and v, the values are very similar but the fluctuation patterns do not track very well. In the geostrophic quantities, the winter months are more prominent relative to the rest of the year when compared to the fluctuations for the total OSCAR. Looking at the difference in mean flow map in Figure 4-3 for the area of the northern IAS, it appears this difference may come from the

area off the east coast of the U.S. In winter, cold fronts frequently come off of North America with strong winds behind the front blowing to the southeast. The Ekman flow from these winds (to the right of the wind) would contribute -u and -v, making the OSCAR magnitudes smaller in winter months and resulting in the smaller range between seasons for total OSCAR.

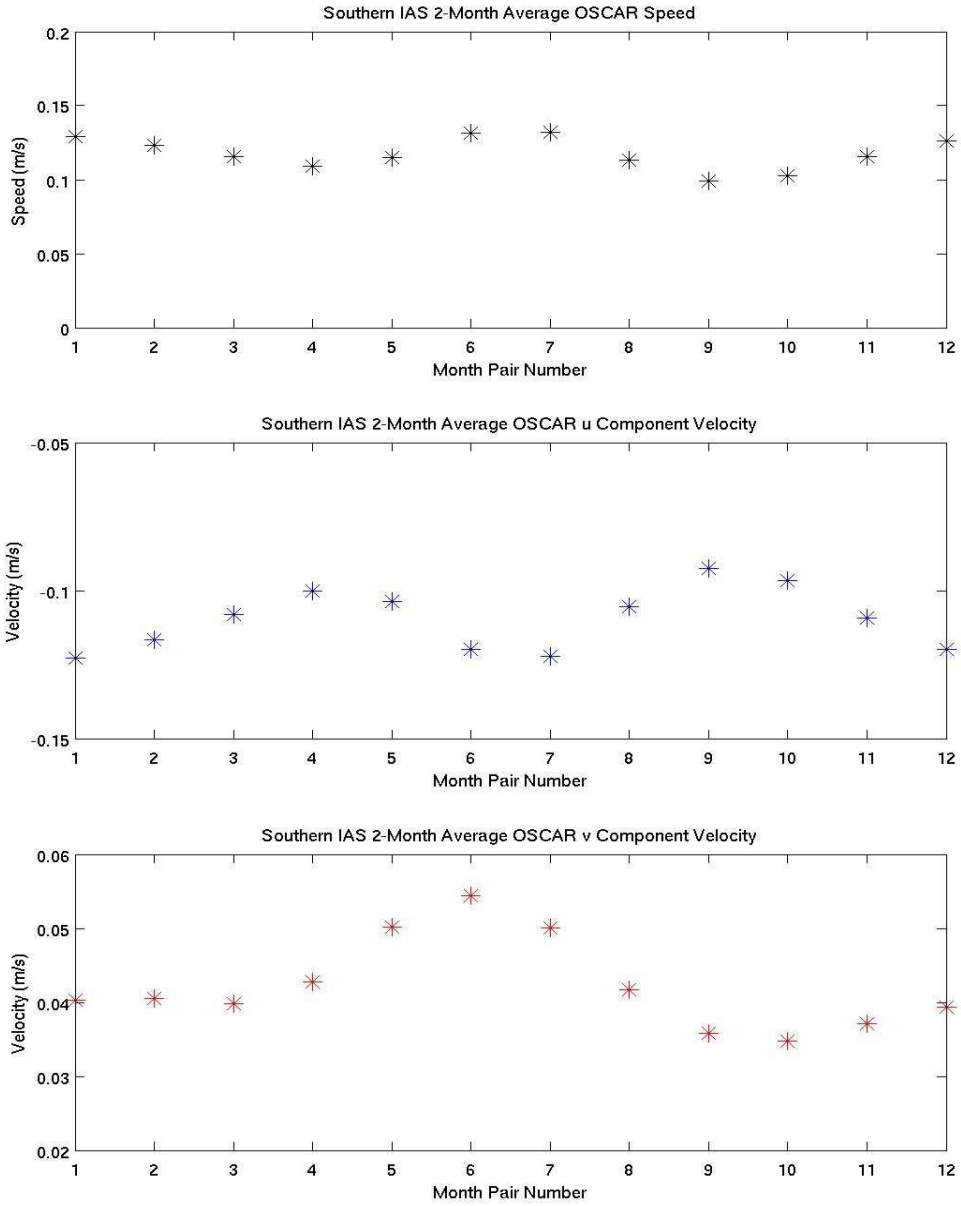


Figure 4-6a. Southern IAS overlapping 2-month averages for total OSCAR speed (top), u component velocity (middle) and v component velocity (bottom)

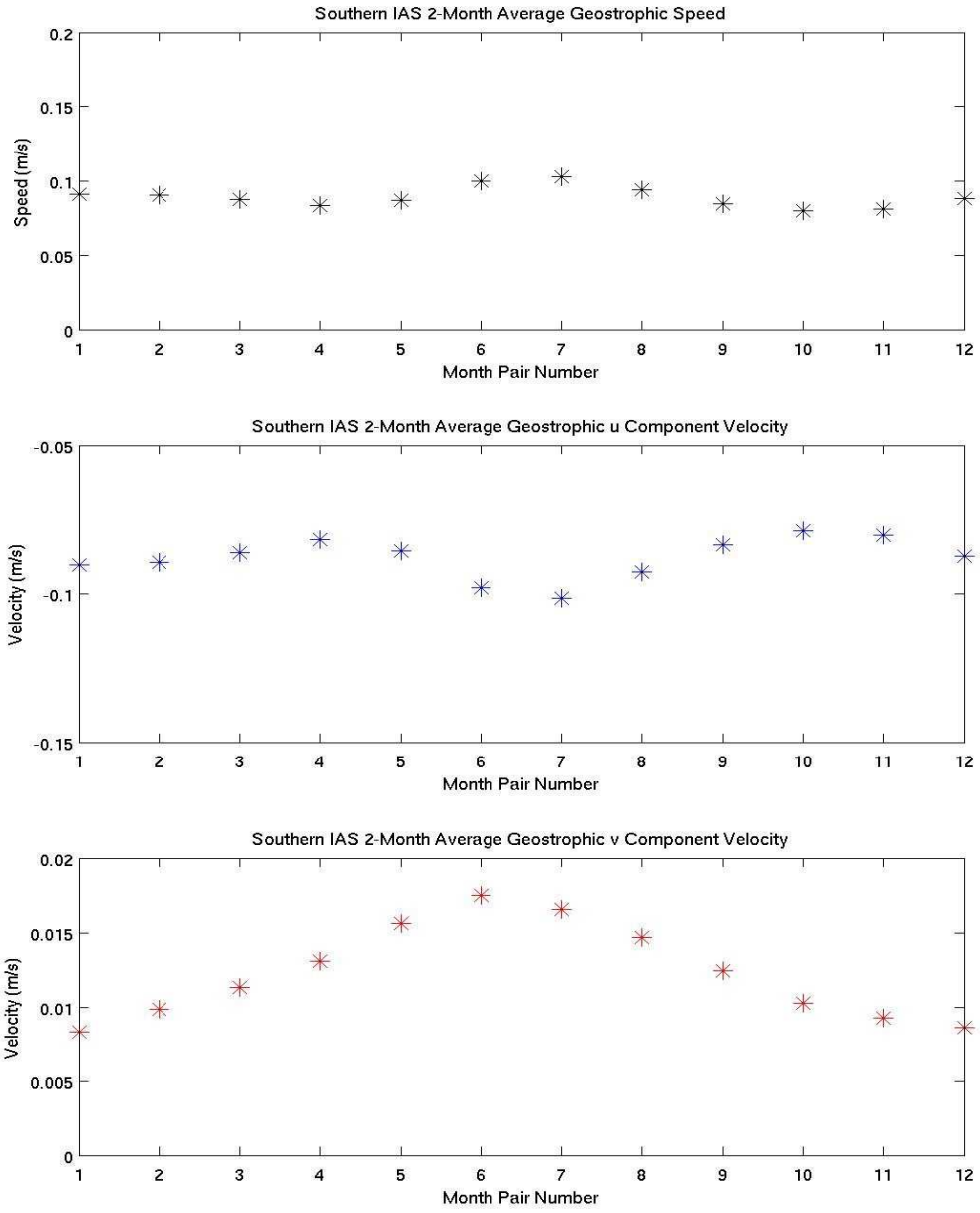


Figure 4-6b. Southern IAS overlapping 2-month averages for geostrophic from AVISO speed (top), u component velocity (middle) and v component velocity (bottom)

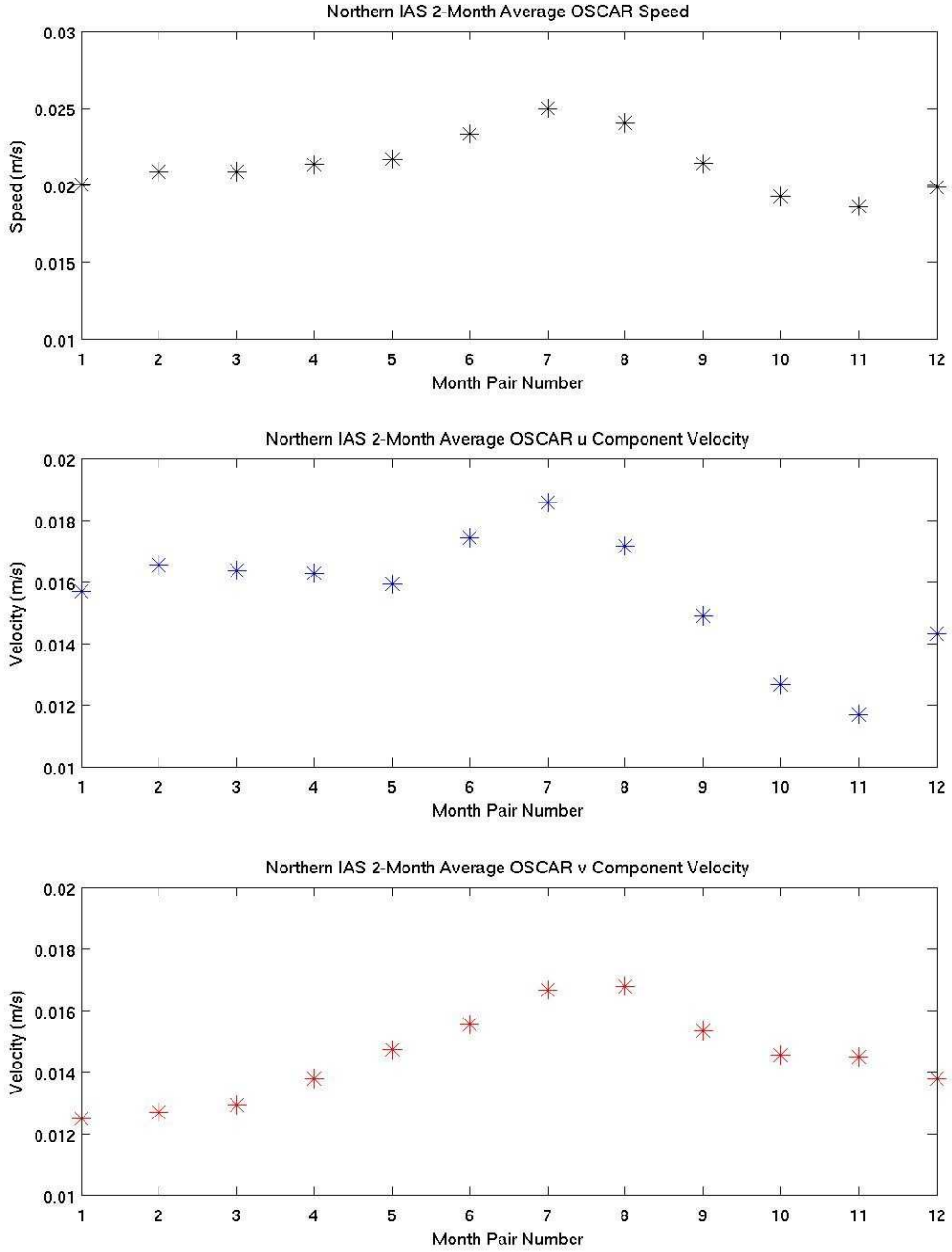


Figure 4-7a. Northern IAS overlapping 2-month averages for total OSCAR speed (top), u component velocity (middle) and v component velocity (bottom)

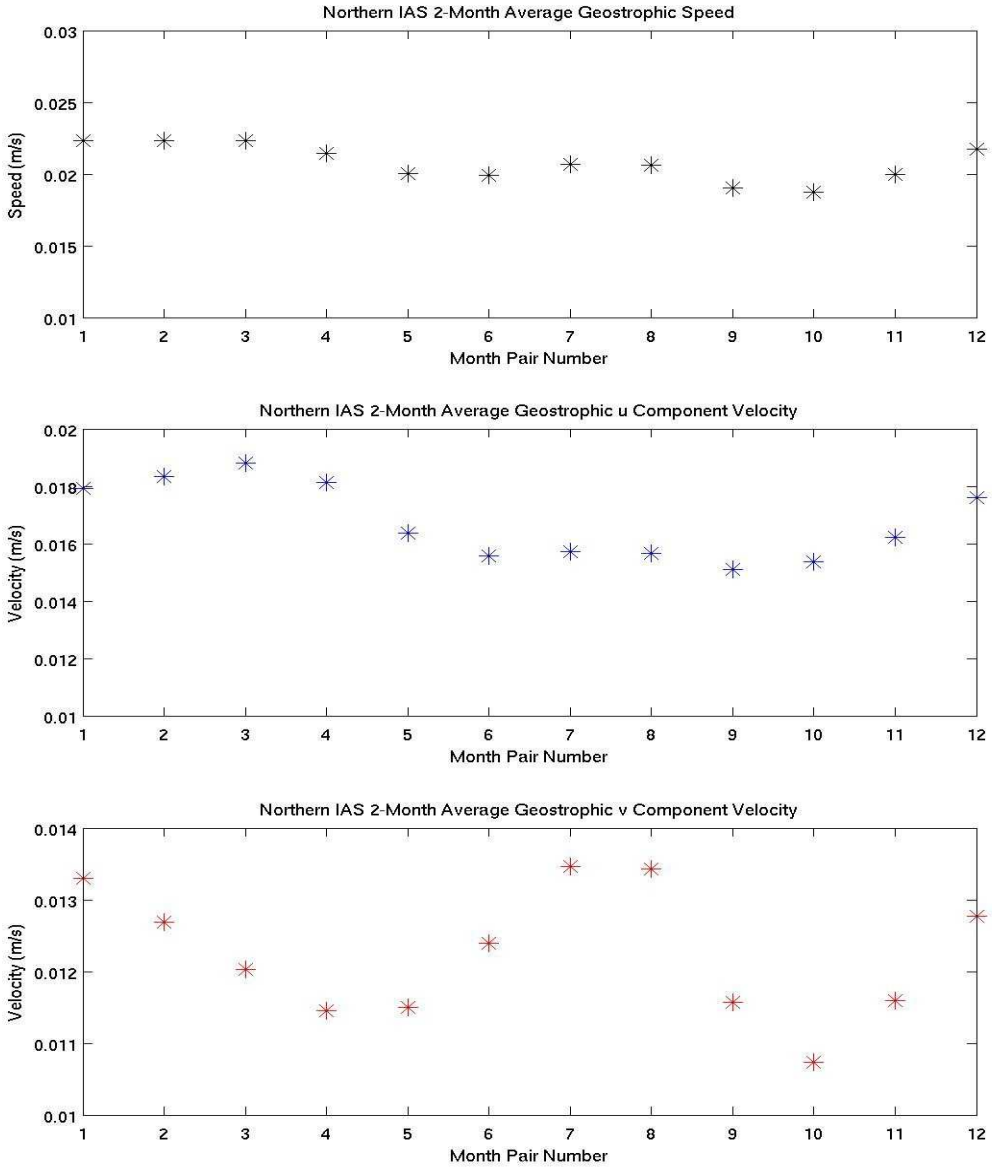


Figure 4-7b. Northern IAS overlapping 2-month averages for geostrophic from AVISO speed (top), u component velocity (middle) and v component velocity (bottom)

Later on in the analysis, the actual geostrophic and Ekman components of OSCAR were obtained from ESR. The vector map of the long-term mean of the geostrophic (buoyancy term included) component is in Figure 4-8. To compare this to the mean geostrophic currents computed from AVISO, a calculation performed by Alvera-

Azcarate et al. 2009 is repeated. In their Figure 1, a box in drawn in the eastern Caribbean. In this box, a maximum instantaneous flow speed of 1.28 m/s, mean 0.82 m/s and standard deviation 0.14 m/s are determined. The same calculation on the OSCAR geostrophic component gives maximum instantaneous flow speed 1.18 m/s, mean 0.72 m/s and standard deviation 0.12 m/s. The means are consistent within one standard deviation though the OSCAR geostrophic values are slightly suppressed compared to the geostrophic from AVISO values computed by Alvera-Azcarate et al. 2009. This could be a result of the additional smoothing done to OSCAR in producing the product.

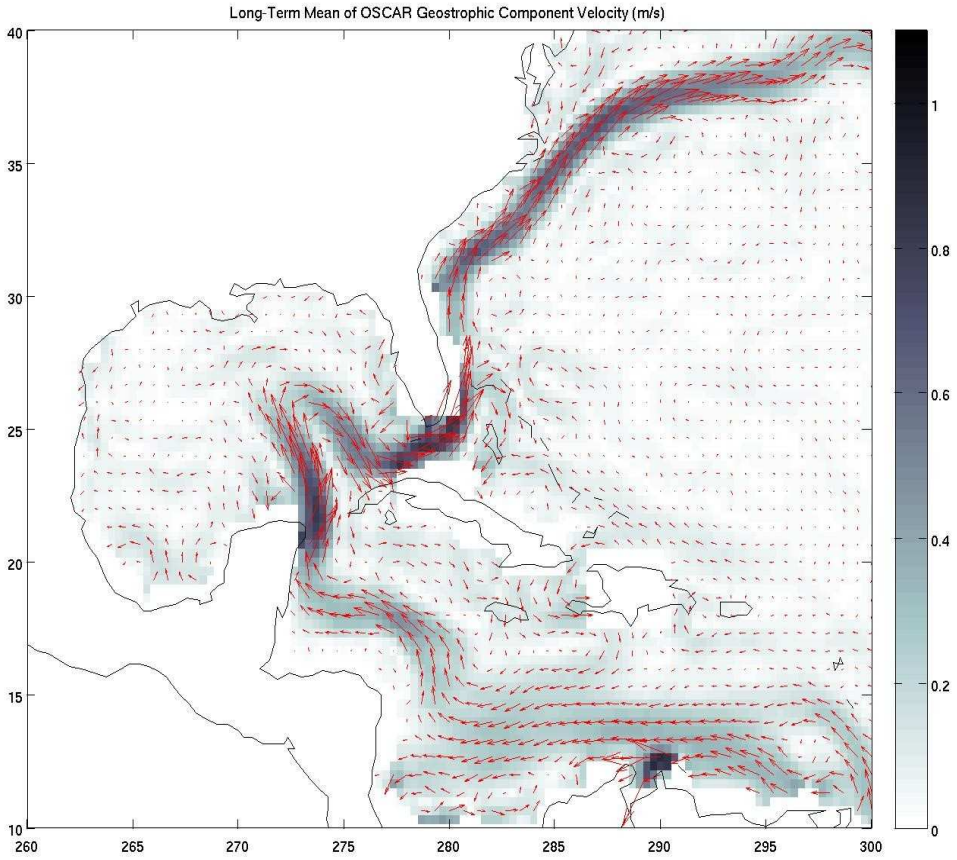


Figure 4-8. Long-term mean of the geostrophic (buoyancy term included) component velocity (m/s) for OSCAR.

To check what the Ekman flow should look like in the IAS, the long-term mean from NCEP winds is computed over the area and then converted to wind stress. Figure 4-

9a maps the wind stress values from NCEP interpolated using nearest neighbor interpolation to the OSCAR third-degree grid. The resolution of NCEP is much coarser, explaining the appearance of repeated patterns in the wind stress field. The long-term mean Ekman component from OSCAR is mapped in Figure 4-9b. The magnitudes fall on a scale of 0 to 0.1, making them an order or magnitude smaller than the geostrophic OSCAR component magnitudes. This is consistent with the differences from Figure 4-3 compared to the geostrophic velocities computed from AVISO. The overall pattern of the mean Ekman component from OSCAR agrees with the difference of the total OSCAR and geostrophic from AVISO, but there are some details in the difference that are not represented in the Ekman component field, especially in the far eastern Caribbean, the eddy-dominated area south of Cuba, around the Loop Current and along the Gulf Stream. This is consistent with the geostrophic velocities computed from AVISO being less smooth (having more spatial detail) than the geostrophic component from OSCAR and that shows in the difference from total OSCAR. The pattern agrees with the classical Ekman model, with the Ekman flow to the right of the wind stress.

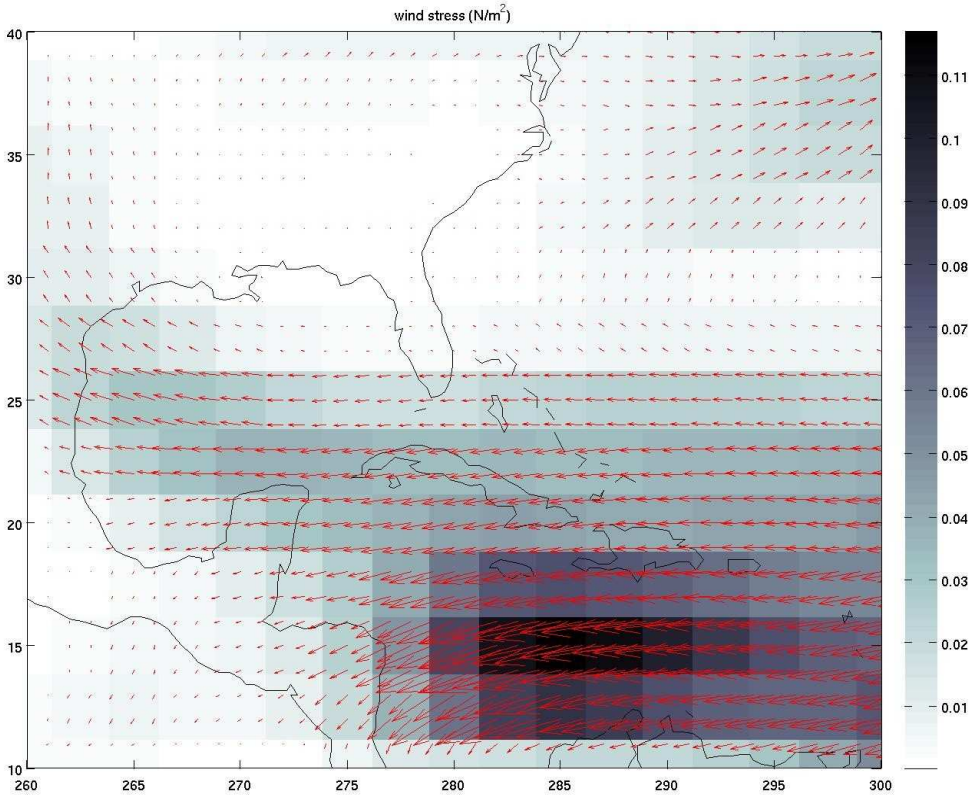


Figure 4-9a. Wind stress values (N/m^2) computed from long-term mean NCEP winds interpolated using nearest neighbor interpolation to the OSCAR third-degree grid. The resolution of NCEP is much coarser, explaining the appearance of repeated patterns in the wind stress field.

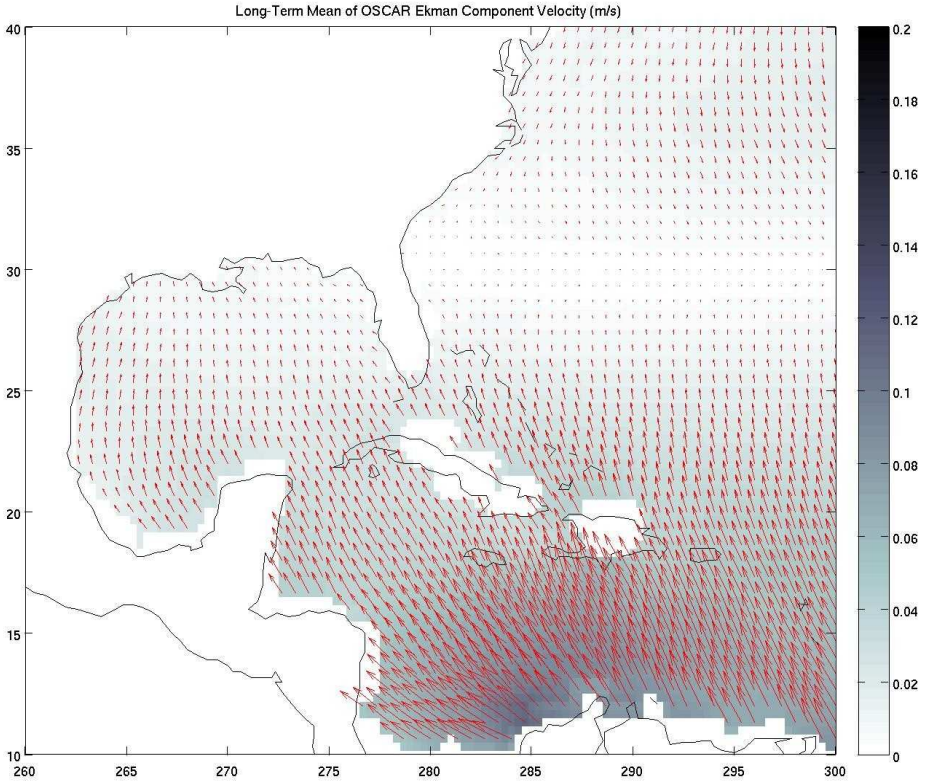


Figure 4-9b. Long-term mean Ekman component velocity (m/s) from OSCAR.

To quantify the rotation, the clockwise rotation angle that minimizes the squared difference between the Ekman component OSCAR and the rotated wind stress vectors is shown on the map in Figure 4-10a. Angles generally result in Ekman flow in a NW or NNW direction. The background grid pattern is again a result of interpolating the NCEP winds to a finer grid. Figure 4-10b shows an average clockwise rotation value over each grid latitude. There is a clear relationship with latitude in the rotation angle, in agreement with classical Ekman model for u and v components of Ekman velocity being dependent on latitude. The classical Ekman model gives a 1/f (where f is the Coriolis parameter) relationship with Ekman flow and therefore also with the angle of Ekman rotation for flow at the surface. This is because as the Coriolis force (the vertical rotation vector)

decreases with latitude, the Ekman spiral rotates more slowly with depth relative to that at higher latitudes, making the initial angle at the surface smaller (measured in the counter-clockwise direction). There is a direct relationship between f and clockwise rotation angle in this plot because rotation in the clockwise direction is negative.

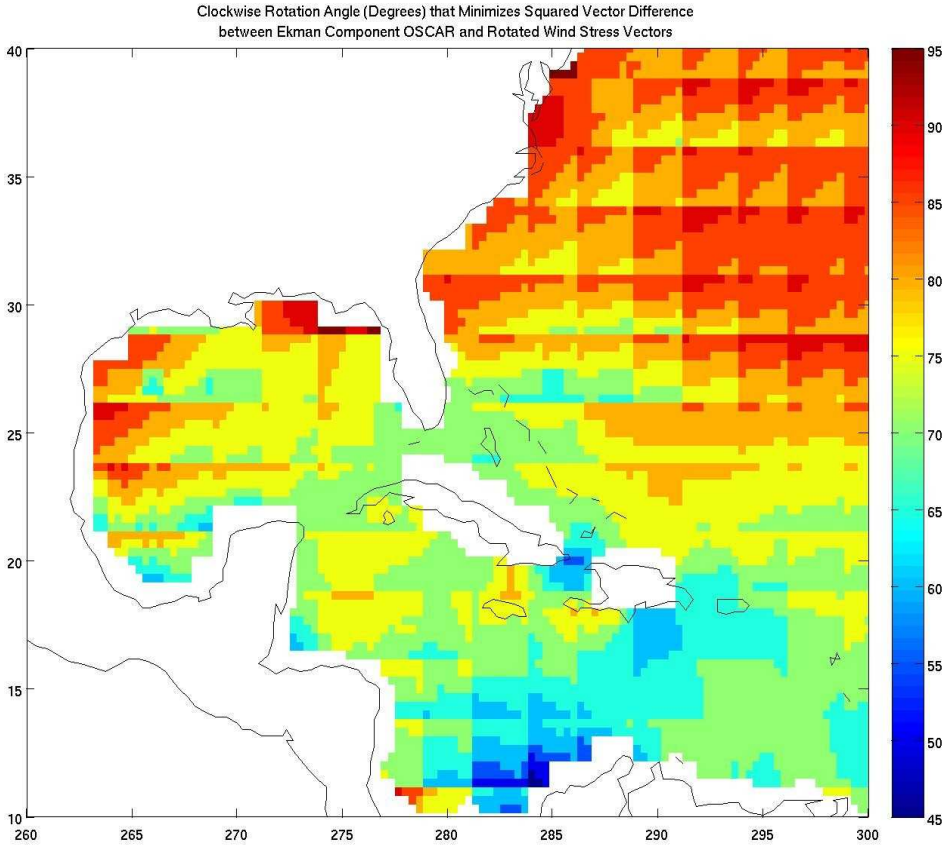


Figure 4-10a. Clockwise rotation angle (degrees) that minimizes the squared difference between the Ekman component OSCAR and rotated wind stress vectors.

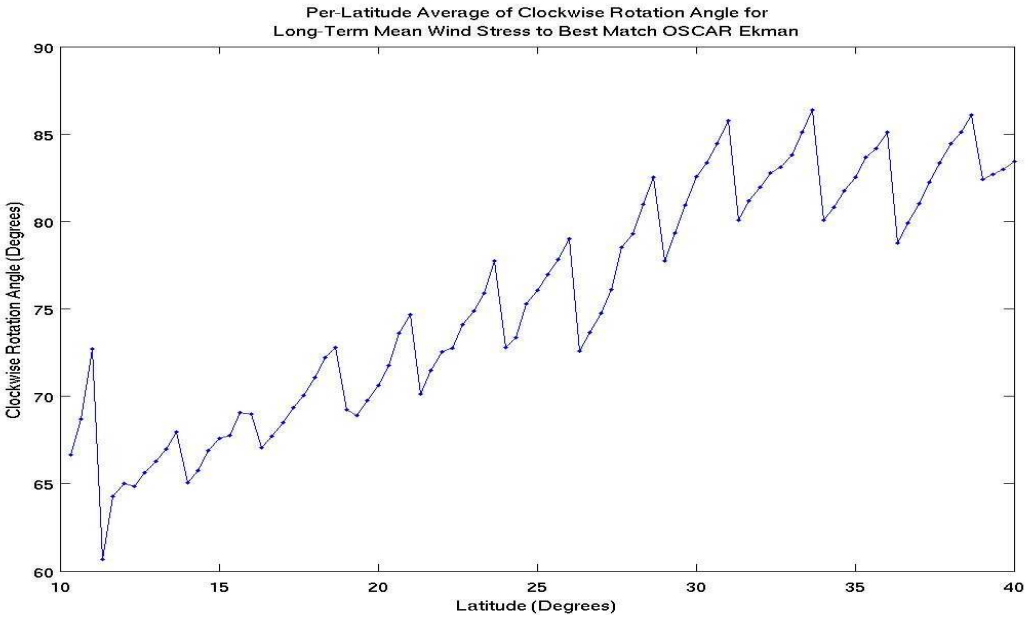


Figure 4-10b. Per latitude averages of the clockwise rotation angle (degrees) from Figure 4-10a. Note the 1/f dependence from the classical formulation of Ekman flow from wind stress. A direct relationship is seen here because the rotation is measured clockwise and is negative rotation.

4.2.2 Component Contributions to Total OSCAR

The OSCAR product has three major components: the geostrophic, Ekman and thermal (also called buoyancy or shear) terms as described in the formulations of Bonjean and Lagerloef 2002 discussed in Section 1.2.2. The relative contributions of each of these components should be understood in order to characterize circulation features in OSCAR. The ratio of the standard deviation about the mean for each of the three components with total OSCAR and the ratio of the standard deviation about the mean for the Ekman to the geostrophic component are mapped in Figure 4-11.

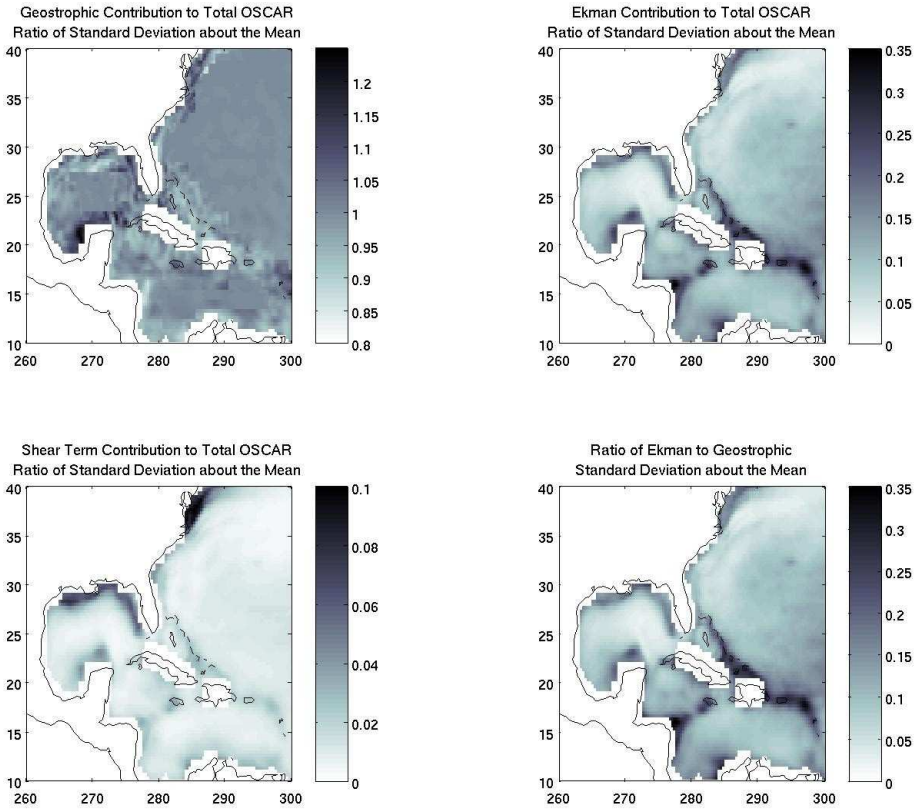


Figure 4-11. Ratio of the standard deviation about the mean for (top left) geostrophic OSCAR to total OSCAR, (top right) Ekman OSCAR to total OSCAR, (bottom left) thermal component OSCAR to total OSCAR and (bottom right) Ekman OSCAR to geostrophic OSCAR.

The geostrophic to total OSCAR plot (top left) shows that OSCAR accounts for at least 80% of the variability in the total OSCAR in most locations. The similarity of the mean geostrophic field to the mean total OSCAR field has already shown it accounts for most of the mean flow. The Ekman to total OSCAR (top right) shows highest contribution to the variability in areas near to land and/or shallow water areas. In these areas, the Ekman component contributes in excess of 20% to the variability about the mean compared to the total OSCAR. The thermal term to total OSCAR (bottom left) contributes less than 10% of variability in most places in the IAS. Areas contributing about 10% are very near to land confined to the northern Gulf of Mexico and the mid-Atlantic coast of the U.S.

The ratio of the standard deviation about the mean for Ekman to geostrophic OSCAR shows that there is little difference when the thermal term is not included. For further comparisons, the thermal term is included in the geostrophic component of OSCAR. The purpose of these maps is to help identify what areas should be searched for moored or shipboard ADCP velocity data so the importance of the Ekman component being added to the geostrophic to produce the total OSCAR can be determined.

4.3 The Importance of Adding the Ekman Component

4.3.1 Comparison with Ship Data Based on Contribution to Total Variability

Comparison locations are chosen by identifying in-situ velocity data that fall in the areas where the ratio of the standard deviation about the mean for Ekman to total OSCAR is > 0.2. There are several ship tracks that run through the area to the northeast of the Greater Antilles through the Bahamas. OSCAR points that match the location and time of a ship data point that falls where the Ekman contribution to the total variability is "large" (ratio of the standard deviation about the mean > 0.2) are plotted against the ship data values in Figure 4-12a and 4-12b. The ship versus OSCAR Ekman speed (top), OSCAR geostrophic speed (middle) and OSCAR total speed (bottom) are in Figure 4-12a. The Ekman speed itself approximately an order of magnitude smaller than the ship data speed at any given space-time location. Comparing the ship data versus geostrophic and total speed plots shows that there is no marked improvement seen here in the agreement of the values with the addition of the Ekman to the geostrophic. The corresponding directions for each set of vector data and OSCAR points are in Figure 4-12b. Looking at the ship data versus geostrophic OSCAR directions (middle), there is some hint at agreement when the directions are in the range from approximately -50° to

50°. Now looking at the ship data versus Ekman OSCAR directions (top), it is obvious that there is something going on with the OSCAR points chosen, as there directions are all tightly clustered around ~110°.

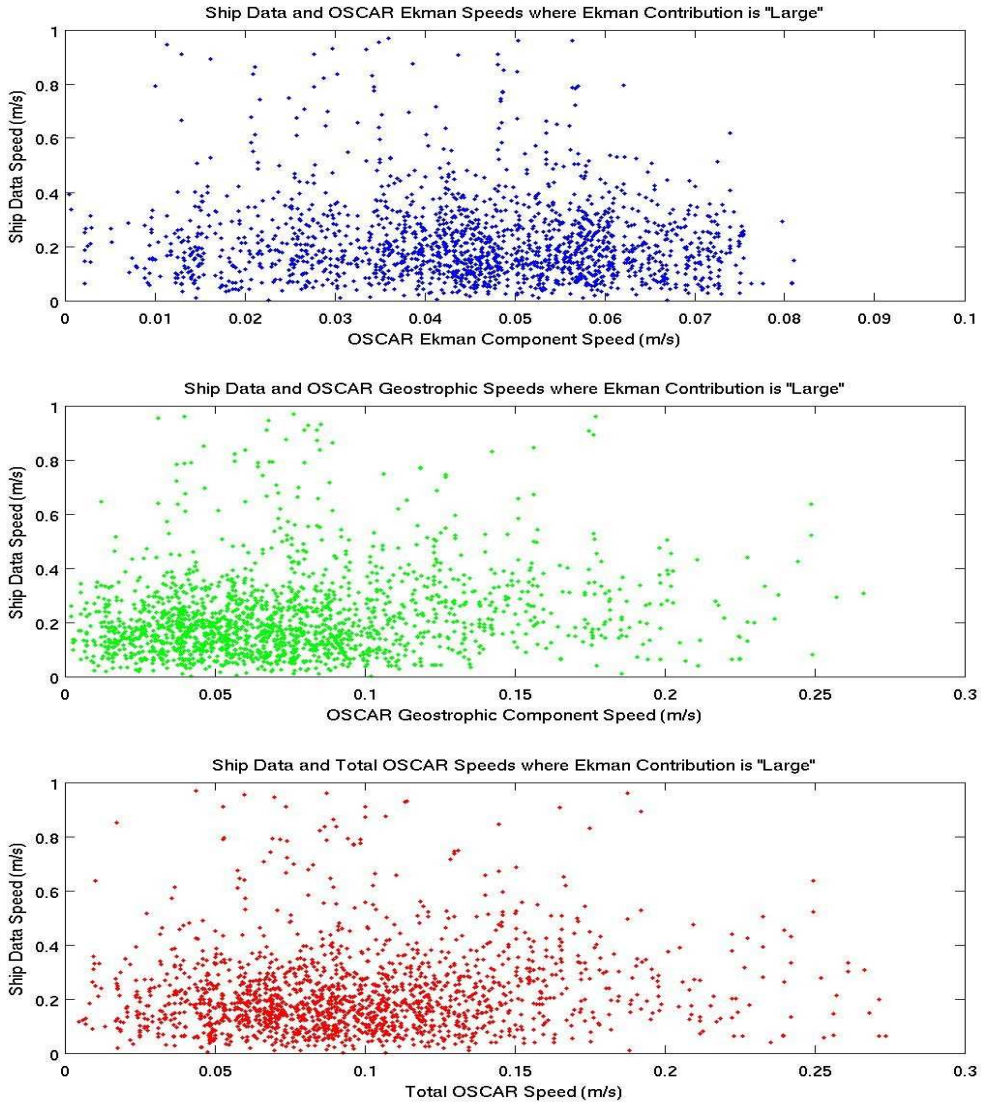


Figure 4-12a. Ship versus OSCAR Ekman speed (top), OSCAR geostrophic speed (middle) and OSCAR total speed (bottom) matched to shipboard ADCP data locations falling in the area where the Ekman contribution to the total variability is "large" (ratio of the standard deviation about the mean > 0.2).

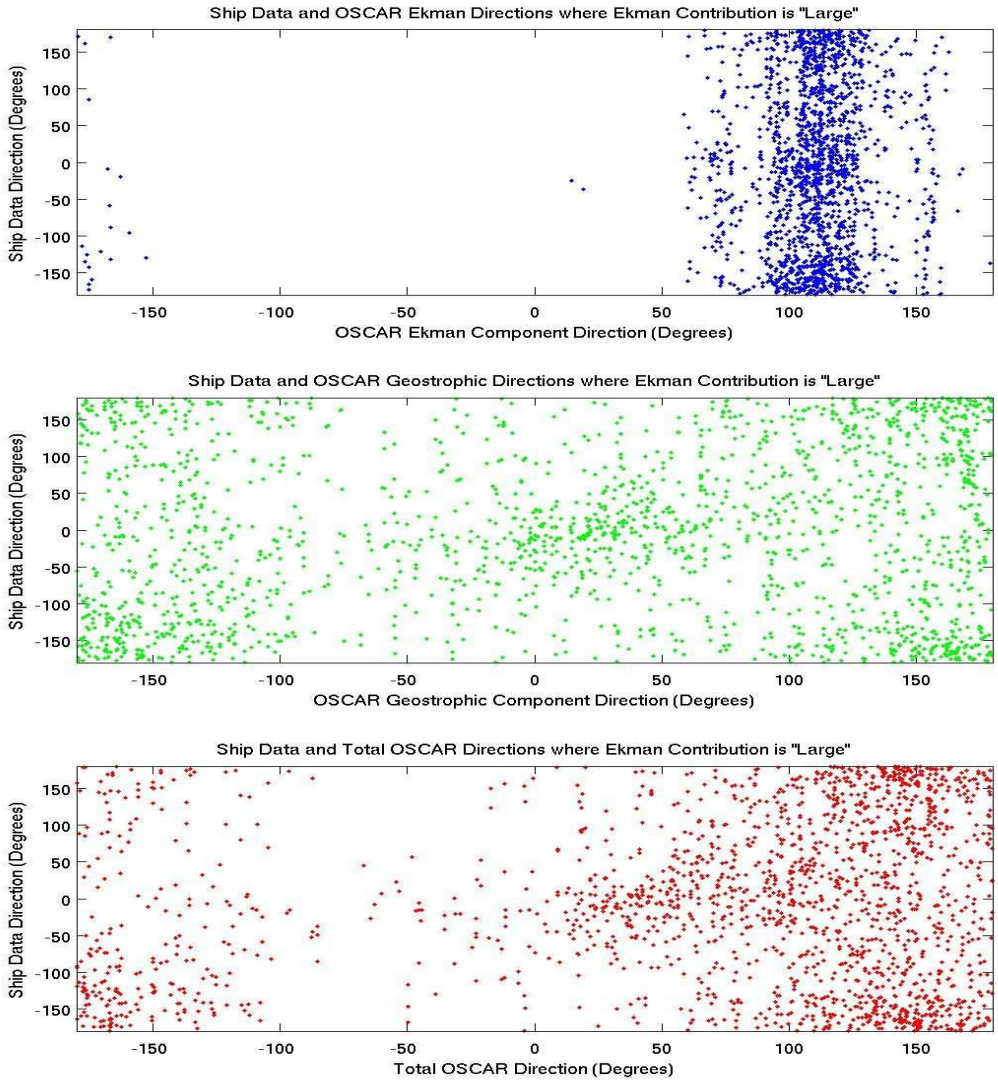


Figure 4-12b. Ship versus OSCAR Ekman direction (top), OSCAR geostrophic direction (middle) and OSCAR total direction (bottom) matched to shipboard ADCP data locations falling in the area where the Ekman contribution to the total variability is "large" (ratio of the standard deviation about the mean > 0.2).

The properties of the distribution of the OSCAR Ekman component velocities are explored to so the Ekman field used for OSCAR can be characterized. The distribution of the directions of all OSCAR Ekman component velocities in the IAS is shown in Figure 4-13a and that of the directions of the corresponding anomalies (time mean removed) is shown in Figure 4-13b. The most common direction for the Ekman

velocities in the IAS is centered at ~110°, the value seen for the OSCAR Ekman points matched in space and time to the ship velocity data points. The directions of the Ekman velocity anomalies are much more evenly distributed across most angle values. If the time-mean velocity vector mostly points in one direction and the anomalies are more randomly distributed, it could mean that the OSCAR Ekman component field does not change much in time and effectively is adding a mean offset to the geostrophic component to create the total OSCAR.

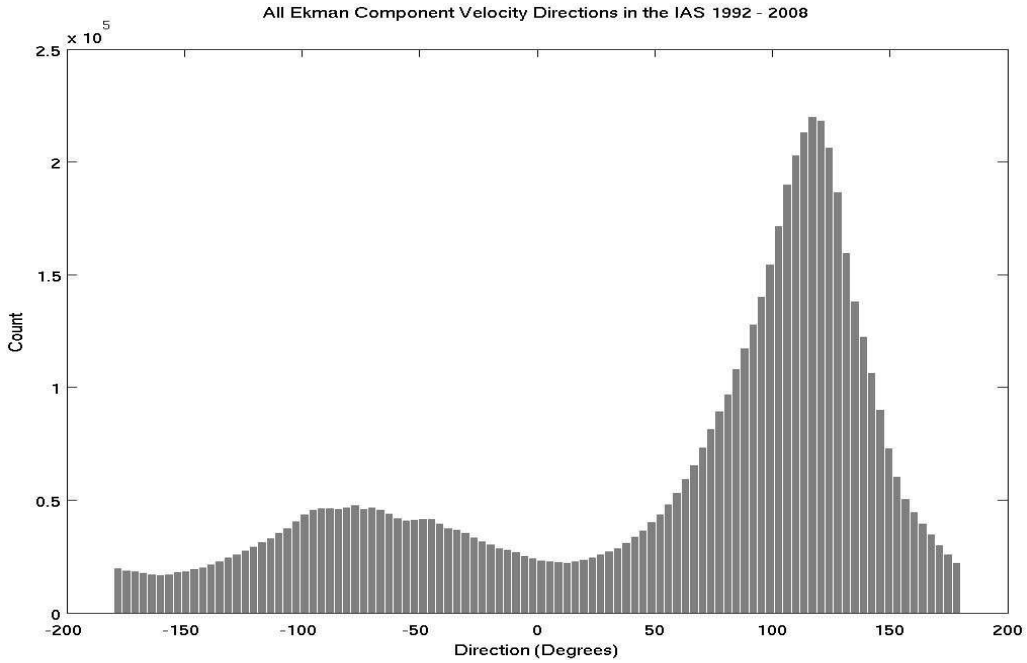


Figure 4-13a. Distribution of the directions of all OSCAR Ekman component velocities in the IAS.

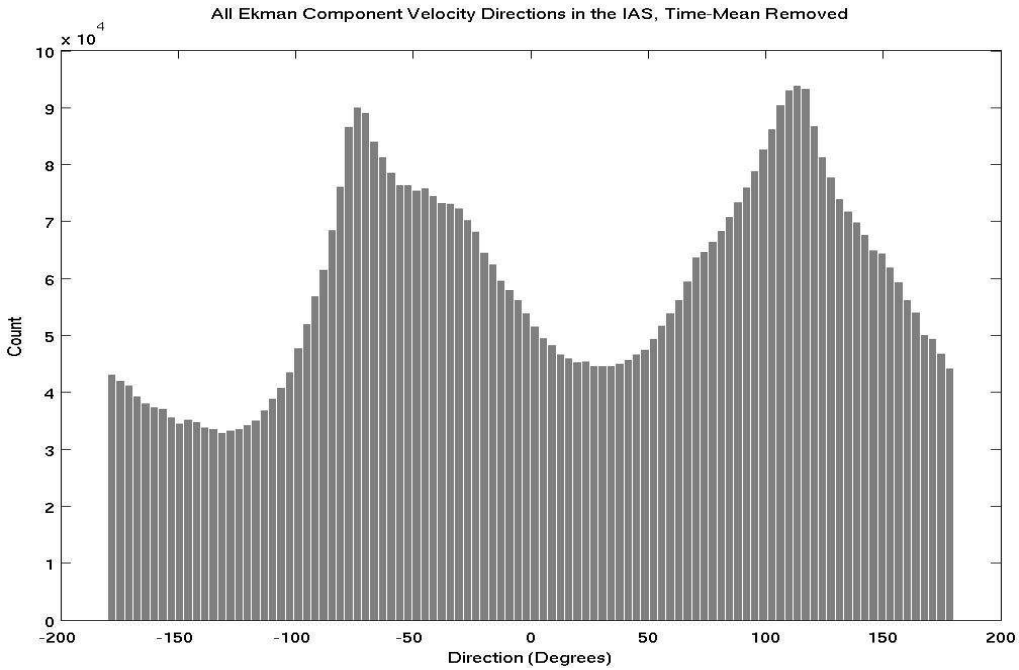


Figure 4-13b. Distribution of the directions of all OSCAR Ekman component velocity anomalies (time-mean removed) in the IAS.

To independently confirm these properties of OSCAR Ekman, the NCEP wind field in the IAS is revisited. The distribution of the directions in the long-term mean wind stress (computed from NCEP winds 1948 – 2004) shows a clustering around a particular direction (Figure 4-14a) while the distribution of the directions of the anomalies at all times for the wind stress is more evenly spread among the directions (Figure 4-14b). Both of these are consistent with the behavior of the Ekman directions.

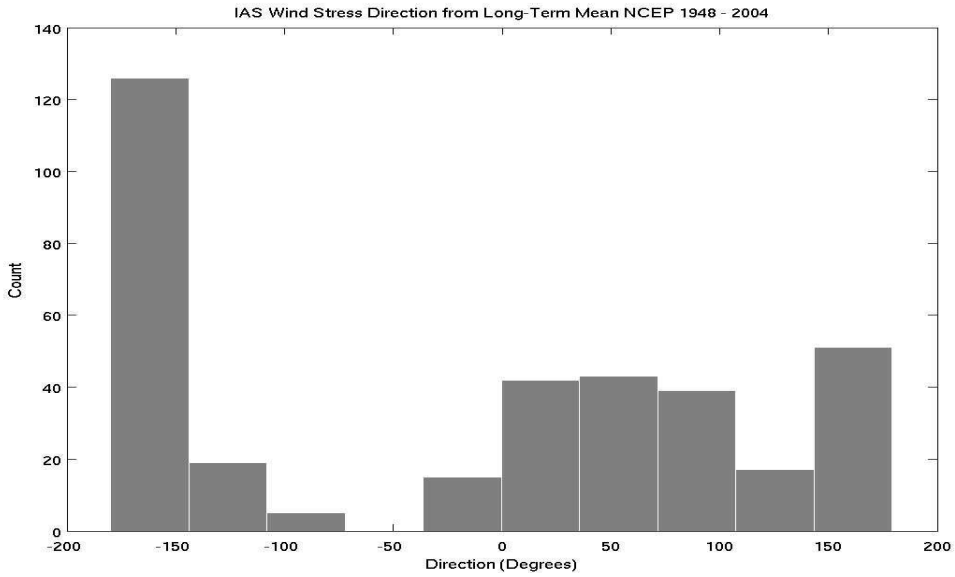


Figure 4-14a. Distribution of the directions in the long-term mean wind stress (computed from NCEP winds 1948 – 2004) in the IAS.

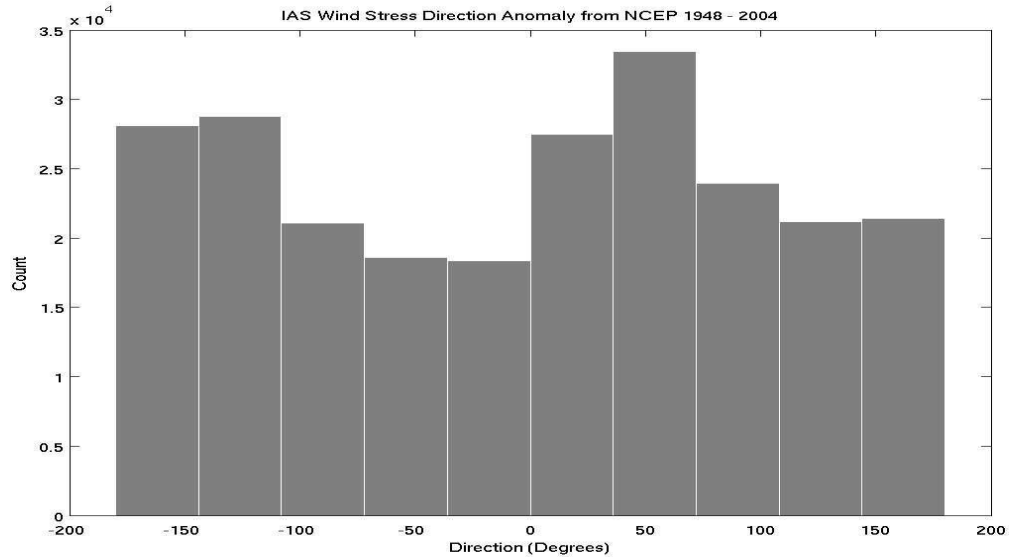


Figure 4-14b. Distribution of the directions wind stress anomalies (time-mean removed) in the IAS.

A map of the OSCAR Ekman time-mean and standard deviation of the anomaly vectors (Figure 4-15) shows that the magnitude of the variability in the anomalies is comparable to the magnitude of the mean in the area of the points matching ship data where the Ekman contribution to the total variability is "large" (ratio of the standard deviation about the mean > 0.2, northeast of the Greater Antilles), so a lack of variability

in the Ekman time-mean field does not explain the clustering of the directions of the Ekman component velocities when matched with the available ship data points.

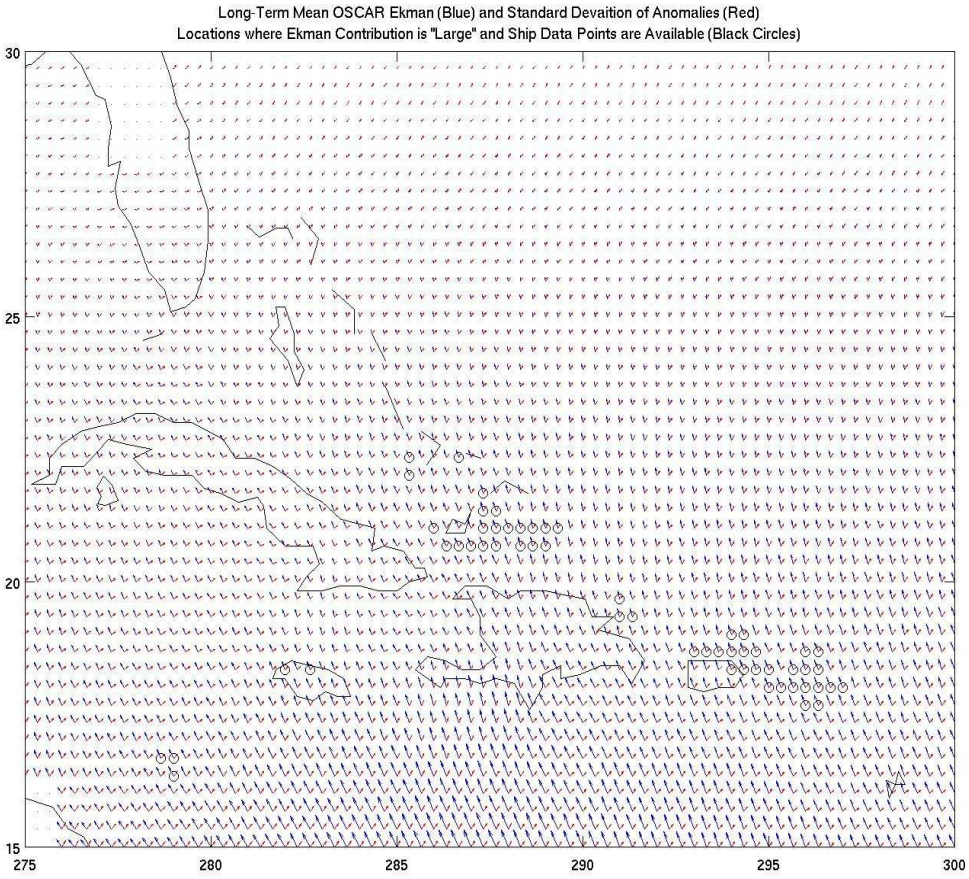


Figure 4-15. Map of the OSCAR Ekman time-mean and standard deviation of the anomaly vectors (time-mean removed). The magnitude of the variability in the anomalies is comparable to the magnitude of the mean in the area of the points matching ship data where the Ekman contribution to the total variability is "large" (ratio of the standard deviation about the mean > 0.2).

Looking at the times of the ship data points themselves, a possible explanation takes shape. Figure 4-16 shows the distribution of the month number during which a ship data point occurs. There is a temporal bias in the shipboard ADCP measurements in this area to the spring and summer months. To show how this affects the distributions of the corresponding OSCAR Ekman component directions, the distribution of direction at OSCAR locations where a ship data point is available, the Ekman contribution to the

variability is “large” and for all OSCAR times (Figure 4-17a), at OSCAR locations where a ship data point is available, the Ekman contribution to the variability is “large” and for OSCAR times corresponding to ship data times (Figure 4-17b) and at OSCAR locations where a ship data point is available, the Ekman contribution to the variability is “large” and only for OSCAR times corresponding to ship data times in the spring and summer months April through August (Figure 4-17c). This progression toward a more and more specific choice of OSCAR Ekman component points corresponding to ship data points in this region shows why the directions are so tightly clustered at those locations and times —ship data points are biased toward the spring and summer and OSCAR Ekman component direction in those months is biased to ~110°.

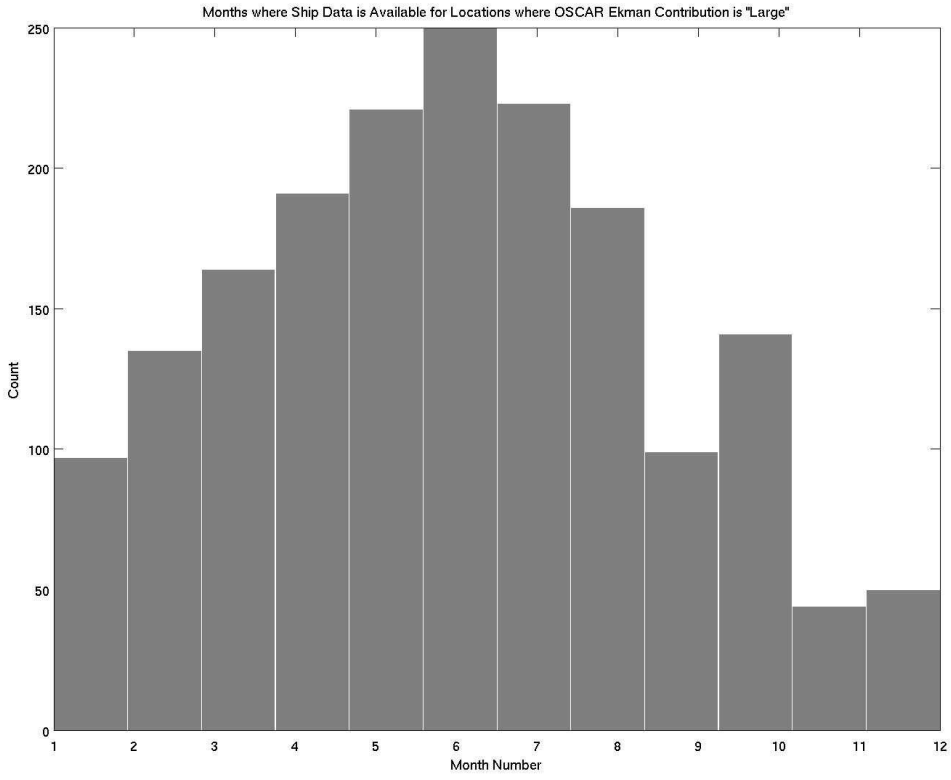


Figure 4-16. Distribution of the month number during which a ship data point occurs where the Ekman contribution to the total variability is “large” (ratio of the standard deviation about the mean > 0.2). Measurements are biased to spring and summer months.

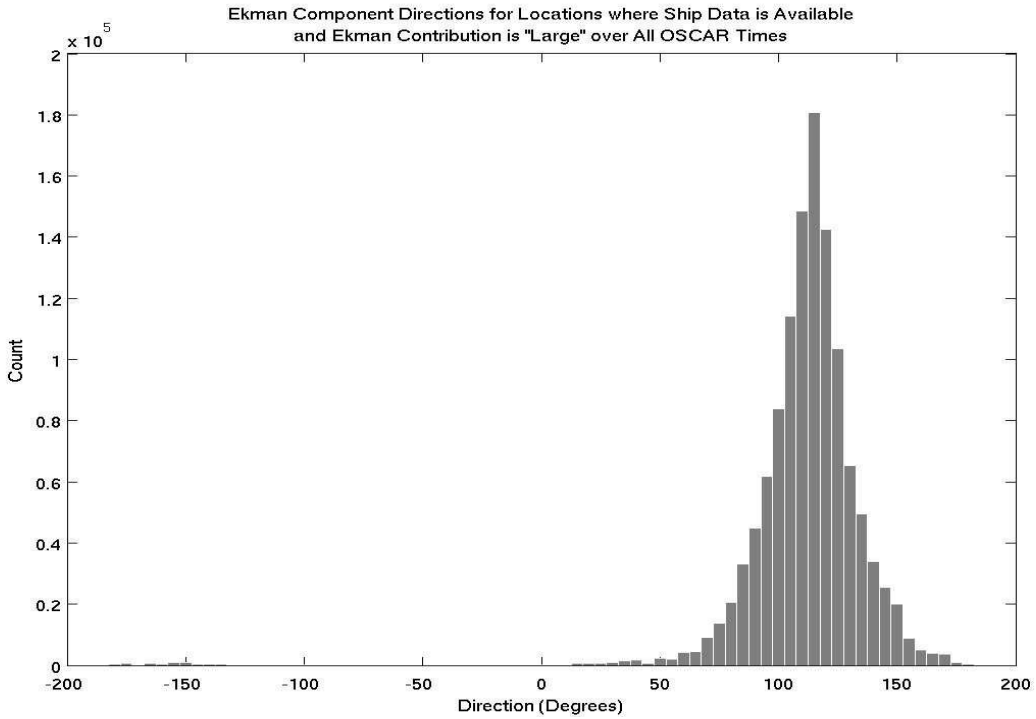


Figure 4-17a. Direction at OSCAR locations where a ship data point is available, the Ekman contribution to the variability is "large" and for all OSCAR times.

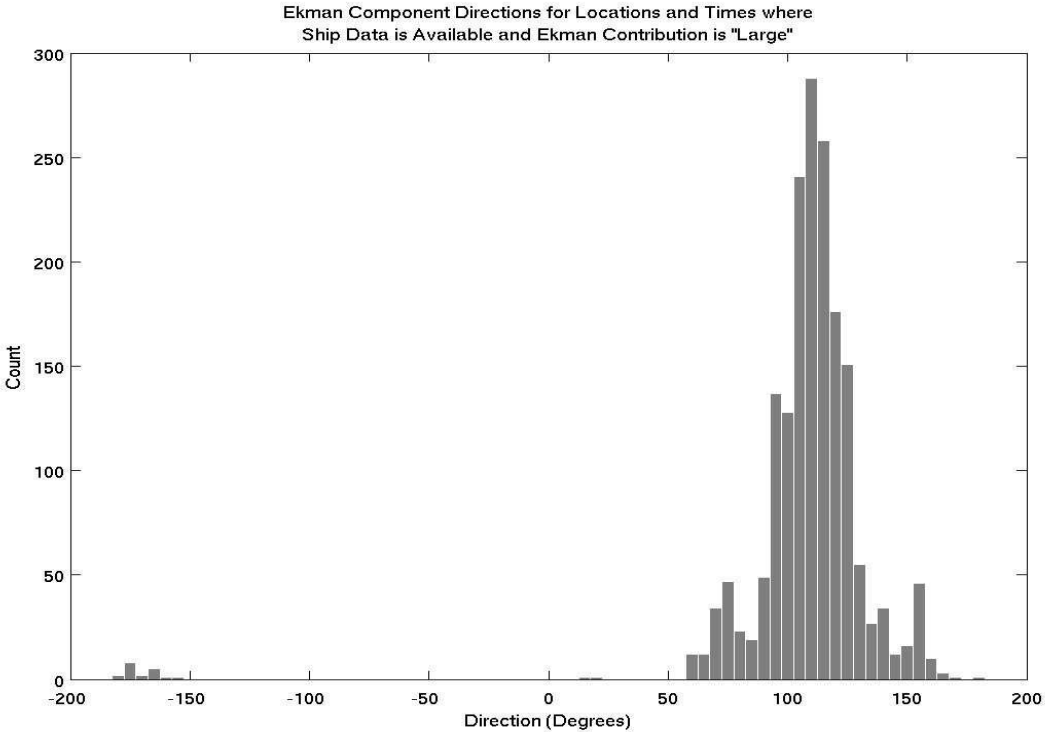


Figure 4-17b. Direction at OSCAR locations where a ship data point is available, the Ekman contribution to the variability is "large" and for OSCAR times corresponding to ship data times.

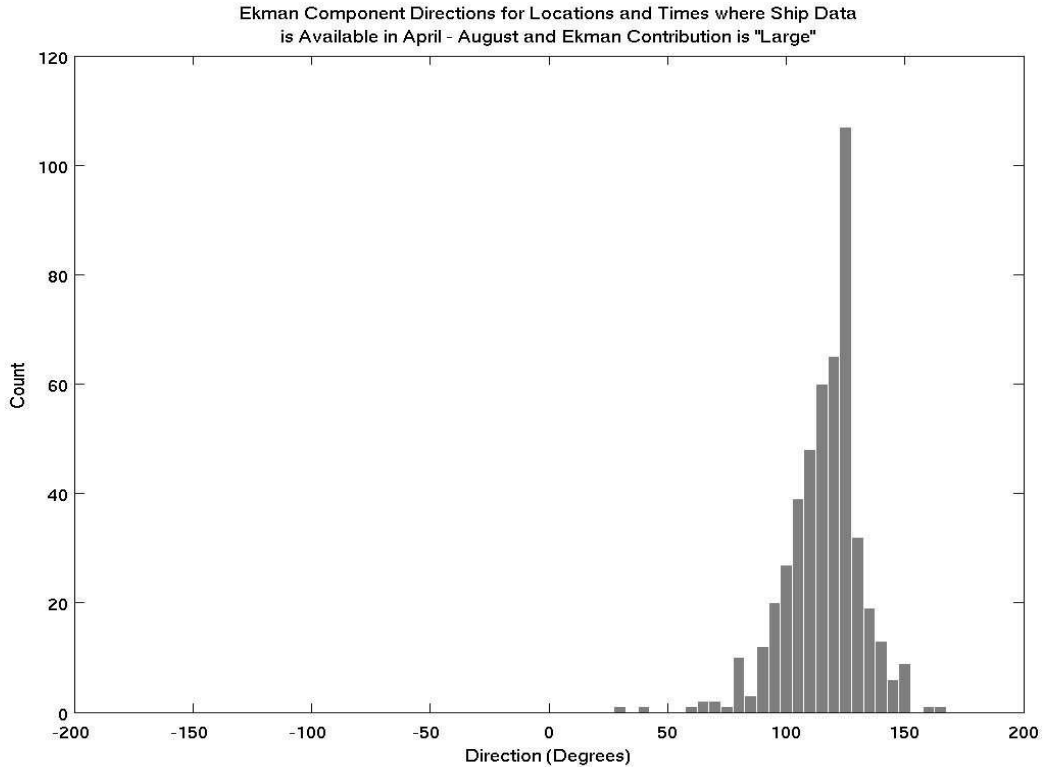


Figure 4-17c. Direction at OSCAR locations where a ship data point is available, the Ekman contribution to the variability is "large" and only for OSCAR times corresponding to ship data times in the spring and summer months April through August. Ship data points are biased toward the spring and summer and OSCAR Ekman component direction in those months is biased toward ~110°.

4.3.2 Comparison with Shipboard and Moored ADCP Data

Comparison locations to determine if adding the Ekman component makes OSCAR output velocities agree better with measured velocities any better than the geostrophic component alone, two additional quantities are computed. The first is the ratio of the Ekman component and total long-term mean speeds. The distribution (top of Figure 4-18) is divided into thirds, color-coded and mapped in the IAS (bottom of Figure 4-18). The second is the ratio of the Ekman component long-term mean speed and the total OSCAR standard deviation (no mean removed). The distribution (top of Figure 4-19) is mapped in grayscale (bottom of Figure 4-19) with the values for small, medium and large Ekman to total long-term mean speeds superimposed. Two criteria are set for

choosing ship data tracks that will be directly compared to instantaneous geostrophic component only and total OSCAR because they will ensure that these are areas where the Ekman contribution are most important: the ratio of the Ekman component and total long-term mean speeds is “large” (> 0.19) and the ratio of the Ekman component long-term mean speed and the (no mean removed) total OSCAR standard deviation is > 0.15.

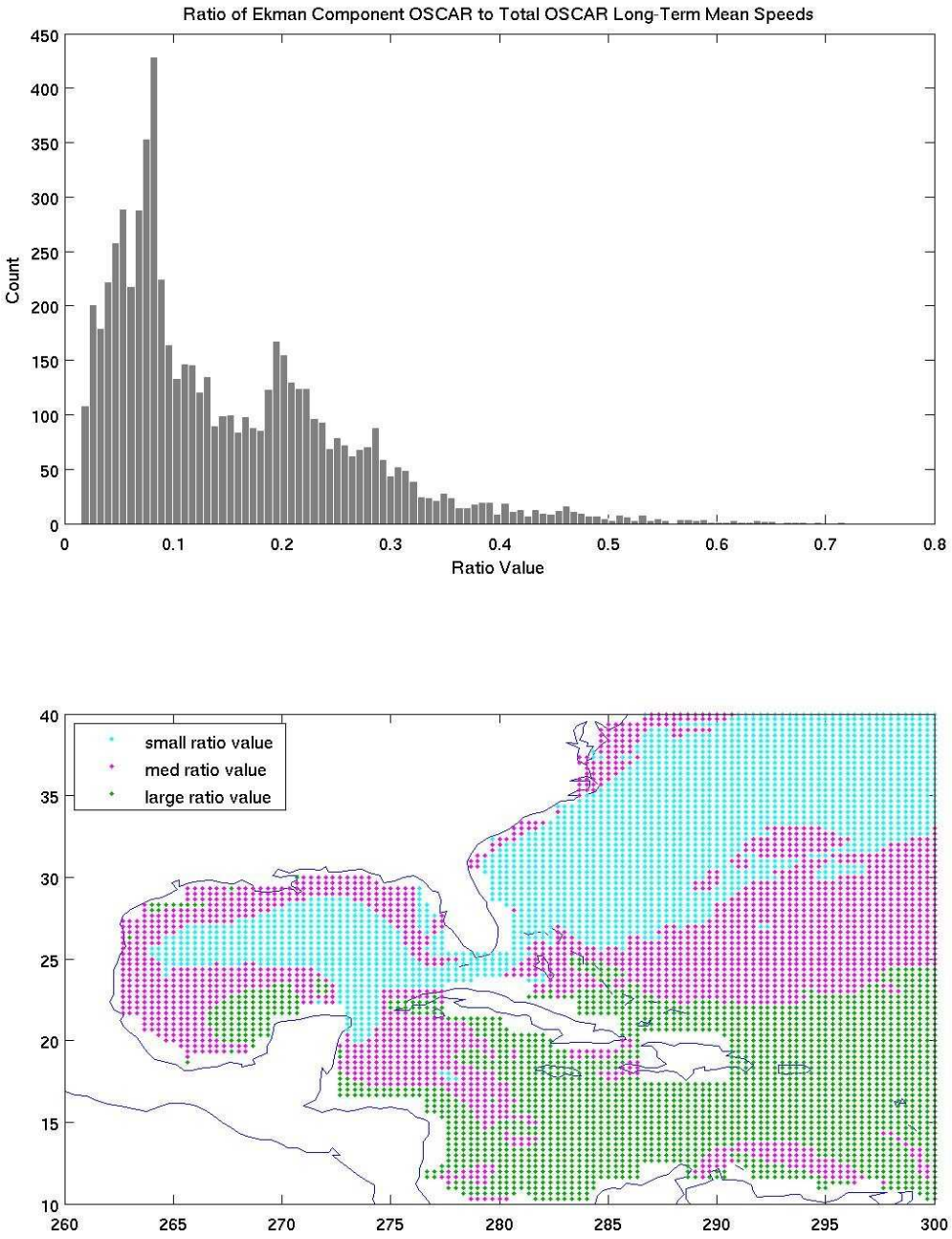


Figure 4-18. Distribution of the ratio of the Ekman component and total long-term mean speeds (top) divided into thirds, color-coded and mapped in the IAS (bottom).

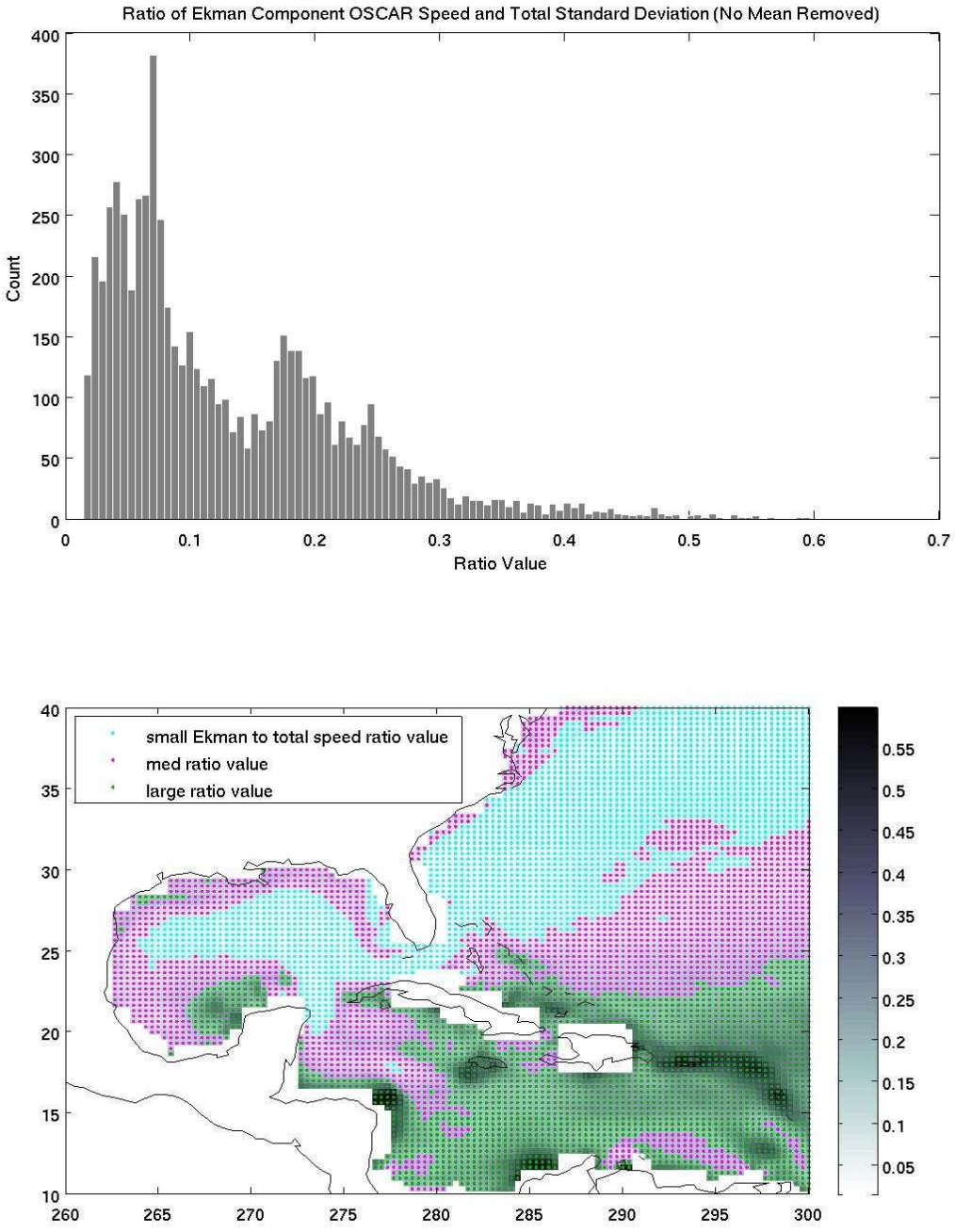


Figure 4-19. Distribution of the ratio of the Ekman component long-term mean speed and the total OSCAR standard deviation with no mean removed (top) mapped in grayscale (bottom) with the values for small, medium and large Ekman to total long-term mean speeds from Figure 4-18 superimposed.

The scatter plot of the instantaneous speeds chosen by these criteria is Figure 4-20. The geostrophic OSCAR points are gray and the total OSCAR points red. By eye, it appears

there is some improvement when the Ekman component is added to the geostrophic for a better agreement with the ship data values.

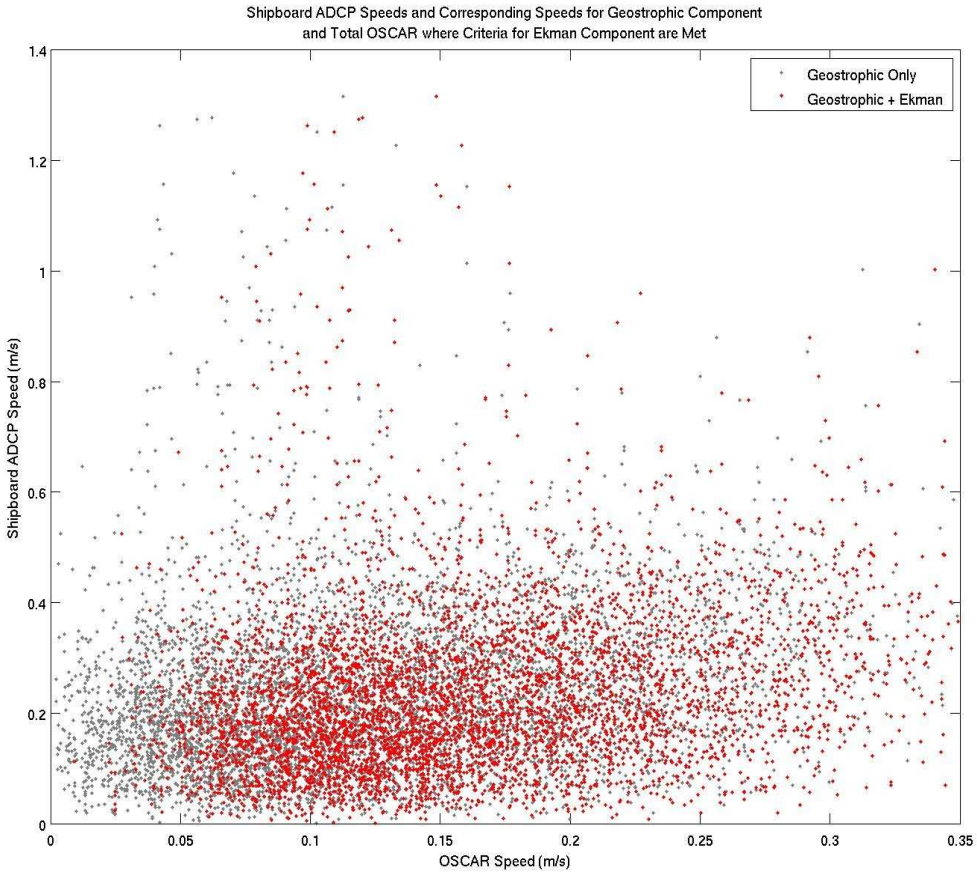


Figure 4-20. Instantaneous geostrophic-only and total OSCAR and shipboard ADCP speeds chosen by the selection criteria ratio of the Ekman component and total long-term mean speeds is "large" (> 0.19) and ratio of the Ekman component long-term mean speed and the (no mean removed) total OSCAR standard deviation is > 0.15.

To quantify the improvement, some comparison values are computed. The average difference between u and v components of each ship track and corresponding geostrophic only and total OSCAR velocities are found and averaged over all of the selected tracks. The error in these values is found by computing the standard deviation of the series of differences for all of the tracks and dividing by the square root of the total number of selected tracks, 260. This is an estimate of the number of degrees of freedom in the average of the differences. The results for the ship track data compared to geostrophic

only OSCAR are:

$$\overline{\delta_{gu}} = 0.0080 \pm 0.0092\ m/s \quad \overline{\delta_{gv}} = 0.0434 \mp 0.0086\ m/s\ .$$

The results for the ship track data compared to the total OSCAR are:

$$\overline{\delta_{tu}} = 0.0220 \mp 0.0092\ m/s \quad \overline{\delta_{tv}} = 0.0080 \mp 0.0085\ m/s\ .$$

Because the average of the differences for the geostrophic u component an the total v component are similar in magnitude and because the error on both the geostrophic and total for the u and v components are of similar magnitude, it is easy to see that there is a slight improvement in using the total OSCAR over the geostrophic only OSCAR to compare to data from ship tracks that fall into the geographic areas where the Ekman contribution is important as defined by the selection criteria.

While the given uncertainty in the mean difference is not true at any individual point, it does estimate the systematic errors in this area. Beal et al. 2008 estimated their bias error during a single Florida current crossing due to heading errors to be ~ 4 cm/s, but our calculation in this geographical area suggests that the bias error here is much lower. The ship likely spent much more time under steady steam along this track relative to the time spent maneuvering (see Section 3.8.1 for an explanation of heading errors during maneuvering and under steady steam). Beal et al. 2008 only used one crossing while the bias error here was computed using several tracks, possibly averaging out some intermittent bias errors. In a time-mean sense, the ocean velocities in this area tend to be oriented in a direction similar to (or directly opposite to) the direction of the ship's forward motion. As per the discussion in Section 3.8.1, bubble contamination is probably the largest source of error in this area. Bubble contamination errors in single-ping data points were addressed by those at RSMAS providing the data and should not be much of

a factor, however, bubble contamination biases due to rough sea conditions do remain in the data and are probably contributing to the bias error in the mean difference estimated here.

The moored ADCP velocity data locations are in places where the selection criteria are not met. Each series of moored velocity data is matched to the OSCAR series that best corresponds in time and location. It is not necessarily expected that the latitude and longitude locations in a gridded product such as OSCAR will match up with the actual locations of the data. To remedy this, the four OSCAR series nearest to the data location are selected and the best-matched series will be determined by vector correlation squared (see Section 3.6.1). The data series must also be low-pass filtered so that their spectral characteristics are similar to those of the OSCAR series. In the analysis in Chapter 4, a typical high-frequency roll-off value for OSCAR at the moored ADCP data locations is ~ 30 days. A filter is designed to be applied to the data series to remove frequencies higher than ~ 30 days. The time means of the u and v components of the filtered data series and the geostrophic-only and total OSCAR are computed for each data location. To place error bars on these quantities, an estimate of the degrees of freedom for each series must be made. The degrees of freedom for each series used in the comparison is estimated by

$$dof \approx dof_s * \left(\frac{N}{N_s} \right) * p_{frac}$$

where dof_s is the number of degrees of freedom for the longest continuous sub-section of the data series, N is the total series length, N_s is the length of the longest continuous sub-section of the data series and p_{frac} is the fraction complete for each data series. Note that

if a series has no gaps, dof = dof_s. The quantity used for the comparison is the magnitude of the vector difference for the time mean data series and the geostrophic-only and total OSCAR. Using propagation of error through each of those quantities, the error is computed. Figure 4-21 shows the magnitudes of the vector difference between geostrophic-only OSCAR and total OSCAR with the moored ADCP velocity along with a one-sigma error bar. The error bars overlap for all of the series used in the comparison and neither set of values is different from zero within 2-sigma. Adding the Ekman component to geostrophic OSCAR does not improve the agreement with velocity data over the geostrophic component of OSCAR alone.

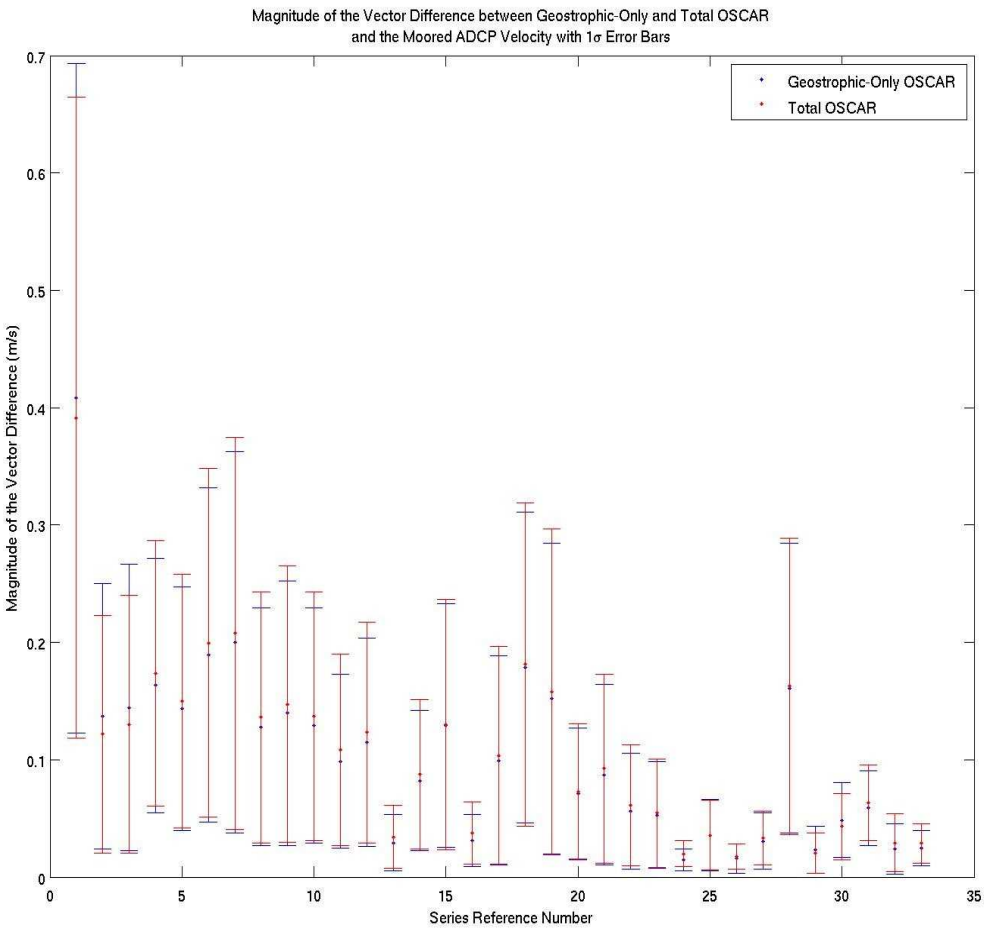


Figure 4-21. Magnitudes of the vector difference (m/s) between geostrophic-only OSCAR and total OSCAR and the moored ADCP velocity along with a one-sigma error bar for the series used in this comparison. The error bars overlap for all of the series and neither set of values is different from zero within 2-sigma.

5 Summary of Results and Concluding Remarks

5.1 Real-Time Height Fields from Altimetry

Proposed improvements for the second-generation OSCAR product were to expand to a global, real-time product with the use of available satellite-derived datasets for heights and winds and to provide a method of creating heights for an operational product accessible on the internet. The USF contribution to OSCAR identifies and combines available gridded altimeter products using a predictive method to bring the height fields to real-time and provides a daily product that runs automatically and includes quality checks. The current available version of OSCAR is global but is provided with a temporal resolution of every 5 days and is therefore considered near real-time.

This thesis details the methods used to create the height fields and evaluate the current OSCAR product in the complex region of the Intra-American Sea. The outcome of this thesis shows it is possible to take two different gridded altimetry products and place them on the same datum (with remaining differences noted in areas that cannot be helped) and use that combined result to apply a predictive method that out-performs persistence through the 10-day span deemed necessary. The daily height fields with automatic checks developed by USF have been incorporated into the operational OSCAR system at ESR.

5.2 Evaluation of OSCAR

The comparison of gridded altimeter heights from AVISO and tide gauge measurements show the series agree reasonably well over frequencies longer than about 1/45 days. At higher frequencies, the altimetry product is unable to reproduce the energy captured by the tide gauges due to the time window and spatial scales over which the product is effectively smoothed during creation of the optimally interpolated gridded product. Exceptions to the agreement at lower frequencies include tidal aliasing frequencies where the AVISO product is deficient in a relatively narrow frequency band corresponding to those frequencies. Since a global tidal model is used to remove the energy aliased by the altimeters from tides, in many locations where tide gauges tend to placed, local tides may differ greatly from the tidal model. The removal of the tidal aliasing energy is also apparently removing some energy not associated with the tidal aliasing from the altimetry data at these locations. Parsing out the identified tidal aliasing frequencies by altimeter and geographic locations shows T/P data is the most prominent over the IAS with the exception of the Gulf of Mexico, where AVISO is biased toward ERS data.

The geographical distribution of the high-frequency roll-off values indicates that the amount of spatial smoothing in AVISO varies with location. This is an important result because ocean dynamics being studied at those higher frequencies will be discernible in some areas and not others strictly due to the processing of the gridded altimetry product and not related to physical differences from one area to another. Because the AVISO DM product is used to create the geostrophic current component of OSCAR, these issues are propagated into OSCAR.

For the moored ADCP locations, the depth of the measurement does not appear to make a difference in the series' agreement with OSCAR. This first shows that there is not a lot of vertical shear in this area of the Gulf of Mexico over the upper 100 meters. This result also suggests that the mean Ekman component is not a significant contributor to the total mean velocity in that geographical area. OSCAR, representing the near-surface where the Ekman component should be the largest in magnitude with depth, also has an Ekman component contribution that is relatively small in this area compared to others in the IAS.

OSCAR velocity series are least correlated with measured velocities in shallow water and near to land. Some of the lowest correlation values found for shallow water station comparisons can be accounted for simply by being near to land. All stations nearest to land have low correlation values. Other low correlations in shallow water but farther from land are possibly attributable to dynamics occurring over higher frequencies than OSCAR can reproduce. Both of these deficiencies in the OSCAR product are directly related to the use of altimetry to compute the geostrophic velocity component of OSCAR.

The rotary auto-spectral analysis provides values for HFRO for OSCAR compared to data series with a range and distribution of values consistent with the tide gauge to AVISO comparison. Like that comparison, the HFRO values vary with geographic location, again indicating that the amount of effective spatial smoothing in OSCAR is not constant. Rotary cross-spectral analysis of co-rotating components for OSCAR and the moored velocity data shows the two are coherent over frequencies corresponding to periods of monthly and longer. This is consistent with the tide gauge to

AVISO comparison and shows the limitation of OSCAR in producing periodically varying signals on monthly or longer timescales due to the smoothing done in the processing.

While the moored velocity comparisons showed OSCAR does not do well nearest to land and in shallow water, the shipboard ADCP velocity comparisons show that effect is overshadowed by OSCAR's inability to reproduce the highest-magnitude speeds in the IAS along a path from the Yucatan to Miami where the current is narrow. This occurs because when the flow is dominated by the geostrophic component as it is along this path, the spatial smoothing used to create the gridded altimetry product from along-track data comes into play. The interpolated geostrophic speeds derived for the gridded height fields are bounded by the values on either side of the grid space being interpolated. The result is decreased magnitudes of velocity from OSCAR compared to shipboard ADCP data that is measuring values from inside the narrow flow. The comparison did show that OSCAR is able to reproduce the direction of the flow as it agrees reasonably well with the direction of the measured velocities. It is reasonable to suggest that OSCAR will have problems accurately representing the magnitude of any current that is relatively narrow compared to the OSCAR grid size (spatial smoothing scale).

The importance of adding the Ekman component to the geostrophic to create total OSCAR is gauged by the relative size of Ekman contribution to the time mean total current. In geographical areas where the mean Ekman contribution is most important relative to the mean total OSCAR, a slight but statistically significant improvement in the ability of OSCAR to reproduce velocities in agreement with in-situ data over a comparison to geostrophic-only OSCAR can be discerned. In geographical areas where

Ekman is not as important, no improvement in agreement of total OSCAR to in-situ data versus geostrophic-only OSCAR is detectible to within error bars. In neither of these cases could OSCAR be considered having the same values as the measured velocities (indicated by a mean difference equal to zero) within errors.

5.3 Concluding Remarks

The results from this analysis show that OSCAR has a variety of problems in accurately reproducing currents in the IAS. Many of these issues are tied to the smoothing in OSCAR that is introduced into the product by the AVISO gridded altimetry and the spatial smoothing of the height gradient field in processing. Some issues are a result of inaccuracies related to altimetry when nearest to land or in shallow water. This highlights the need for better and more observing and modeling systems for the oceans, providing in-situ data with adequate spatial and temporal coverage that can supplement the use of gridded altimetry when necessary.

ESR continues to produce OSCAR with a temporal resolution of every 5 days based on averages of daily product runs. The OSCAR publicly accessible product recently updated from 1-degree spatial resolution to finer third-degree. The most up-to-date information on OSCAR and access to the product can be found at http://www.oscar.noaa.gov/.

6 List of References

Alvera-Azcárate, A., A. Barth and R. H. Weisberg, The Surface Circulation of the Caribbean Sea and the Gulf of Mexico as inferred from Satellite Altimetry, *J. Phys. Oceanogr.*, *39*, 640-657, 2009.

Beal, L.M., J.M. Hummon, E. Williams, O.B. Brown, W. Baringer and E.J. Kearns, Five Years of Florida Current Structure and Transport from the Royal Caribbean Cruise Ship *Explorer of the Seas*, *J. Geophys. Res., 113, C06001*, 2008.

Bloomfield, P., *Fourier Analysis of Time Series: An Introduction*, 2nd Edition, pp. 224-226, John Wiley & Sons, Inc., 2000.

Bonjean, F. and G. S. E. Lagerloef, Diagnostic Model and Analysis of the Surface Currents in the Tropical Pacific Ocean, *J. Phys. Oceanogr.*, *32*, 2938-2957, 2002.

Centurioni, L. R. and P. P. Niiler, On the Surface Currents of the Caribbean Sea, *Geophys. Res. Letters*, *30 (6)*, 1279-1282, 2003.

Crosby, D. S., L. C. Breaker and W. H. Gemmill, A Proposed Definition for Vector Correlation in Geophysics: Theory and Application, *J. Atmos. And Oceanic Tech.*, *10*, 355-367, 1993.

Dohan, K., G. S. E. Lagerloef and J. Gunn, *OSCAR-2 Ocean Surface Current Analysis Real-time: The next generation data products and services*, prepared at Earth and Space Research, Seattle, WA, 2008.

Emery, W. J. and R. E. Thomson, *Data Analysis Methods in Physical Oceanography*, 2nd Ed., pp. 425-432 and 494-500, Elsevier B. V., 2004.

Jacobs, G. A., C. N. Barron, D. N. Fox, K. R. Whitmer, S. Klingenberger, D. May and J. P. Blaha, Operational Altimeter Sea Level Products, *Oceanography, 15 (1)*, 13-21, 2002.

Johnson, E. S., F. Bonjean, G. S. E. Lagerloef, J. T. Gunn, G. T. Mitchum, Validation and Error Analysis of OSCAR Sea Surface Currents, *J. Atmos. and Oceanic Tech., 24*, 688-701, 2007.

Lagerloef, G. S. E., G. T. Mitchum, R. B. Lukas and P. P. Niiler, Tropical Pacific near-surface currents estimated from altimeter, wind and drifter data, *J. Geophys. Res., 104*, 23, 313-23, 326, 1999.

Mooers, C. N. K., A Technique for the Cross Spectrum Analysis of Pairs of Complex-Valued Time Series, with Emphasis on Properties of Polarized Components and Rotational Invariants, *Deep-Sea Res., 20*, 1129-1141, 1973.

National Snow and Ice Data Center. "All About Sea Ice: 'Characteristics: Arctic vs. Antarctic'." http://nsidc.org/seaice/characteristics/difference.html. Accessed 08 August 2011.

Numerical Recipes in FORTRAN 77: The Art of Scientific Computing, 1st Edition, pp. 444-446, Cambridge University Press, 1986.

Pratt, R. W. and G. A. Maul, Sea Surface Height Variability of the Intra-Americas Sea from Topex/Poseidon Satellite Atimetry: 1992-1995, *Bull. Of Marine Sci., 67 (2)*, 687-708, 2000.

Richardson, P. L., Caribbean Current and Eddies as Observed by Surface Drifters, *Deep-Sea Res. II, 52*, 429-463, 2005.

SSALTO/DUACS User Handbook: (M)SLA and (M)ADT Near-Real Time and Delayed Time Products, version 1, rev. 8, 2007.