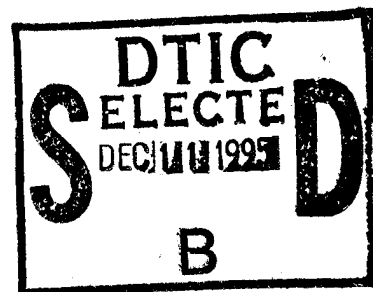


Report No. CG-D-34-95

THE MARIANO GLOBAL SURFACE VELOCITY ANALYSIS 1.0

A. J. Mariano
E. H. Ryan
Rosenstiel School of Marine and Atmospheric Science
University of Miami
4600 Rickenbacker Causeway
Miami, FL 33149
and
Analysis and Technology, Inc.
P.O. Box 220
North Stonington, CT 06359



B.D. Perkins and S. Smithers
U.S. Coast Guard Research and Development Center
1082 Shennecossett Road
Groton, CT 06340-6096



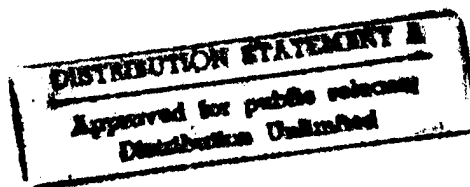
FINAL REPORT
JULY 1995

19951206 094

This document is available to the U.S. public through the
National Technical Information Service, Springfield, Virginia 22161

Prepared for:

U.S. Department of Transportation
United States Coast Guard
Office of Engineering, Logistics, and Development
Washington, DC 20593-0001



DTIC QUALITY INSPECTED 1

NOTICE

This document is disseminated under the sponsorship of the Department of Transportation in the interest of information exchange. The United States Government assumes no liability for its contents or use thereof.

The United States Government does not endorse products or manufacturers. Trade or manufacturers' names appear herein solely because they are considered essential to the object of this report.

The contents of this report reflect the views of the Coast Guard Research & Development Center. This report does not constitute a standard, specification, or regulation.



G. T. Gunther
Technical Director, Acting
United States Coast Guard
Research & Development Center
1082 Shennecossett Road
Groton, CT 06340-6096

Technical Report Documentation Page

1. Report No. CG-D-34-95		2. Government Accession No.		3. Recipient's Catalog No.	
4. Title and Subtitle The Mariano Global Surface Velocity Analysis 1.0				5. Report Date July 1995	
				6. Performing Organization Code	
7. Author(s) A. J. Mariano, E.H. Ryan, B.D. Perkins, and S. Smithers				8. Performing Organization Report No. R&DC 34/95	
9. Performing Organization Name and Address Rosenstiel School of Marine and Atmospheric Science University of Miami 4600 Rickenbacker Causeway Miami, FL 33149 Analysis & Technology, Inc. P.O. Box 220 North Stonington, CT 06359				10. Work Unit No. (TRAIS)	
				11. Contract or Grant No.	
				13. Type of Report and Period Covered Final Report	
12. Sponsoring Agency Name and Address U.S. Coast Guard Research and Development Center 1082 Shennecossett Road Groton, Connecticut 06340-6096 Department of Transportation U.S. Coast Guard Office of Engineering, Logistics, and Development Washington, D.C. 20593-0001				14. Sponsoring Agency Code Commandant (G-NRS) USCG Headquarters Washington, D.C. 20593-0001	
15. Supplementary Notes The R&DC technical point of contact is LCDR B.D. Perkins.					
16. Abstract This report details the construction of the Mariano Global Surface Velocity Analysis (MGSVA) 1.0 from the Maury Ship Drift Data. The generalized median filter and the parameter matrix objective analysis algorithm used in the analysis are detailed. The applicability and limitations of the MGSVA are discussed, and specific recommendations are made for using the MGSVA for U.S. Coast Guard Search and Rescue operations.					
17. Key Words surface velocity search surface current CASP objective analysis search planning combining velocity field Monte Carlo climatological environmental			18. Distribution Statement Document is available to the U.S. public through the National Technical Information Service, Springfield, Virginia 22161		
19. Security Classif. (of this report) UNCLASSIFIED		20. SECURITY CLASSIF. (of this page) UNCLASSIFIED		21. No. of Pages 55	
				22. Price	

METRIC CONVERSION FACTORS

Approximate Conversions to Metric Measures				Approximate Conversions from Metric Measures			
Symbol	When You Know	Multiply By	To Find	Symbol	When You Know	Multiply By	To Find
LENGTH				LENGTH			
in	inches	* 2.5	centimeters	mm	millimeters	0.04	inches
ft	feet	30	centimeters	cm	centimeters	0.4	inches
yd	yards	0.9	meters	m	meters	3.3	feet
mi	miles	1.6	kilometers	km	kilometers	1.1	yards
						0.6	miles
AREA				AREA			
in ²	square inches	6.5	square centimeters	cm ²	square centimeters	0.16	square inches
ft ²	square feet	0.09	square meters	m ²	square meters	1.2	square yards
yd ²	square yards	0.8	square meters	km ²	square kilometers	0.4	square miles
mi ²	square miles	2.6	square kilometers	ha	hectares (10,000 m ²)	2.5	acres
	acres	0.4	hectares				
MASS (WEIGHT)				MASS (WEIGHT)			
oz	ounces	28	grams	g	grams	0.035	ounces
lb	pounds	0.45	kilograms	kg	kilograms	2.2	pounds
	short tons (2000 lb)	0.9	tonnes	t	tonnes (1000 kg)	1.1	short tons
VOLUME				VOLUME			
tsp	teaspoons	5	milliliters	ml	milliliters	0.03	fluid ounces
tbsp	tablespoons	15	milliliters	l	liters	0.125	cups
fl oz	fluid ounces	30	milliliters	l	liters	2.1	pints
c	cups	0.24	liters	l	liters	1.08	quarts
pt	pints	0.47	liters	l	liters	0.26	gallons
qt	quarts	0.95	liters	m ³	cubic meters	35	cubic feet
gal	gallons	3.8	liters	m ³	cubic meters	1.3	cubic yards
ft ³	cubic feet	0.03	cubic meters				
yd ³	cubic yards	0.76	cubic meters				
TEMPERATURE (EXACT)				TEMPERATURE (EXACT)			
°F	Fahrenheit temperature	5/9 (after subtracting 32)	Celsius temperature	°C	Celsius temperature	9/5 (then add 32)	Fahrenheit temperature

* 1 in = 2.54 (exactly).

TABLE OF CONTENTS

	<u>Page</u>
LIST OF ILLUSTRATIONS	vi
LIST OF TABLES	vi
1. INTRODUCTION	1
2. MAURY SHIP DRIFT DATA	2
3. METHODOLOGY	5
3.1. GENERALIZED MEDIAN FILTER	5
3.2. OBJECTIVE ANALYSIS	8
4. MARIANO GLOBAL SURFACE VELOCITY ANALYSIS 1.0	18
5. LIMITATIONS AND RECOMMENDATIONS	47
ACKNOWLEDGEMENTS	50
BIBLIOGRAPHY	51
ABBREVIATIONS AND ACRONYMS	55

Accession For	
NTIS GRA&I	☒
DTIC TAB	☐
Unannounced	☐
Justification	
By _____	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	4

LIST OF ILLUSTRATIONS

<u>Figure</u>	<u>Page</u>
1a. An OA Map of the Florida Current Region Using the Unedited Data	6
1b. The Same OA Map After Removing Data Outliers Before the OA.....	6
2a. The Velocity Scale as a Function of Latitude (The software for the curly vector plots was developed at the Navy Research Lab South.)	19
2b. The OA Map for the Atlantic Ocean for the Month of May	20
2c. The Corresponding Error Map (Note that most of the basin, except for the sub-polar and polar regions, has low estimation error.).....	21
3a. The OA Map for the Pacific Ocean for the Month of December	22
3b. The Corresponding Error Map (Note that the North Pacific has much lower estimation error than the South Pacific.)	24
4a-d. The Weighted Annual Average of the Monthly Surface Velocity Estimates (The weights were inversely proportional to the error estimates associated with each monthly estimate.).....	25
5a-d. The Weighted Seasonal Average of the Monthly Surface Velocity Estimates for the Months of December, January, and February	29
6a-d. The Weighted Seasonal Average of the Monthly Surface Velocity Estimates for the Months of March, April, and May	33
7a-d. The Weighted Seasonal Average of the Monthly Surface Velocity Estimates for the Months of June, July, and August.....	37
8a-d. The Weighted Seasonal Average of the Monthly Surface Velocity Estimates for the Months of September, October, and November.....	41
9a. The Histogram of the East-West Velocity Component, u , Differences Between the Atlantic Minus the Indian Basins	45
9b. Same as (a) But for v	45
9c. A Histogram of the Speed Differences, $\sqrt{(u_{Atlantic} - u_{Indian})^2 + (v_{Atlantic} - v_{Indian})^2}$	45
9d. A Polar Plot of the Vector Differences Between the Atlantic and Indian Basin Velocities	45
10a. An Estimate of the Streamfunction Field for the Gulf Stream Region.....	48
10b. Same as (a) But Another Field Estimate.....	48
10c. Using the Classic Optimal Estimator for Combining Fields From (a) and (b).....	48
10d. Using Contour Analysis.....	48

LIST OF TABLES

<u>Table</u>		<u>Page</u>
1. Number of Edited Ship-Drift Observations for the Three Major Basins.....		3

1. Introduction

One of the hardest problems in all of physical oceanography is real-time Lagrangian prediction. Accurate, reliable, real-time Lagrangian prediction of the movement of overboard people, distressed ships, downed aircraft, drifting icebergs, pollutants, etc. in ocean surface waters is crucial for U.S. Coast Guard (USCG) activities related to Search and Rescue (SAR), International Ice Patrol, and environmental protection. In particular, SAR missions are global and require global environmental data sets for planning optimal search strategies.

Accurate and contemporaneous estimates of environmental variables such as wave height, wind speeds, and, most importantly, sea surface velocity are required to achieve the highest probability of a successful rescue. However, in many situations such data are not available and historical estimates and their associated variances are used in computer models for identifying regions of highest probability of containment. These computer models require historical sea surface velocity and their variances on a regular grid.

Two recent studies (Perkins and Marsee, 1994; Viekman and Jordan, 1991) in the Florida Current region, Gulf Stream, and California Current region have documented serious deficiencies in the present historical data base used in SAR operations. The major problems are that (1) the surface velocities in the present operational data base are too slow in the major currents analyzed and (2) the velocity fields are too smooth. For instance, Viekman and Jordan (1991) calculated a 40-90 *cm/s* difference in zonal and meridional velocity components between the historical data base and surface drifters. Both studies recommended that a new historical data base be constructed, for USCG SAR operational use, from ship drift data archived by the U.S. Navy. This report details the construction of a new SAR historical data base from the Maury Ship Drift Data.

The largest historical data base of sea surface velocity is the Maury Ship Drift Data (MSDD) compiled by the U.S. Naval Oceanographic Office (NAVOCEANO). These data, described in Section 2, (1) have an irregular distribution in space and time; (2) have large measurement errors; and (3) contain extreme unrealistic data outliers that must be removed. Section 3 contains a description of a generalized median filter for removing data outliers and an objective analysis (OA) technique that was used for space/time interpolation of the irregularly distributed data onto a regular grid. Examples of the analysis velocity fields, and the seasonal and annual averages of the sea surface velocity analysis are presented in Section 4. The limitations of the MSDD and our analysis methodology and recommendations for future improvement of sea surface velocity estimates and applications for SAR activities are discussed in Section 5.

2. Maury Ship Drift Data

The sea surface velocity data base used here consists of historical ship drift measurements obtained from NAVOCEANO. This data base, known as the Maury Ship Drift Data, is named after Lt. Matthew Fontaine Maury who, during the years 1842-1861, inaugurated the tabulation of ship drift from historical merchant logbooks and a hydrographic reporting program among shipmasters (Bowditch, 1984). This information has been used to compile pilot charts under the auspices of NAVOCEANO (1955, 1966, 1979a, 1979b, 1979c).

Ship drift is computed as the vector difference between dead-reckoning position and the actual position over a 12-24 hour period. Averaging over 24 hours practically eliminates tides, inertial motion, and small-scale ocean variability, and yields an estimate that resolves oceanic scales on the order of a few hundred kilometers (Wyrтки *et al.*, 1976; Richardson and McKee, 1984).

Strictly speaking, drift velocities differ from true sea surface velocities mainly because of windage effects on the ships. The differences between these two quantities are detailed in Section 3.2 when errors are assigned to the ship drift measurements. Nevertheless, a number of oceanographic studies have used this data set to estimate surface mean and eddy kinetic energy levels (Wyrтки *et al.*, 1976; Richardson and McKee, 1984), to study seasonal circulation patterns (Stidd, 1975; Meehl, 1982; Richardson and Reverdin, 1987), and to evaluate numerical models (Richardson and Philander, 1987; Arnault, 1987). These studies and the comparison between ship drift, buoys, and current meters by McPhaden *et al.* (1991) clearly demonstrate that ship drift-based velocity estimates are a surprisingly good indicator of mean sea surface velocities and their seasonal variations.

The data base used here, acquired from NAVOCEANO by USCG Research and Development Center (R&D Ctr), was last updated November 1991. The number of observations is 4.16 million ship drift estimates. There may be up to 0.7 million missing observations from this data base because of probable accidental deletion during file transfers (Richardson, 1989). Most of the missing observations (63 percent) are probably in the data-sparse South Pacific. Finding the missing data and adding the new data, especially surface drifters (see Section 5), should be given high priority by the USCG R&D Ctr.

Data from 1900 to 1991 are used in this analysis. Navigation errors were significantly reduced during the early part this century through a combination of improved optics in the sextants and more stable chronometers. Comprehensive statistics for this data base are given in Richardson (1989) and references therein. Table 1 presents the number of edited observations for each month for each of three major basins.

Table 1. Number of Edited Ship-Drift Observations for Three Major Basins

	<u>Atlantic</u>	<u>Indian</u>	<u>Pacific</u>
<i>January</i>	147077	48307	70596
<i>February</i>	99752	38351	54918
<i>March</i>	165768	48008	109170
<i>April</i>	109285	37518	70534
<i>May</i>	182938	42725	94765
<i>June</i>	88293	41457	65195
<i>July</i>	163220	41640	54801
<i>August</i>	174074	42551	57113
<i>September</i>	164577	41942	54276
<i>October</i>	120494	39441	54510
<i>November</i>	149695	38689	84438
<i>December</i>	77015	40980	49562

Note in Table 1 the low number of observations for June and December for the Atlantic Ocean. Richardson (1989) also noted this, and argued for probable deletion or omission of the data file in the North Atlantic. In addition, Richardson noted in the Pacific data that there were a small number of observations for the months of February, April, June-October, and December for the North Pacific and that only three relatively data-rich months, March, April, and June, are available in the South Pacific.

A monthly temporal resolution was chosen for our estimates. This frequency allows for adequate resolution of the important semi-annual and annual components of the flow field and barely resolves the circulation features associated with the energetic 50-60 day wind events. The important 1-10 day weather scale was not resolved and should be modeled as the random component in the trajectory simulations. Higher temporal resolution would significantly lower the confidence in the estimates and would create many data-void gaps, especially in the southern oceans. Instead of the twelve monthly estimates, only one annual estimate was made south of 50° S because of data sparsity. In general, the North Atlantic has the highest data density and, consequently, the most reliable estimates, and the South Pacific has the lowest data density and the least reliable estimates.

The monthly data sets were also binned into twelve basins for the analysis. This binned data set is known as the Monthly MSDD (MMSDD). This was done for two primary reasons: (1) to make the computations more manageable so that they could be done on a workstation; and (2) to prevent artificial leaking of information during the interpolation procedure over land masses. For example, given the 300 km correlation scales used (see Section 3.2), Caribbean Sea velocities would be used in the estimation of Pacific velocities in the region surrounding the Isthmus of Panama, and vice versa if the data were not binned into separate files. Also, velocity data from coastal seas, with much different dynamics and smaller decorrelation scales, would not influence the estimates in nearby oceanic regimes.

The twelve basins are the: Antarctic Ocean, Atlantic Ocean, Baltic Sea, Black Sea, Caribbean Sea, South/East China Sea, Indian Ocean, Sea of Japan, Mediterranean Sea, Pacific Ocean, Persian Gulf, and the Red Sea. Data overlap regions at the boundaries connecting the major basins, such as the region south of the Cape of Good Horn, are needed to ensure that the final field estimates blend together smoothly. Data overlap regions are on the order of $(1000 \text{ km})^2$ for the OA calculations. This area is greater than three times the correlation length scales in each direction (west-east, south-north) so that bigger overlap regions would not add any new information to the estimates at the "seams" where basins meet. A comparison of the estimates at the basin seams is discussed in Section 4.

3. Methodology

As mentioned earlier, the MSDD must be edited to remove outliers and then space-time interpolated onto a regular grid in space and time for USCG operational use. The spatial interpolation grid was chosen by USCG R&D Ctr and is currently the grid in operational use. Monthly estimates for all basins, except Antarctic, at these grid points were determined to be sufficient for present operational use. The next two subsections discuss the generalized median filtering algorithm used for data editing (Section 3.1) and the parameter matrix objective analysis algorithm of Mariano and Brown (1992), hereafter MB92, used for data interpolation (Section 3.2).

3.1 Generalized Median Filter

An area encompassing the Gulf Stream off the coast of Florida was chosen for initial testing because it contained both the well-defined Florida current and because the SAR historical data base was unrealistic in this very important area (Perkins and Marsee, 1994). Fig. 1a is a representative OA map of surface velocities using all the May data for this area. It is quite evident that some of the velocity estimates are poor even after smoothing by the OA scheme (though for this map, the smoothing was kept to a minimum). Inspection of the input data revealed extreme data outliers. The presence of outliers was expected with the MSDD based on previous work by Richardson and colleagues. Fig. 1b shows the OA map after the outliers were removed using a simple procedure that calculated the arithmetic mean and the standard deviations of u and v , and that removed every data point that was more than three standard deviations from the mean. A technique for removing outliers from the basin-wide data subsets was formulated, based on the encouraging results of this initial mapping experiment.

Most methods for removing outliers require a measure of central tendency, $\bar{Q}$, such as the arithmetic mean or median value, and a measure of variation or spread, such as the standard deviation, σ_Q . An envelope of acceptable data values, (Q_1, Q_2) , is constructed by calculating

$$Q_1 = \bar{Q} - a\sigma_Q, \quad (3.1a)$$

$$Q_2 = \bar{Q} + a\sigma_Q, \quad (3.1b)$$

where a is a factor controlling the width of the acceptable data envelope. In many applications, a is 2-3. If the data value Q_i is between Q_1 and Q_2 , viz.

$$Q_1 < Q_i < Q_2,$$

then Q_i is deemed a good data point and is kept for analysis. Otherwise, it falls outside the acceptable data value range and is removed from the analysis.

In many oceanographic applications, including this one, there are a number of practical difficulties in using this simple formula. These difficulties are encountered

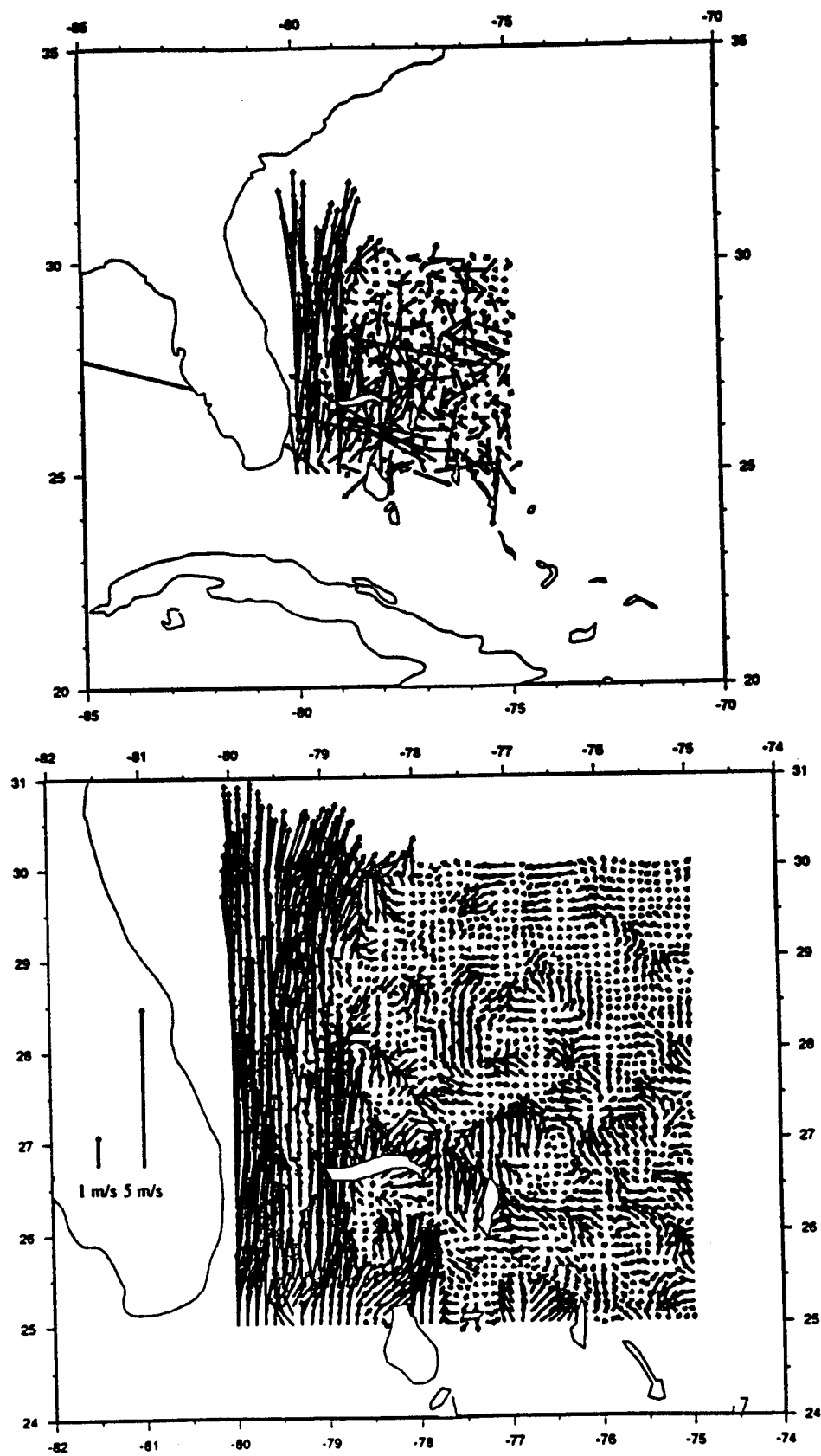


Fig. 1 (a) An OA map of the Florida current region using unedited data. (b) The same OA map after removing data outliers before the OA.

because: (d1) the ocean is energetic over a wide spectrum of scales and its variability is greater than the mean in most places; (d2) the ocean is heterogeneous-the mean flow in the subtropical gyre is a few cm/s while in the Gulf Stream, the flow is 100-250 cm/s ; (d3) the variation in σ is also large, *e.g.*, the surface velocity Eddy Kinetic Energy (EKE) maps of Wyrski *et al.* (1976) vary by a factor of six to seven; and (d4) anomalous ocean events, order of ten standard deviations from the mean, occur (McWilliams *et al.*, 1983).

If the measurement error were small, a large anomaly would be real if measured, and difficulty (d4) would vanish. However, the measurement error for the MMSDD is large and the necessity of removing outliers was demonstrated in Section 2. Thus we must live with (d4) and an algorithm dealing with difficulties (d1-d3) was constructed. This algorithm flags good data as bad data for extreme events, but this is unavoidable with this data set.

In light of (d1-d3), the basin-wide application of eqn. (3.1) requires $\bar{Q}(x, y)$ and $\sigma_Q(x, y)$ for $Q = u, v$. Note that the temporal dependency in these terms is not needed since the data have been binned into months. In general, a could also be a function of x and y , but this was not tried. Empirical determination of a is discussed below.

The MMSDD was binned into $5^\circ \times 5^\circ$ bins for outlier removal and OA calculations. The selection of bin size was a compromise between a small bin size that would better resolve the field and a large bin size so that $\bar{Q}(x, y)$ and $\sigma(x, y)$ are estimated from a large sample size. For each bin k , the standard deviation of u and v , denoted by σ_u and σ_v , respectively, are estimated using the standard formulas

$$\sigma_u = \left(\frac{\sum_{i=1}^n (u_i - \bar{u}_k)^2}{(n-1)} \right)^{1/2},$$

$$\sigma_v = \left(\frac{\sum_{i=1}^n (v_i - \bar{v}_k)^2}{(n-1)} \right)^{1/2},$$

where,

$$\bar{Q} = \sum_{i=1}^n Q_i / n,$$

for $Q = u_k$ and v_k . σ_u and σ_v are assigned to the midpoint of the k^{th} bin. A smooth bicubic spline (Inoue, 1986; MB92) is fitted to all assigned σ_u values and all assigned σ_v values generating two sets of spline coefficients, one for σ_u , one for σ_v , for each of the twelve basins and for each of the twelve months.

The outlier procedure is described next for one basin and one month. For each $5^\circ \times 5^\circ$ bin k , $\bar{u}_k$ and $\bar{v}_k$ were estimated using a median filter since this measure of central tendency is better suited than the arithmetic average for data sets with extreme outliers. Let the position of the i^{th} data point be denoted by (x_i, y_i) . $\sigma_u(x_i, y_i)$ and $\sigma_v(x_i, y_i)$ are calculated by evaluating the bicubic spline for σ_u and σ_v , respectively, at each position (x_i, y_i) . $\bar{u}_k$ and $\bar{v}_k$ for each of the k bins are calculated by taking the median value, defined by ordering the data values,

$$Q_1 \leq Q_2 \leq \dots \leq Q_j \leq \dots \leq Q_{n-1} \leq Q_n,$$

and defining if n is odd

$$j = j + 1 = (n + 1)/2,$$

and if n is even

$$j = n/2, \quad j + 1 = (n + 2)/2.$$

The median value, $\tilde{Q}$, is then defined as

$$\tilde{Q} = 1/2(Q_j + Q_{j+1}).$$

Let s_i denote the speed associated with the i^{th} data point, $s_i = \sqrt{u_i^2 + v_i^2}$. Let $\bar{s}_k$ denote the speed associated with the median values of $\bar{u}_k$ and $\bar{v}_k$, $\bar{s}_k = \sqrt{\bar{u}_k^2 + \bar{v}_k^2}$. Let $\sigma_s = \sqrt{\sigma_u^2 + \sigma_v^2}$ represent the standard deviation of the field evaluated at the point (x_i, y_i) . If

$$\bar{s}_k - a\sigma_s(x_i, y_i) < s_i < \bar{s}_k + a\sigma_s(x_i, y_i),$$

is true then the data point (u_i, v_i) is deemed good and is used in the analysis. It is removed from the analysis if it fails this condition.

The value for a was determined empirically. If a is too small, too many good data points are removed. If a is too large, too many bad data points are retained for analysis. For example, if the velocity components were truly Gaussian distributed, then a value of $a \approx 2$ (3) would remove 5% (1%) of the good data points. A range of a from 1.5-4.0 was tested by viewing the fields and noting that the largest a value contained no obvious outliers. The best value of a , for this data set, was determined to be 3.

3.2 Objective Analysis

Objective Analysis, hereafter OA, is used here for interpolating irregularly spaced asynoptic observations onto a regular grid at a common estimation time for subsequent analysis. OA has been used extensively for mapping oceanic fields (*e.g.*, Bretherton *et al.*, 1976; Freeland and Gould, 1976; Carter, 1983; Davis, 1985; McWilliams *et al.*, 1986; Carter and Robinson, 1987; Robinson *et al.*, 1987; Watts *et al.*, 1989; and Mariano and Brown, 1992). OA has its roots in the analysis of meteorological variables (*e.g.*, Gandin, 1963; Thiébaux, 1974; Ghil *et al.*, 1981; Thiébaux and Pedder, 1987; Daley, 1991). In general, an OA of data requires (1) estimates of the covariance (correlation) function of the variable of interest, (2) the covariance function of the data variable, and (3) the cross-covariance function between the variable of interest and the data. A brief introduction to the theory of OA is presented next in terms of a scalar variable T that represents temperature since

the software was developed for the estimation of global Sea Surface Temperature (SST). Velocity components u and v are mapped independently by the following method outlined for T .

Since a perfectly accurate temperature can never be measured, the i^{th} measurement $\hat{T}_i$ is denoted as

$$\hat{T}_i = T_i + e_i,$$

where e_i is the i^{th} measurement error. $\langle e_i \rangle$ is the average error at the i^{th} location.

Our goal is to find the best estimator of the form

$$\hat{T}_o = \sum_{i=1}^n a_i \hat{T}_i = \sum_{i=1}^n a_i (T_i + e_i), \quad (3.2)$$

that takes into consideration both error and distance information. The usual choice for the criterion of merit for the notion of best is a minimum variance unbiased estimator.

First, we examine whether the estimator is unbiased. The following relation

$$\langle \hat{T}_o \rangle = \langle \sum_{i=1}^n a_i \hat{T}_i \rangle = \sum_{i=1}^n a_i \langle \hat{T}_i \rangle$$

shows that, if $\langle \hat{T}_i \rangle = 0$, then $\langle \hat{T}_o \rangle = 0$ and the estimator is unbiased.

Our estimator will always be unbiased if we consider only the deviations of the i^{th} measurement from its average state,

$$\hat{T}'_i = \hat{T}_i - \langle \hat{T}_i \rangle,$$

because it will always have an average of zero, since by definition

$$\langle \hat{T}'_i \rangle = \langle \hat{T}_i \rangle - \langle \langle \hat{T}_i \rangle \rangle = \langle \hat{T}_i \rangle - \langle \hat{T}_i \rangle = 0.$$

The middle equality uses the fact that, for statistically stationary fields, $\langle \hat{T}_i \rangle$ is a constant (it may be a different constant for different values of i) and the average of a constant is obviously the constant itself. Thus deviations from an average state will always be assumed for any measurement and the prime notation is dropped for convenience. In practice, only an estimate of the mean field is available. The estimated mean field is referred to as the trend.

The variance of our estimator is

$$\langle (\hat{T}_o - \langle \hat{T}_o \rangle)^2 \rangle = \sigma_{\hat{T}_o}^2.$$

Since our estimator is unbiased (i.e., $\langle \hat{T}_o \rangle = T_o$), the variance of the estimator can be written

$$\langle (\hat{T}_o - T_o)^2 \rangle = \langle (\sum_{i=1}^n a_i \hat{T}_i - T_o)^2 \rangle = \langle [\sum_{i=1}^n a_i (T_i + e_i) - T_o]^2 \rangle, \quad (3.3)$$

where we recall that T_o is the true value.

The unknown weighting coefficients, a_i , are found by first minimizing the variance in (3.3). Derivatives with respect to each coefficient a_j yield a set of linear equations for $j = 1, 2, \dots, n$,

$$\left\langle 2 \left[\sum_{i=1}^n a_i (T_i + e_i) - T_o \right] [T_j + e_j] \right\rangle = 0. \quad (3.4.1)$$

(3.4.1) can be written as either

$$\left\langle \sum_{i=1}^n a_i (T_i + e_i) (T_j + e_j) \right\rangle = \langle T_o (T_j + e_j) \rangle, \quad (3.4.2)$$

or

$$\begin{aligned} \sum_{i=1}^n a_i [\langle T_i T_j \rangle + \langle e_i T_j \rangle + \langle T_i e_j \rangle + \langle e_i e_j \rangle] &= \langle T_o (T_j + e_j) \rangle, \\ &= \langle T_o \hat{T}_j \rangle. \end{aligned} \quad (3.4.3)$$

It is usually assumed that there is no relationship between the error field and temperature field on the average, such that

$$\langle T_o e_j \rangle = \langle e_i T_j \rangle = \langle T_i e_j \rangle = 0. \quad (3.4.4)$$

Consequently, for $j = 1, 2, \dots, n$, (3.4) becomes

$$\sum_{i=1}^n a_i [\langle T_i T_j \rangle + \langle e_i e_j \rangle] = \langle T_o \hat{T}_j \rangle. \quad (3.5)$$

Since T_i , T_j , and T_o are deviations from their average values, $\langle T_i T_j \rangle$ is the covariance function of the true temperature field and $\langle T_o \hat{T}_j \rangle$ is the covariance between the temperature of interest and the measurement $\hat{T}_j$. If the errors are unbiased (*i.e.*, $\langle e_i \rangle = \langle e_j \rangle = 0$), then $\langle e_i e_j \rangle$ is the error covariance. Also, since $\langle T_i e_j \rangle = 0$, it is easily shown that the bracketed ([]) terms in (3.5) can also be written

$$\langle T_i T_j \rangle + \langle e_i e_j \rangle = \langle (T_i + e_i) (T_j + e_j) \rangle = \langle \hat{T}_i \hat{T}_j \rangle, \quad (3.6)$$

which, for all i and j , is the covariance function of the measured temperature field. Both a dynamical component, because the structure of the covariance function is usually determined by mesoscale and/or wave dynamics $\langle T_i T_j \rangle$, and a measurement component $\langle e_i e_j \rangle$, contribute to the covariance function of the observations. These components are uncoupled by (3.4.4).

How accurate is this estimator? In our example, the covariance function of the true temperature field is also needed to determine the error variance associated with our estimator. To see this, substitute (3.6) into (3.5) and expand to give

$$\begin{aligned}
a_1 \langle \hat{T}_1 \hat{T}_1 \rangle + a_2 \langle \hat{T}_1 \hat{T}_2 \rangle + \dots + a_n \langle \hat{T}_1 \hat{T}_n \rangle &= \langle T_1 \hat{T}_o \rangle \\
\vdots & \\
a_1 \langle \hat{T}_n \hat{T}_1 \rangle + a_2 \langle \hat{T}_n \hat{T}_2 \rangle + \dots + a_n \langle \hat{T}_n \hat{T}_n \rangle &= \langle T_n \hat{T}_o \rangle
\end{aligned}$$

which, upon substitution of the following matrix definitions (t is transpose),

$$\begin{aligned}
\hat{T}^t &= (\hat{T}_1, \hat{T}_2, \dots, \hat{T}_n), \\
C_{\hat{T}\hat{T}} &= \begin{pmatrix} \langle \hat{T}_1 \hat{T}_1 \rangle & \dots & \langle \hat{T}_1 \hat{T}_n \rangle \\ \vdots & \ddots & \vdots \\ \langle \hat{T}_n \hat{T}_1 \rangle & \dots & \langle \hat{T}_n \hat{T}_n \rangle \end{pmatrix}, \\
\mathbf{a}^t &= (a_1, a_2, \dots, a_n),
\end{aligned}$$

$$C_{T_o \hat{T}} = (\langle T_o \hat{T}_1 \rangle, \langle T_o \hat{T}_2 \rangle, \dots, \langle T_o \hat{T}_n \rangle),$$

and rearrangement, becomes

$$\mathbf{a} = C_{T_o \hat{T}} C_{\hat{T}\hat{T}}^{-1}. \quad (3.7)$$

Note that $C_{\hat{T}\hat{T}}$ is the covariance between the observations. It is sometimes more rigorously denoted as $C_{\hat{T}\hat{T}^t}$, since the matrix is formed by the expected value of the outer product of $\hat{T}$ and $\hat{T}^t$. The transpose operator, t , is dropped for notational convenience. All the second arguments in our notation for covariance functions are the transpose of that argument. Moreover, in many of the references, the second argument is dropped when it is the same as the first argument.

The variance of our estimator, $C_{\hat{T}_o \hat{T}_o}$, is defined to be

$$C_{\hat{T}_o \hat{T}_o} = \langle (\hat{T}_o - T_o)^2 \rangle.$$

The following formula for estimation variance can easily be derived,

$$C_{\hat{T}_o \hat{T}_o} = C_{T_o T_o} - C_{T_o \hat{T}} C_{\hat{T}\hat{T}}^{-1} C_{\hat{T} T_o}^t. \quad (3.8.1)$$

The square root of (3.8.1), or the standard deviation, is usually referred to alternatively as the estimation error, mapping error, or analysis error. (3.8.1) may also be written as

$$\sigma_{\hat{T}_o}^2 = \sigma_{T_o}^2 - C_{T_o \hat{T}} C_{\hat{T}\hat{T}}^{-1} C_{\hat{T} T_o}^t. \quad (3.8.2)$$

The first term on the right hand side of (3.8.1) is the natural variation in the true temperature at the point of interest, T_o , and the second term on the right hand side of (3.8.1) measures the information content of the observations. If the covariance between T_o and $\hat{T}$ is zero, as it would be if the measurements were very distant from the location of interest, then

- (a) the second term on the right hand side of (3.8.1) and a in (3.7) are zero,
- (b) the best estimate of T_o is the mean value, and
- (c) the error variance associated with our estimate, $\hat{T}_o$, is the field variance at location O , $\sigma_{T_o}^2$.

The measurements added no new information and the estimation error is the square root of the variance of the temperature field at the location of interest. Of course in practice, the true field variance is an estimate based on all *a priori* information; historical data, concurrent data, and theoretical expectations.

A more general estimation problem is the estimation of the temperature field at m different locations given n different measurements. Instead of a vector of coefficients, as we had in (3.7), we now need to find a $m \times n$ matrix of coefficients, A , to estimate the field at the m locations of interest. Define a vector $T_o^t = (T_{o1}, \dots, T_{on})$, which is the true temperature field at the locations of interest, and its associated field of temperature estimates $\hat{T}_o$. Eqn. (3.7) generalizes in a straightforward fashion to

$$A = C_{T_o \hat{T}} C_{\hat{T} \hat{T}}^{-1} \quad (3.9)$$

where $C_{T_o \hat{T}}$ and A are now $m \times n$ matrices. The variance of our estimator $\hat{T}_o$ is

$$C_{\hat{T}_o \hat{T}_o} = C_{T_o T_o} - C_{T_o \hat{T}} C_{\hat{T} \hat{T}}^{-1} C_{\hat{T} T_o}^t. \quad (3.10)$$

Eqns. (3.9) and (3.10) are a version of the Gauss-Markov theorem for a linear minimum mean-square estimate of a random variable. It should be noted here that the Gauss-Markov theorem is not affected by replacing the covariance functions with covariance functions normalized by their value at zero lag or correlation functions. Correlation functions are generally used in practice because of numerical convenience. Thus C in all of the previous formulas can either denote covariance or correlation.

In fact, the matrices associated with the covariance functions can undergo an affine transformation (magnitude change, rotation, translation). Structure functions (e.g., Gandin, 1963)

$$G_{TT} = 2 * \sigma_{T_o}^2 (I_n - C_{TT}) = \begin{pmatrix} 0 & \langle (T_1 - T_2)^2 \rangle & \dots & \langle (T_1 - T_n)^2 \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle (T_n - T_1)^2 \rangle & \langle (T_n - T_2)^2 \rangle & \dots & 0 \end{pmatrix}, \quad (3.11)$$

can be used. C_{TT} is the correlation function. The advantage of structure functions is that a mean is not removed from the observation. Interpolation using structure functions is known as Kriging in geostatistics (e.g., Journel and Huijbregts, 1978), and its use is gaining popularity in oceanography (Denman and Freeland, 1985; Hanson and Herman, 1989). The biggest disadvantages of Kriging are that (1) the present algorithms do not allow for spatial and temporal correlations and (2) the structure function has considerable curvature near zero lag. This extreme curvature near zero lag hampers estimation of parametric forms of the structure

function. Kriging was formulated for mapping ore deposits that do not move over the observing time. When mapping ocean features, especially in strong frontal regimes, phase speeds are needed to move non-synoptic observations to the location where they would be at the estimation time. Also, a temporal correlation factor is needed to account for the loss of correlation in time due to turbulence and/or stochastic forcing.

The primary subjective components of an OA scheme are the selection of the assumed observational noise level, the number of *influential data points*, the mean field, and the correlation function(s). The four subjective components are discussed next.

The numerical value of the error is given as a fraction of total field variance, typically in the range of 0.05-0.50. Fortunately, the accuracy of the objective analysis is only weakly dependent upon the actual value of this difficult-to-estimate parameter. On the average, an OA estimate at an observing location will be within the noise level of the observed value. The larger the assumed noise level, the smoother the OA field. For the case of assumed uncorrelated noise, the noise level parameter appears in the main diagonal of the observation's covariance matrix as an additive term. The assumption of uncorrelated noise is poor when the spacing between observing stations is much less than the distance between the estimation locations. For instance, if you are making large-scale maps using stations that resolve the energetic mesoscale only in data-rich patches, uncorrelated noise is not a good error model (see Clancy (1983) for an excellent discussion of this point). The assumption of uncorrelated noise is also poor for systems that measure with a bias. Bias reduces the amount of independent information that the observations provide, but it enhances gradient calculations (Seaman, 1977). Biased noise also puts a lower limit on the mapping error that does not exist for random noise (Gandin, 1963). See Bergman (1978) and Chelton (1983) for further discussion on correlated errors.

Since the noise level increases the magnitude of the diagonal elements of $C_{\hat{T}\hat{T}}$, a large enough noise level will guarantee a positive semi-definite matrix, and the numerical inversion of $C_{\hat{T}\hat{T}}$ will be stable. Hence, it is always advisable to have some nonzero value for the assumed noise level.

An OA estimate is essentially a weighted average of observations with the weights chosen by eqn. (3.7). The number of observations ("limit") that influence the OA estimate must also be chosen. The maximum number of points to use in this weighted average is the "limit," which typically ranges from 5-20 points. On the one hand, large values of "limit" improve the statistical confidence in the estimate because of a large number of degrees of freedom. On the other hand, larger than necessary values of "limit" do not improve the accuracy of the OA estimate and may smear features in heterogeneous regions. (Note, as indicated by eqn. (3.9), that a $n \times n$ matrix must be inverted for each estimation point; n is an upper bound on "limit.") While smaller values of "limit" lead to faster

computation, the OA estimates under these circumstances are more sensitive to noise in the observations. If the correlation representation is nearly singular, a relatively small value of "limit" may be advantageous since the inversion of a small matrix is less likely to be numerically singular than a large one. Some trial and error experimentation with the value of "limit" is a recommended part of the early implementation of an OA. For our analysis, the number of influential data points equals 8.

Ideally, the best influential data points are those data points most correlated with the estimation location and least correlated amongst themselves. The latter requirement can be relaxed by averaging close and highly correlated points before the selection of influential data points. Then only the former requirement of the most correlated data points with the estimation location are used to select influential data points. The most correlated data points are found by sorting the absolute value of the correlations. The absolute value is sorted since, for example, a correlation of -0.9 is more desirable than a correlation of 0.1 .

Our approach for decomposing data containing both large scale and mesoscale components is to remove the mean of the data before the OA, and to objectively interpolate just the deviations from the mean. The final field estimate is the sum of the mean and the OA of the deviation field. In practice, the mean field is a function of longitude, latitude, depth, and time, and is either a trend, climatology, a forecast from a dynamical model, or a previous field estimate. Common trends are either a constant average, linear trends, polynomial surfaces, splines, or different sets of basis functions. A frequently used name, first-guess field, sums up the situation nicely.

OA works best for interpolation. Extrapolation by OA or any other technique can be risky. For instance, an OA might yield unrealistic streamlines running in and out of a coastline or temperature values that are far from the historical range. These problems are usually caused by sparse boundary data, which can affect the analysis at two different stages. First, a spatial trend calculated from sparse and/or patchy data may not represent the true trend over the entire analysis domain. For example, a plane fit to the data in a region that contains a front can give unrealistic estimates near the edge of the analysis domain away from the front. The second problem with sparse data is that only a simple few-parameter correlation function can be estimated from the data. Based on these considerations, a tense bicubic smoothing spline was chosen to represent the mean field in our analysis.

Inoue's least-square finite-element splines were chosen because they have adjustable smoothness and tension parameters, allow variable data errors, and, for computational efficiency, they use a B-spline basis and iterate with binary subdivisions (Inoue, 1986). Because of the rich large-scale structure evident in sea surface velocity fields, a trend consisting of a simple mean or two-dimensional polynomial fit produces unrealistic maps for large analysis domains. MB92 used two-dimensional bicubic splines for the trend surface in the interpolation of sea surface temperature in the tropical Pacific from $30^{\circ}S$ to $30^{\circ}N$ with good results. The use of tense

splines allows reasonable extrapolation from strong frontal regions with large spatial gradients to more homogeneous regions. The spline parameters chosen were the (inverse) smoothing parameter $\rho = 10^{-4}$, the tension parameter $\tau=0.99$, and the initial number of knots for the large basins, $m_{x0} = m_{y0} = 60$.

Typical correlation functions are in the form of an oscillatory term multiplied by a decay term. The oscillatory term models wave behavior and is usually a trigonometric function or a Bessel function. The decay term models turbulent behavior and is usually an exponential (i.e., Gaussian) function or inverse polynomial. Other representative functions are described in Bretherton *et al.* (1976), Carter and Robinson (1987), Julian and Thiébaux (1975), Freeland and Gould (1976), McWilliams *et al.* (1986), Thiébaux (1975, 1976), and Thiébaux and Pedder (1987). A functional form for the correlation function is assumed in the parameter matrix algorithm with the novelty of having nine correlation parameters that can vary in space and time. The parameter matrix algorithm uses an anisotropic, time-dependent correlation function with correlation parameters that vary in space/time, and a time-dependent trend surface for efficient OA of dynamically heterogeneous and non-stationary fields. An observation (subscript o) is decomposed into three components,

$$T_o(x, y, t) = T_m(x, y, t) + T_e(x, y, t) + e_s(x, y, t), \quad (3.12)$$

where the first term is the contribution of the large scale or trend field (subscripted m for mean), the second term is the natural field variability important on the mesoscale or synoptic time scale (subscripted e for eddy), and the third term is the combined effects of unresolved scales, i.e., subgrid-scale noise, and error from the particular sensor (subscript s for subgrid-scale and sensor error).

T_m was estimated by least-square fitting of a bicubic spline surface to the data. The anisotropic and time-dependent correlation model for estimating T_e from the detrended data ($T_o - T_m$) is

$$C(dx, dy, dt) = C(1)[1 - (DX/C(4))^2 - (DY/C(5))^2] * \exp - [(DX/C(6))^2 + (DY/C(7))^2 + (dt/C(8))^2], \quad (3.13)$$

$$DX = dx - C(2) * dt \text{ and } DY = dy - C(3) * dt,$$

where,

- dx is the east-west lag; dy is the north-south lag; and dt is the time lag.
- $C(1)$ is the correlation at zero lag and equals one minus the normalized (by the field variance) measurement variance (which includes a subgrid scale component).
- $C(2)$ and $C(3)$ are the mean phase speeds in the east-west and north-south direction, respectively.

- $C(4)$ and $C(5)$ are the zero-crossing scales in the east-west and north-south direction, respectively.
- $C(6)$ and $C(7)$ are the spatial decay scales (or the e-folding scales) in the east-west and north-south direction, respectively.
- $C(8)$ is the temporal decay scale.

This correlation function can be rotated in space by $C(9)$, an arbitrary angle. This correlation function and its isotropic form, $C(4) = C(5)$ and $C(6) = C(7)$, have been used in a variety of dynamical regimes (*e.g.*, Carter and Robinson, 1987; Robinson *et al.*, 1987; and MB92).

The parameter matrix algorithm starts by dividing both the interpolation grid and data domain into the same rectangular space/time bins. The dimensions of the bins are based on the correlation scales and how the correlation parameters change in space and time: (1) each bin must contain one set of correlation parameters and (2) all the data in a bin should be significantly correlated with the midpoint of the bin to minimize computations. The parameter matrix is formed by assigning the values of the nine correlation parameters to each numbered bin.

The present parameter matrix algorithm consists of the following steps:

- 0) Highly correlated data are averaged.
- 1) Each data point is assigned a bin number.
- 2) The data points are sorted by bin number so that all the data for each bin are stored together with their corresponding position, time, detrended value and error level.
- 3) A marker array is used to record the array position where each bin starts (or ends).
- 4) The OA is performed one interpolation point at a time.
- 5) The bin number of the interpolation point is found and denoted by N .
- 6) Data in bin N and all bins within an influential correlation window from the interpolation point are selected for determining the influential data points.
- 7) The correlations between the interpolation point and each data point from step 6 are tabulated and sorted.
- 8) The most influential data points, *i.e.*, the most correlated with the interpolation point, are used for the OA and usually number between 6 and 20 data points.
- 9) Steps 5-8 are repeated for each interpolation point.

If a data point is in one bin and the interpolation point is in another bin, all the correlation parameters from the two bins are averaged except the phase speeds. The phase speeds of the data bin are used since it is desirable to move the data points to where they would be at the estimation time. For calculating correlation between data in different bins, all the correlation parameters (including the phase speeds) are averaged.

The algorithm allows two types of random errors to contribute to the diagonal elements of $\mathbf{C}_{\hat{\mathbf{T}}\hat{\mathbf{T}}}$ and enter the calculation of $\hat{\mathbf{T}}$. An environmental error due to subgrid-scale processes is assigned to $C(1)$ in the parameter matrix and a sensor

error is assigned to each data point. For these calculations, $C(1) = 0.9$, and the sensor error is assumed to be 0.1 so that the total normalized error level is 0.2.

The exact error is hard to estimate because of the unknown and hard-to-model effects of winds and waves in conjunction with ship shape and size. Another unknown factor is the skill, or lack of it, of the navigators. Richardson and McKee (1984) estimated a random error for a single ship drift speed to be 20 *cm/s*. Systematic and bias errors in dead-reckoning will cancel out in the velocity calculations, but the systematic error due to windage will not. By comparing ship drift, drifting buoys, and current meters, McPhaden *et al.* (1991) estimated an upper bound of windage effects to be 3 percent of the surface wind speed. For SAR applications, the windage bias is probably not a problem and may actually help when looking for distressed ships. On the other hand, windage effects downgrade the ship drift velocity estimates for comparisons with ocean circulation models.

Wind speed information was not used in the calculation of the error level. Given a range of speeds from 0 to 200 *cm/s*, typical maximum speeds of order 100 *cm/s*, and that the random error component for single ship drift estimate is 20 *cm/s* (Richardson and McKee, 1984) leads to an average error on the order of the assumed 20 percent value ($\frac{20 \text{ cm/s}}{100 \text{ cm/s}} \times 100\%$). Future calculations may want to assume that this error is on the order of 10 percent in the western boundaries and on the order of 50 percent in mid-ocean regions where typical eddy speeds are 40-60 *cm/s*. One fact should not be overlooked. In all of the quoted studies, the investigators were surprised at how much useful information was obtained on basin-wide circulation patterns and their seasonal variability after averaging ship drift data. Though the data are very noisy, the law of large numbers and a data set consisting of over four million data points lead to stable estimates of seasonal ocean currents.

The following spatial scales, in units of degrees of longitude for the x direction and degrees of latitude for the y direction, were chosen based on data density considerations and because the vast majority of the data resolve scales on the order of 300-400 *km*. Dynamical considerations would lead to smaller correlation scales and more structure in the final field estimates. Since the bicubic splines capture almost all of the large-scale heterogeneity of the surface velocity fields, the spatial scales were chosen to be isotropic. The temporal scale, in units of days, was chosen to be large since a historical average of data from many different years is wanted (not the typical OA synoptic scale estimate using just one month of data).

$$C(1) = 0.9, C(2) = C(3) = C(9) = 0.0, C(4) = C(5) = 3.0,$$

$$C(6) = C(7) = 2.0, C(8) = 99999.$$

The parameter matrix algorithm results are presented in Section 4.

4. Mariano Global Surface Velocity Analysis 1.0

For the two largest basins, the Pacific Ocean and the Atlantic Ocean, the largest amount of monthly data was in the Atlantic for the month of May, a total of 182,938 good data points, and the smallest amount of monthly data was in the Pacific for the month of December, a total of 49,562 good data points (Table 1). The OA maps and normalized mapping error plots for these two data density extremes are presented in Figs. 2 and 3.

Seasonal and yearly averages of the data are shown in Figs. 4-8. These velocity maps contain many well-known ocean features. For instance, the major western boundary currents—the Gulf Stream, the Brazil Current, the Kuroshio, E. Australian, Somali, Agulhas, Labrador Current, Malvanus, and the Kamchatka current, as well as the equatorial current system—North Equatorial Current (NEC), North Equatorial Counter Current (NECC), and the South Equatorial Current (SEC) in all three major ocean basins, are clearly visible in the annual maps (Fig. 4). Major eastern boundary currents—Guinea, Angola, Benguela, California, Peru, and the Leeuwin currents—are also clearly visible.

The seasonal averages—DJF, MAM, JJA, and SON (Figs. 5-8)—show the well-known reversal of the NEC in the Indian Ocean during the onset of the Southwest Monsoon and the retroflexion of the N. Brazil current into the NECC in the latter half of the year (see Richardson and Walsh, 1986). This agreement with well-known ocean circulation regimes proves, at least qualitatively, that these maps would be of practical use to the USCG.

The reliability of these maps was further studied in a visual comparison with the present historical velocity file in operational use. Based on these comparisons, it was quite clear to R&D Ctr that this data product was far superior to the overly smooth data base currently in use. The monthly space/time interpolated velocity fields were termed the Mariano Global Surface Velocity Analysis, hereafter MGSVA. If anything, some smoothing of the MGSVA in some areas may be needed.

The statistics of the differences between the Atlantic and Indian basins' OA estimates at the boundary are shown in Fig. 9. These calculations were performed for the area 10-30° E, 30-70° S, south of South Africa, for the month of January. Figs. 9a and 9b are the histogram of the velocity differences between the Atlantic and Indian estimates for the u and v velocity component, respectively. These differences are not biased. Fig. 9c is a histogram of the differences in speed between the Atlantic and Indian estimates. Fig. 9d shows the vector differences between the two estimates. Over 90 percent of the speed estimates agree to within 30 *cm/s* (Fig. 9c) and there is no clear-cut directional bias (Fig. 9d). This is encouraging since most of the estimates compared here are from the data-sparse areas with the largest estimation error.

There are no obvious biases between the two different sets of estimates. Consequently, the two sets of estimates for each month and for each seam should be averaged together using weights that are inversely proportional to each error estimate and normalized such that their sum is one. For example, let u_1 be the velocity estimate associated with region 1 and let e_1 be the estimation variance associated

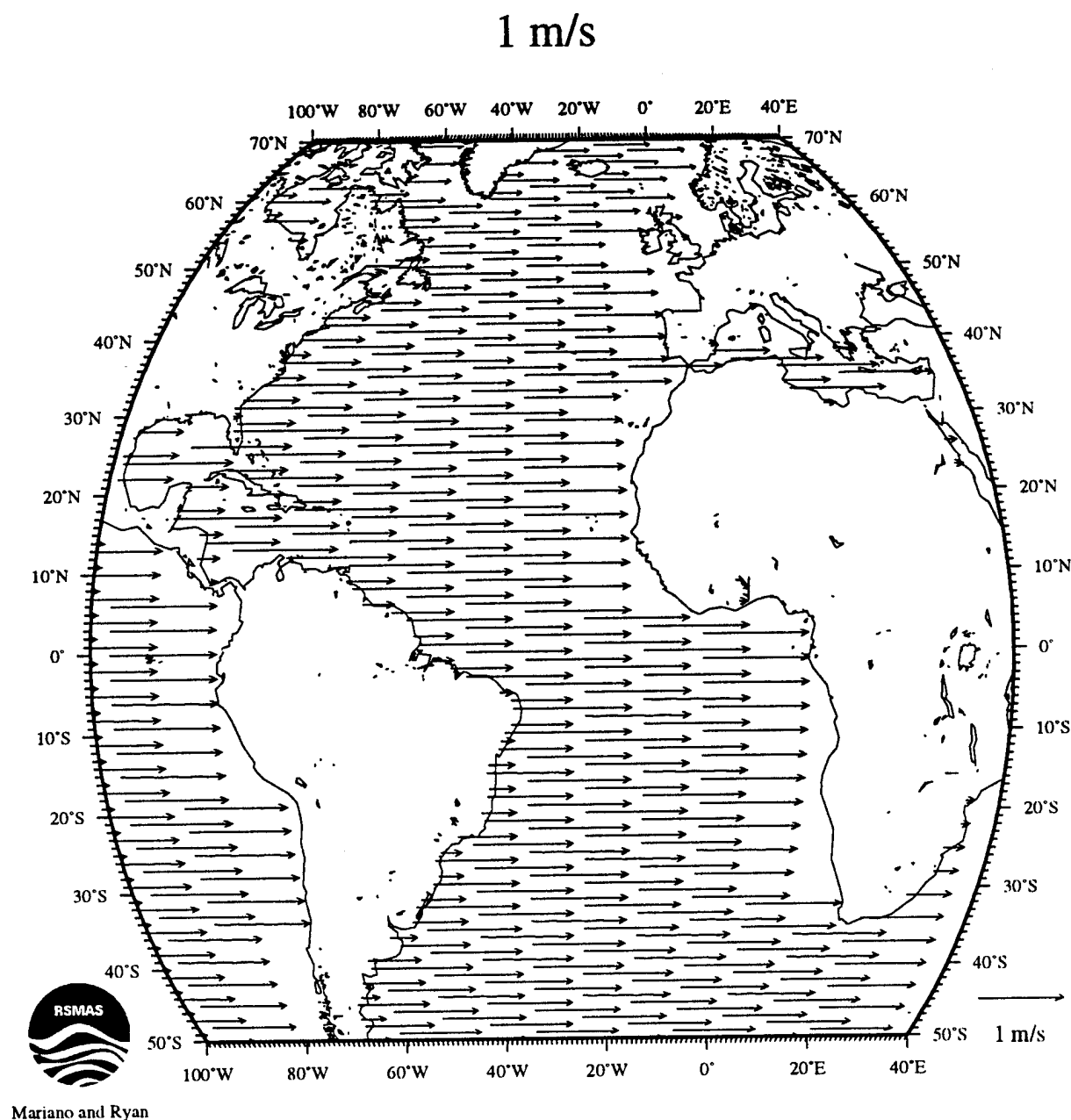
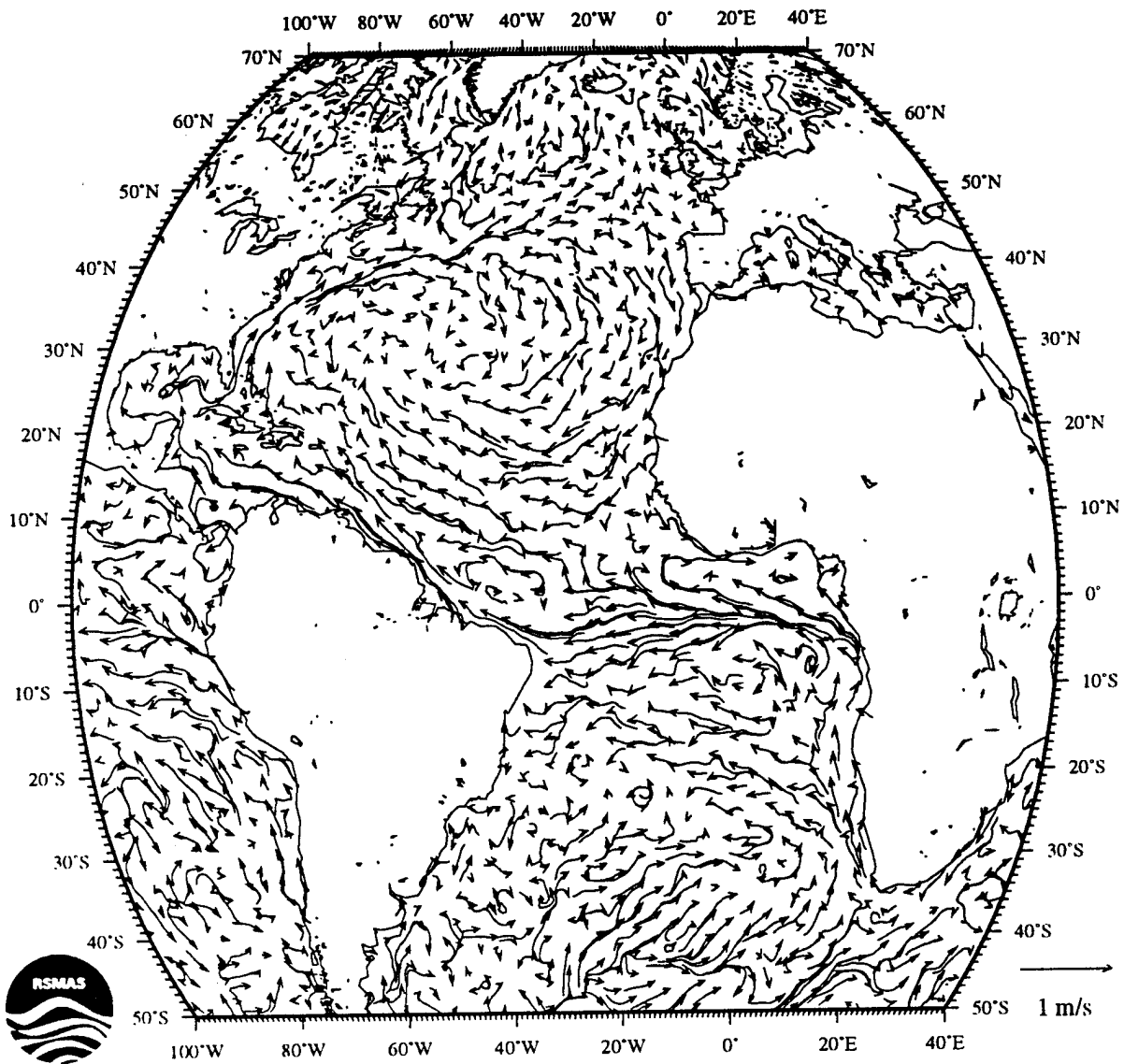


Fig. 2 (a) The velocity scale as a function of latitude. (The software for the curly vector plots was developed at the Navy Research Lab South.) (b) The OA map for the Atlantic Ocean for the month of May. (c) The corresponding error map. (Note that most of the basin, except for the sub-polar and polar regions, has low estimation error.)

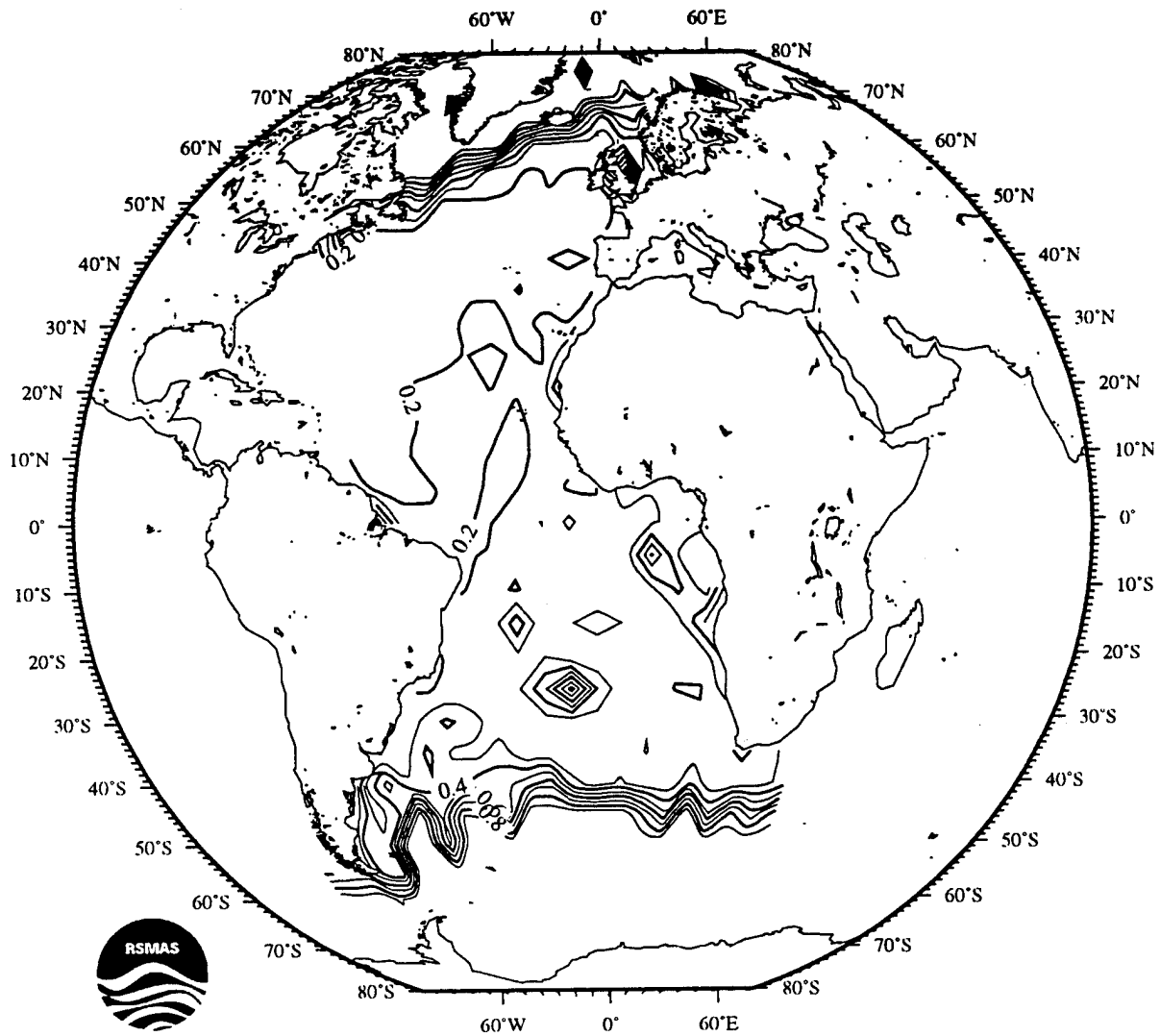
May



Mariano and Ryan

2(b)

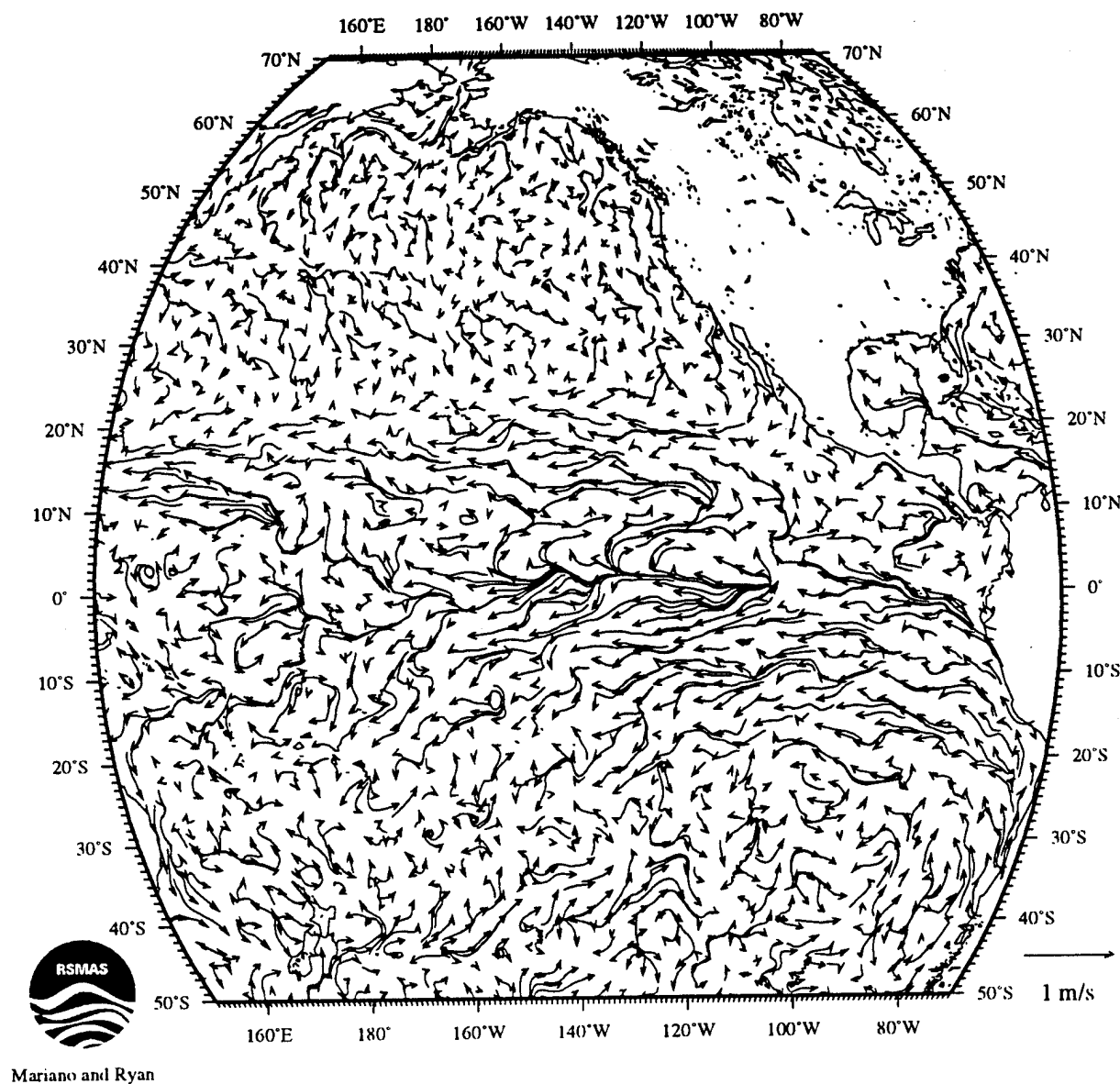
May error levels



Mariano and Ryan

2(c)

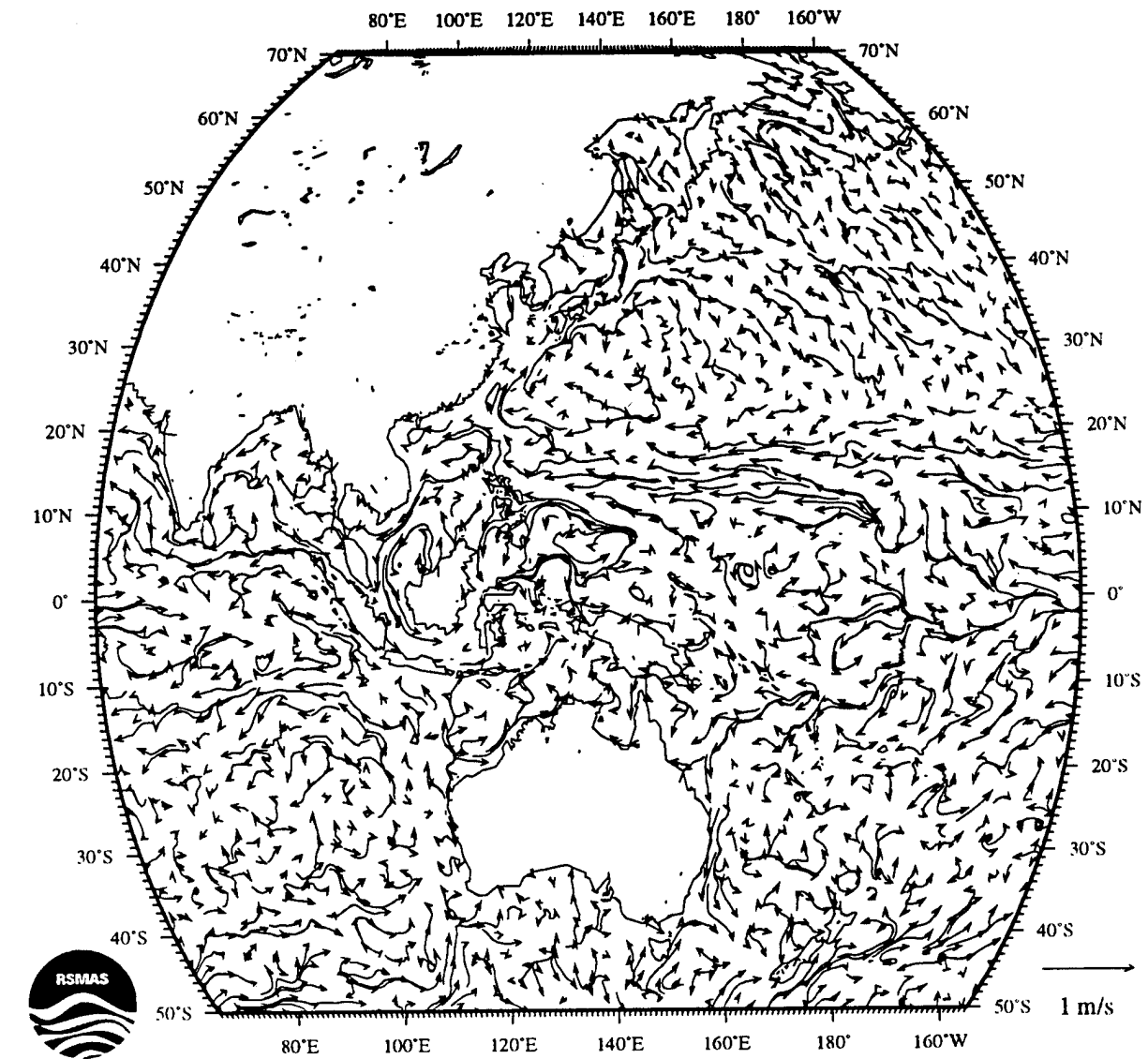
December



3(a)

Fig. 3 (a) The OA map for the Pacific Ocean for the month of December. (b) The corresponding error map. (Note that the North Pacific has much lower estimation error than the South Pacific.)

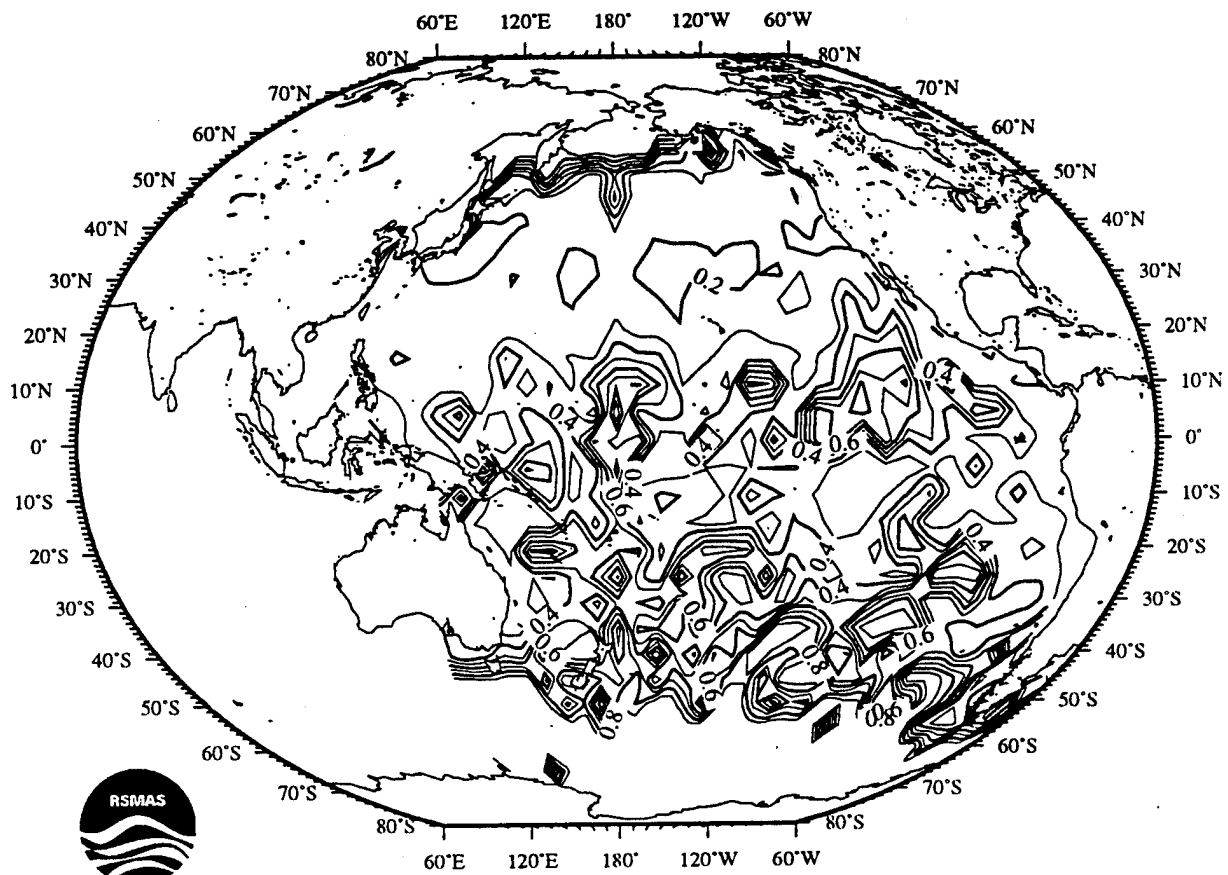
December



Mariano and Ryan

3(a)(continued)

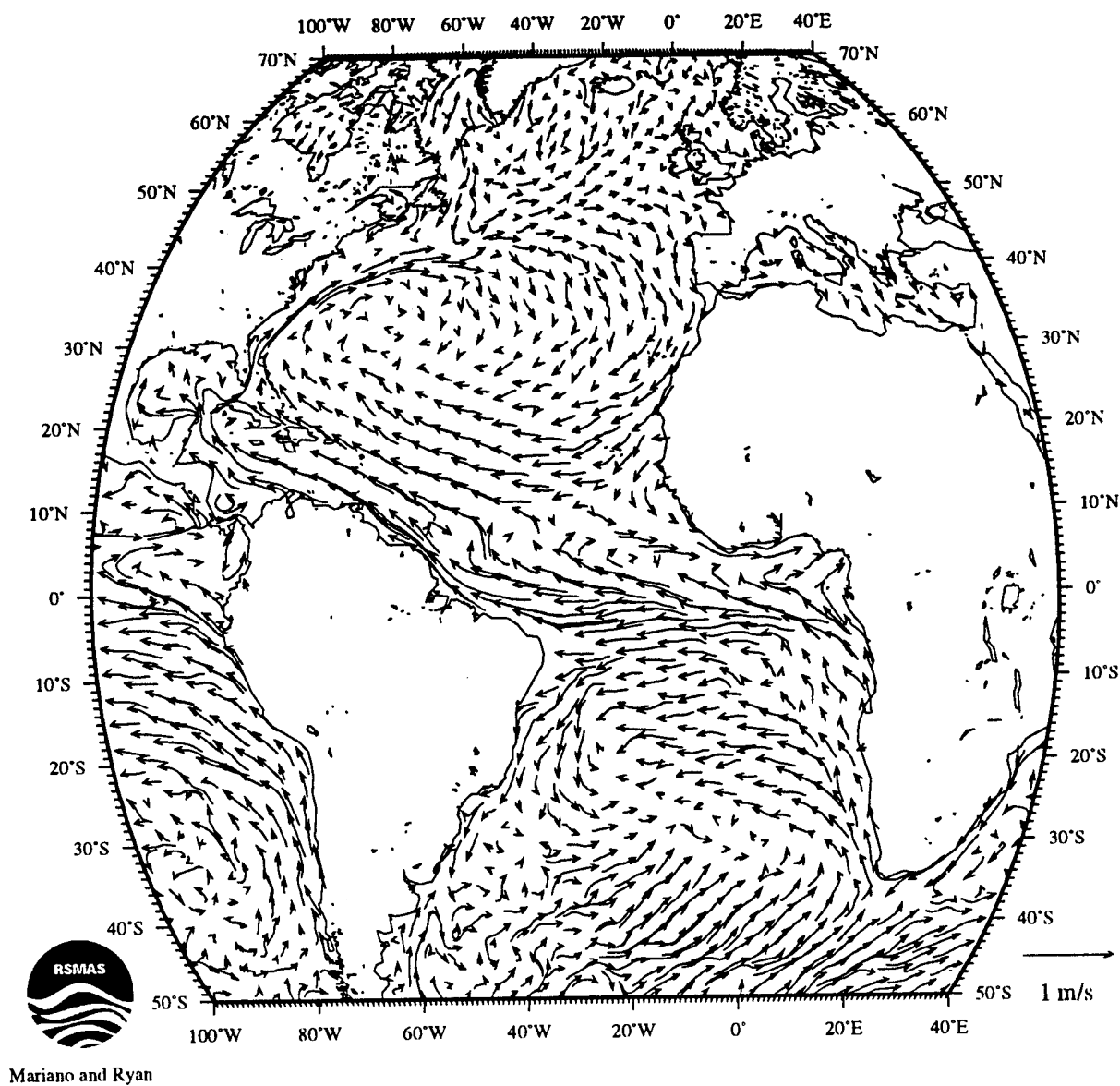
December error levels



Mariano and Ryan

3(b)

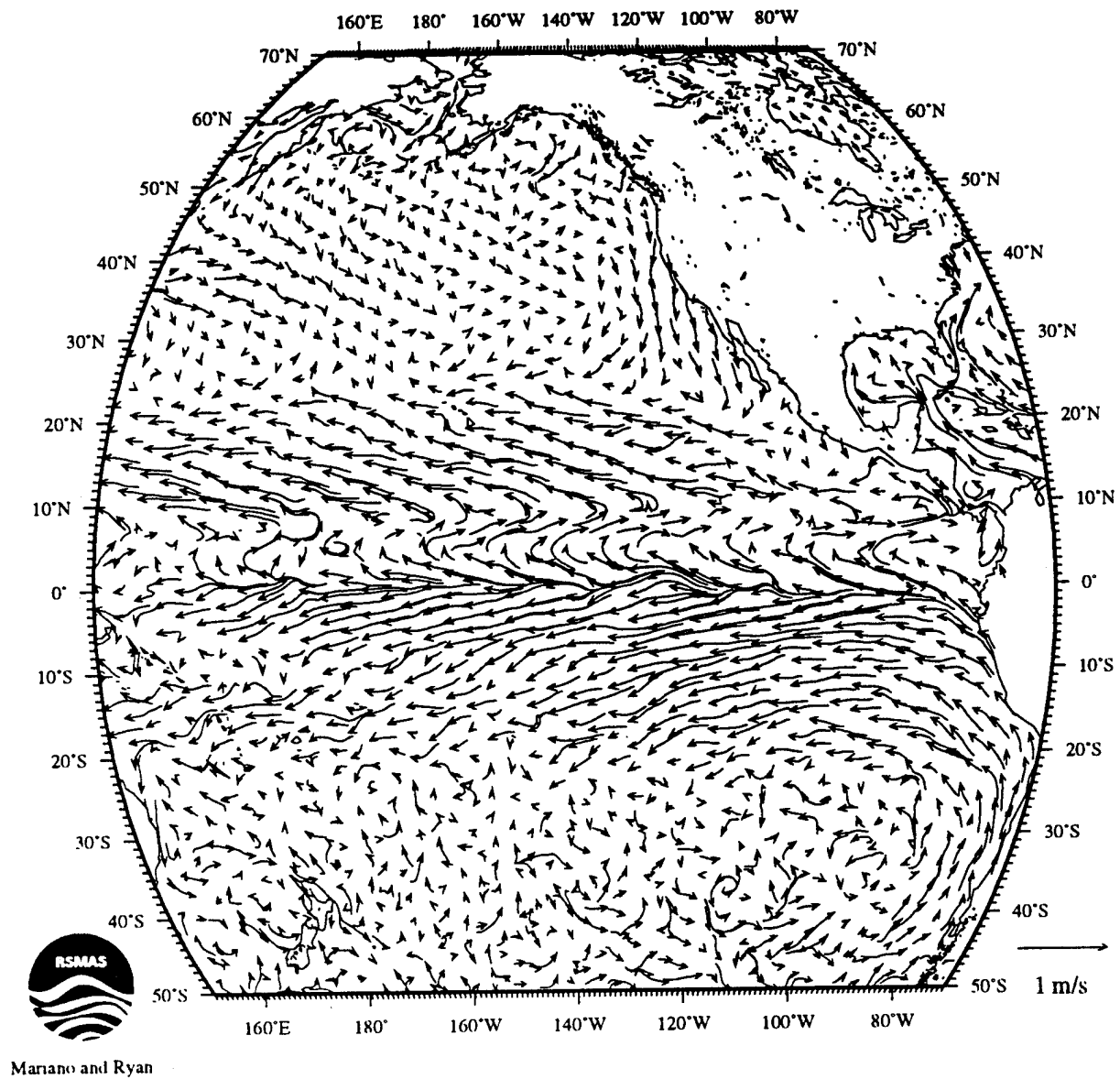
Yearly Average



4(a)

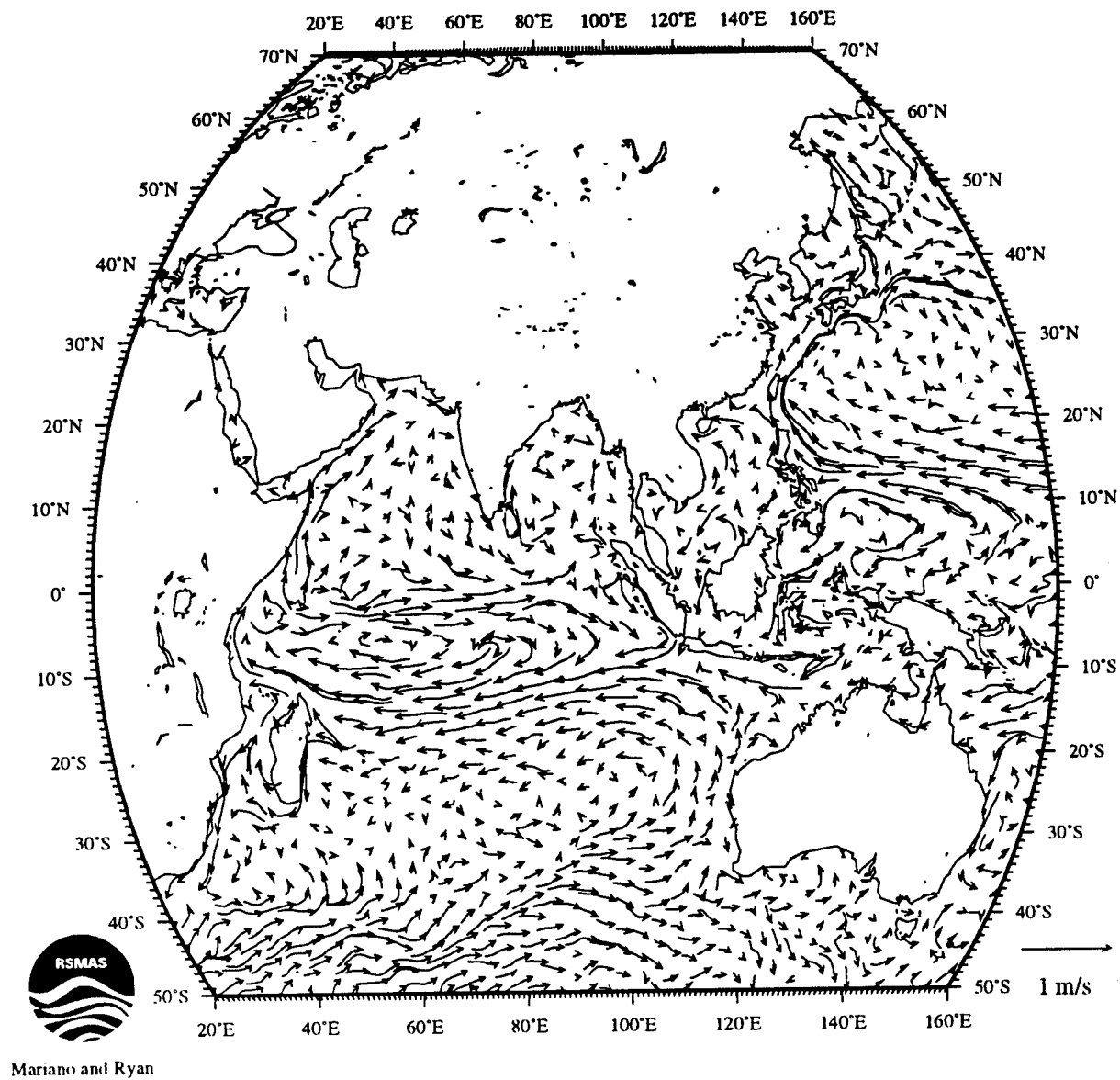
Fig. 4 (a-d) The weighted annual average of the monthly surface velocity estimates. The weights were inversely proportional to the error estimates associated with each monthly estimate.

Yearly Average



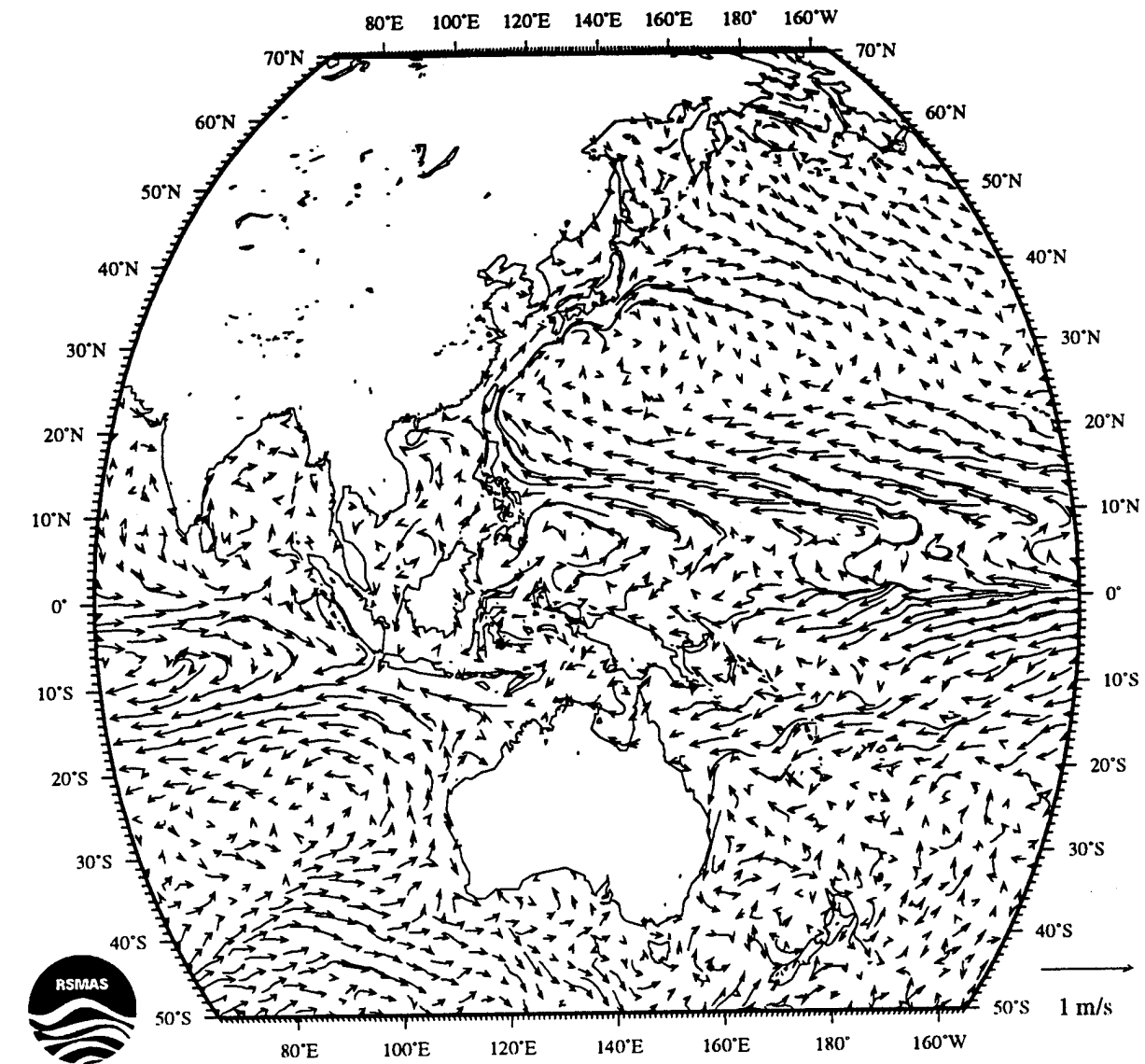
4(b)

Yearly Average



4(c)

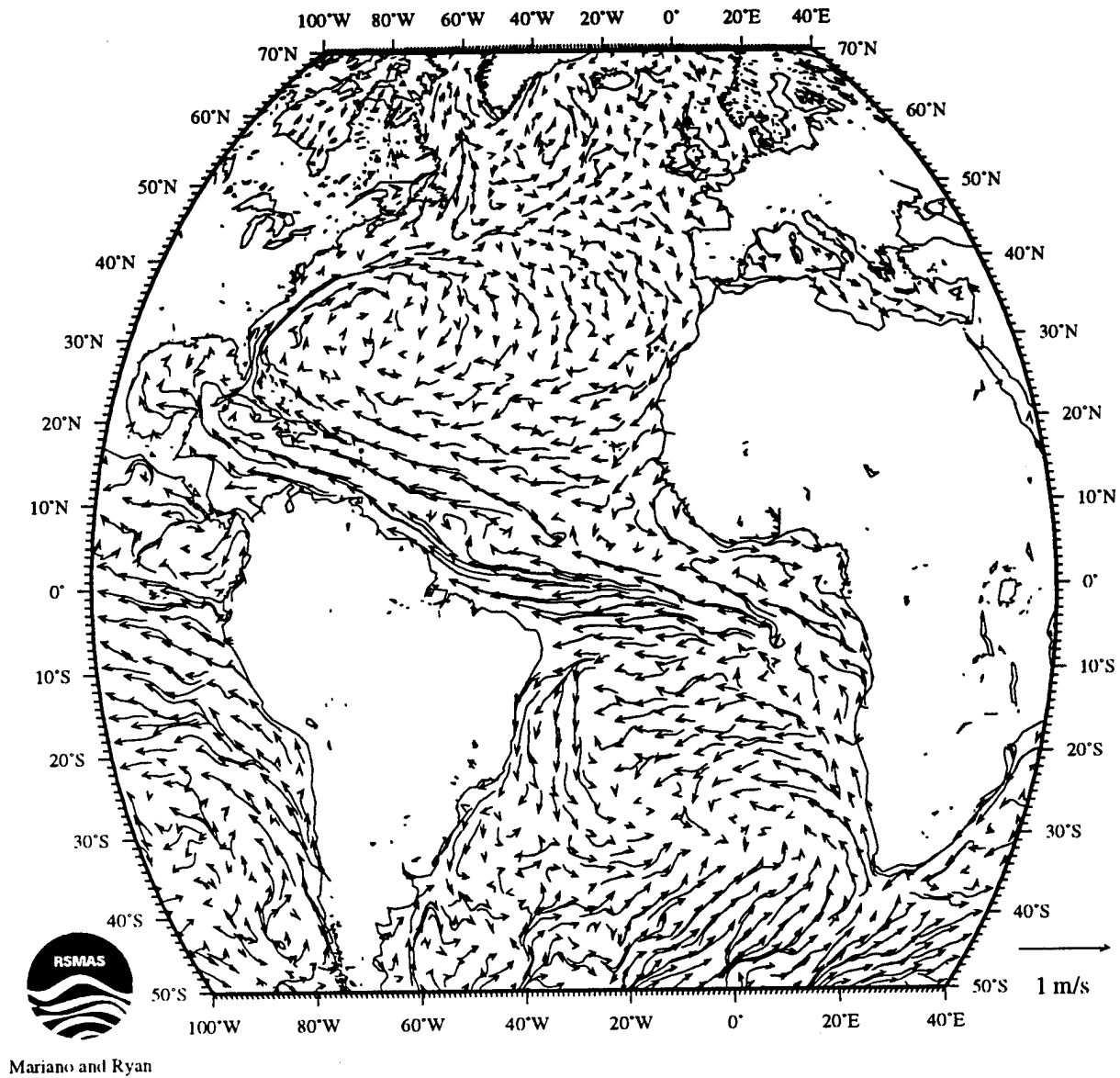
Yearly Average



Mariano and Ryan

4(d)

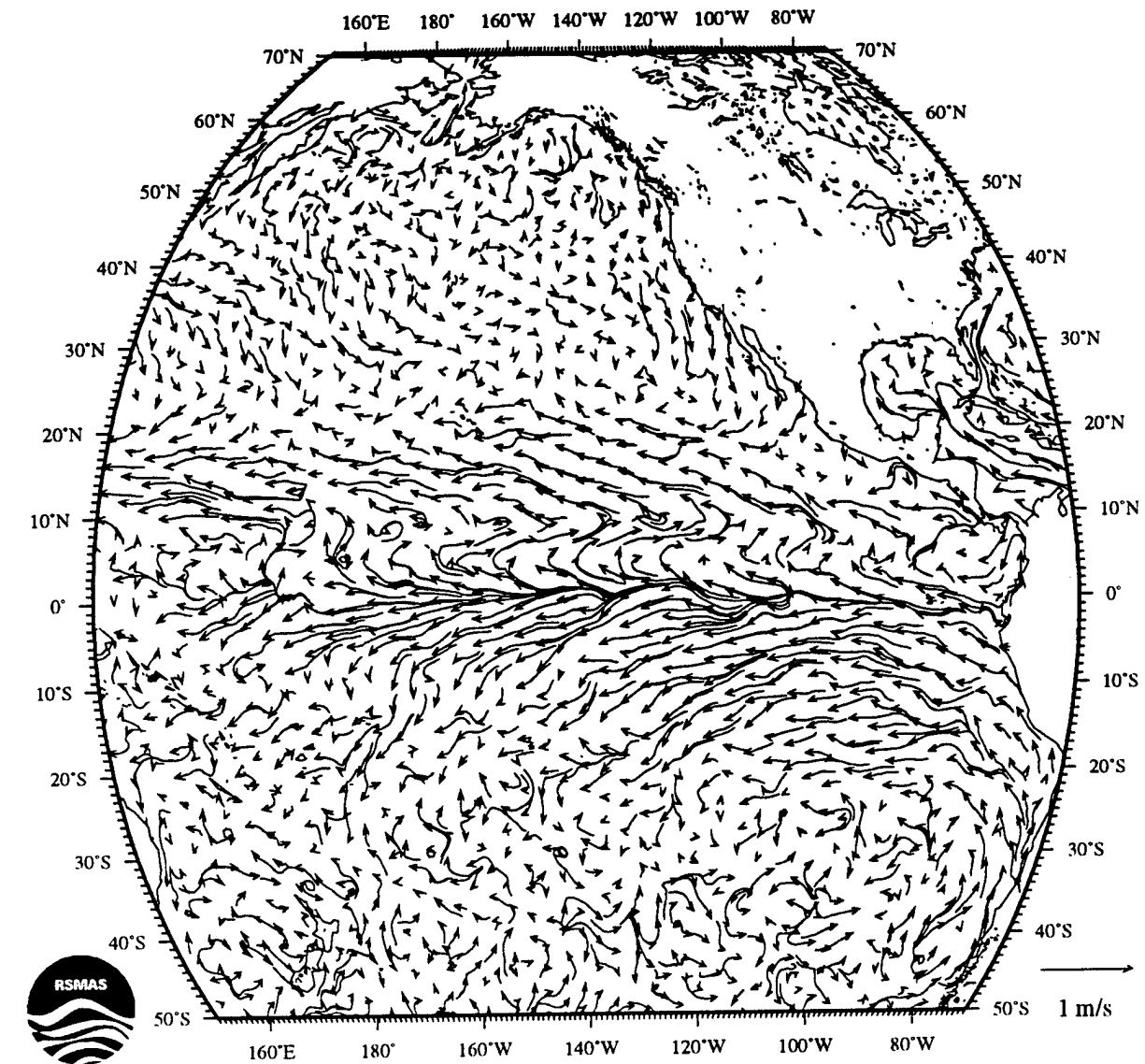
Winter Average



5(a)

Fig. 5 (a-d) The weighted seasonal average of the monthly surface velocity estimates for the months of December, January, and February.

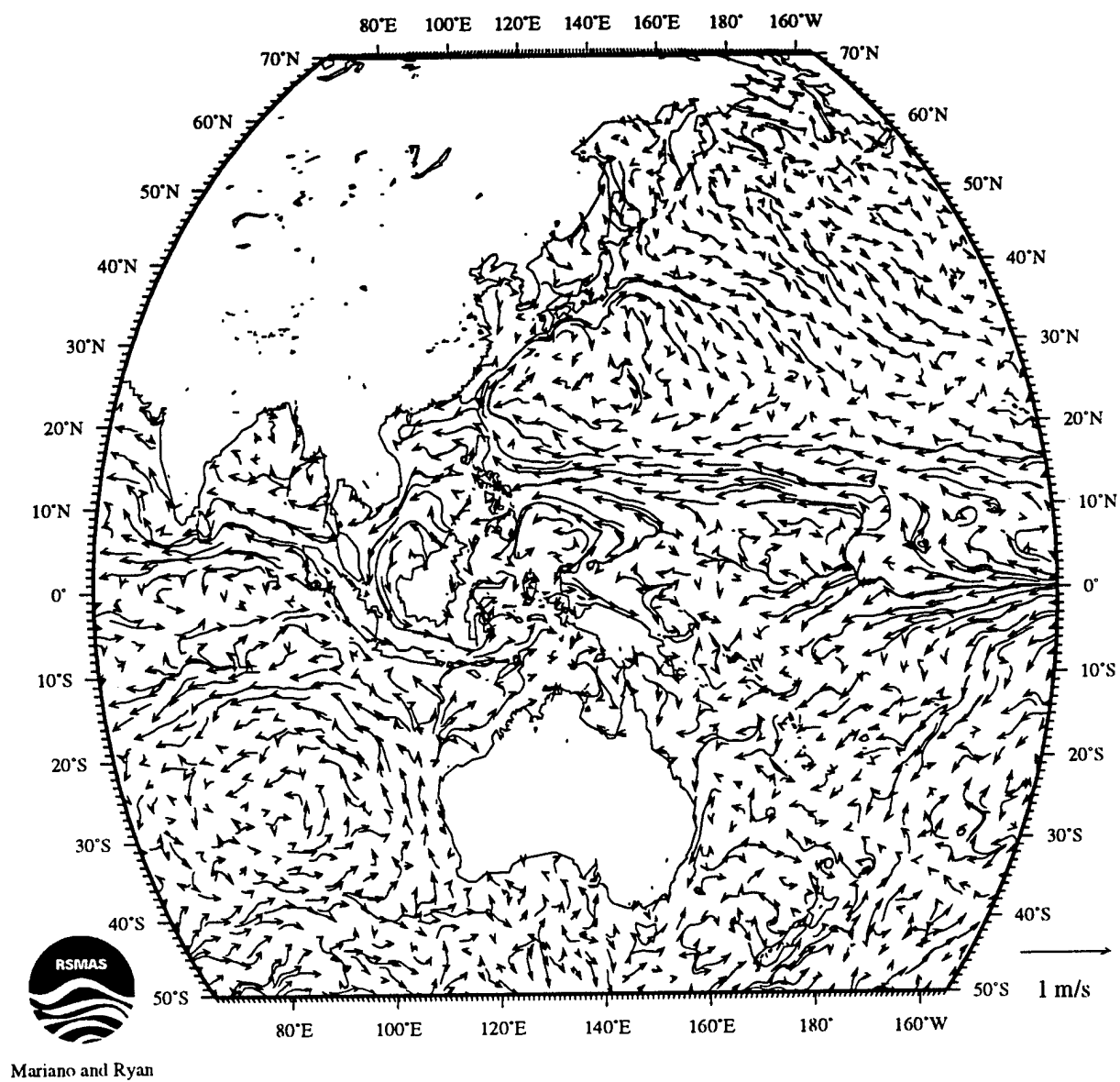
Winter Average



Mariano and Ryan

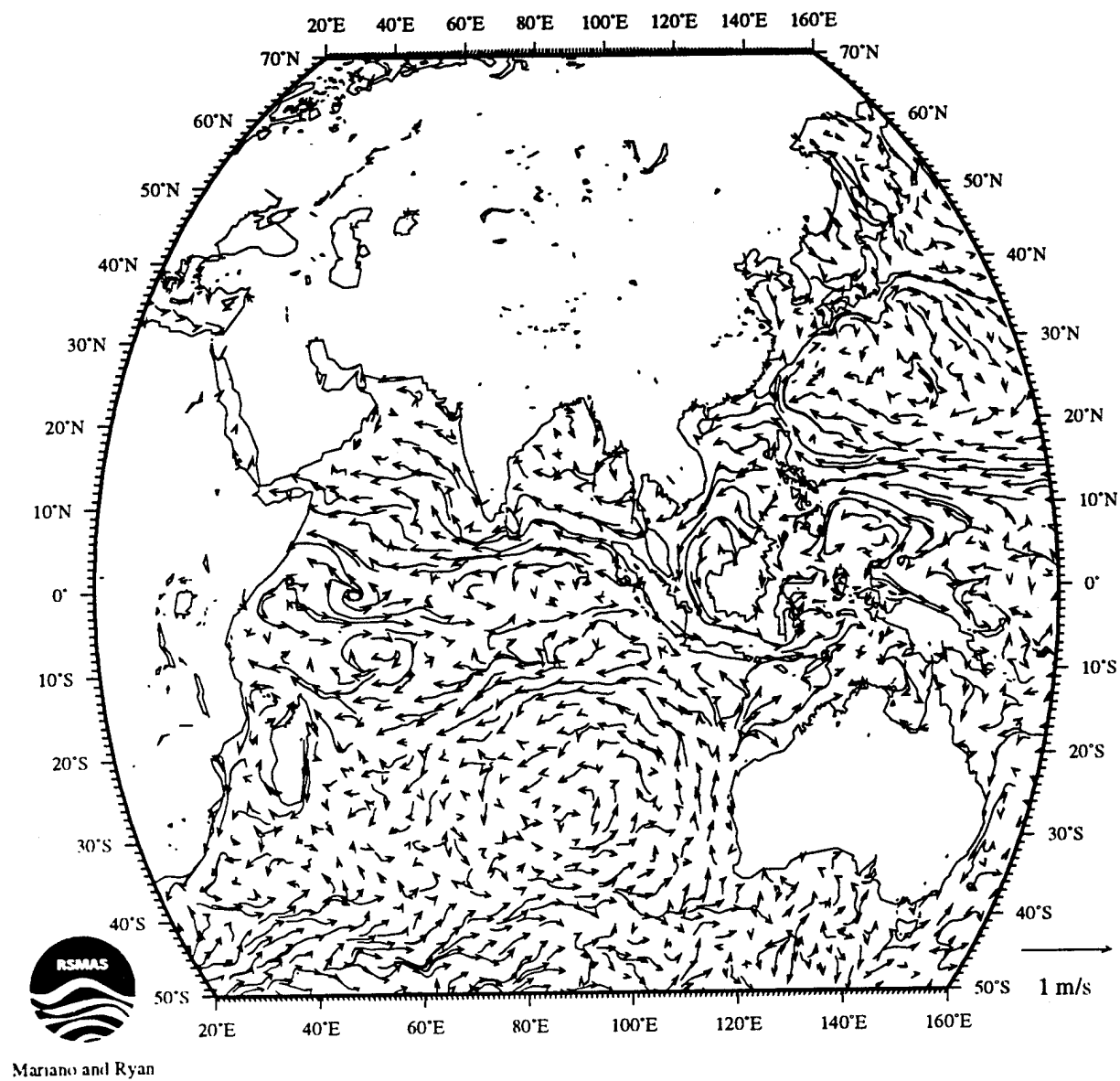
5(b)

Winter Average



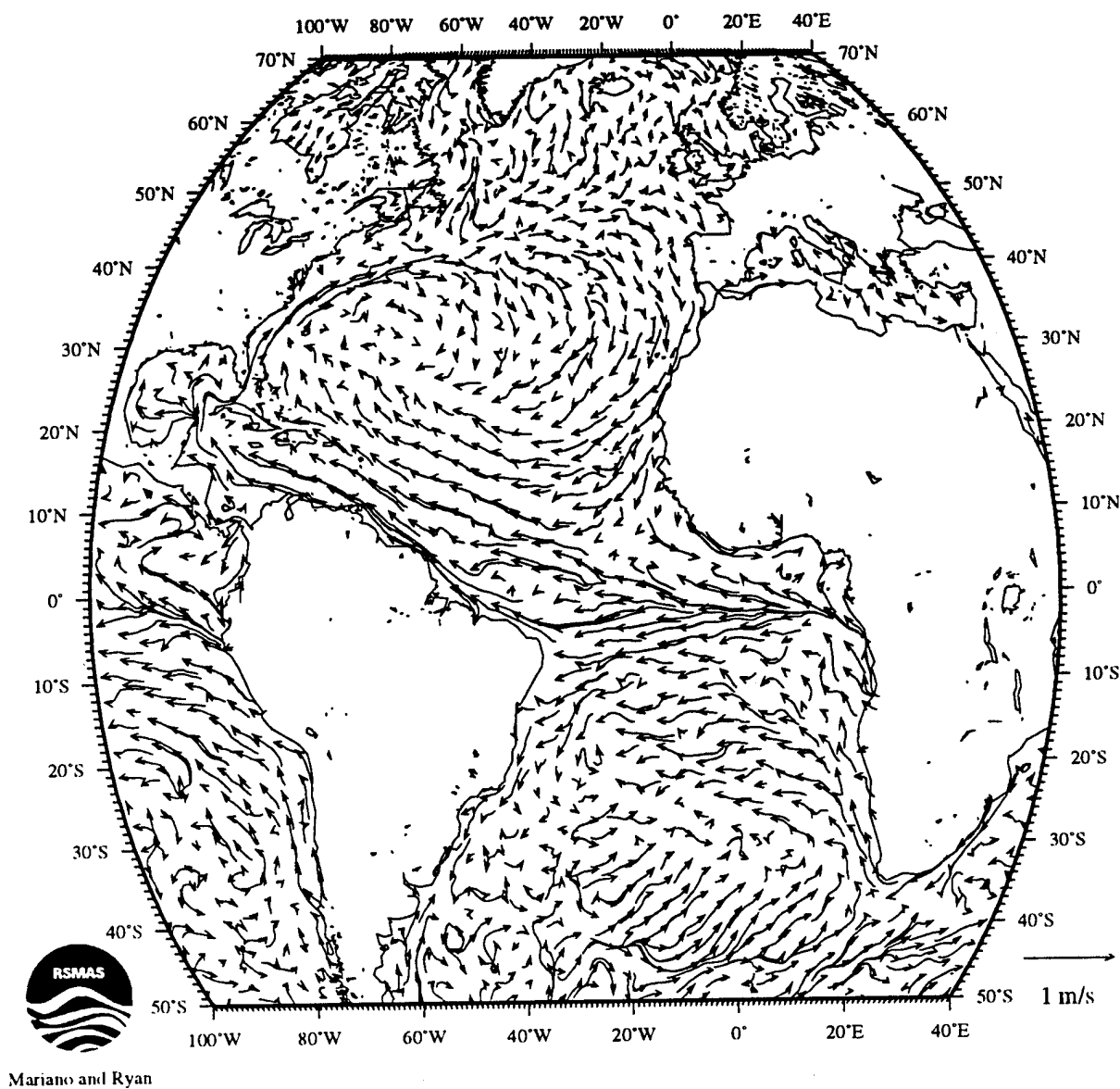
5(c)

Winter Average



5(d)

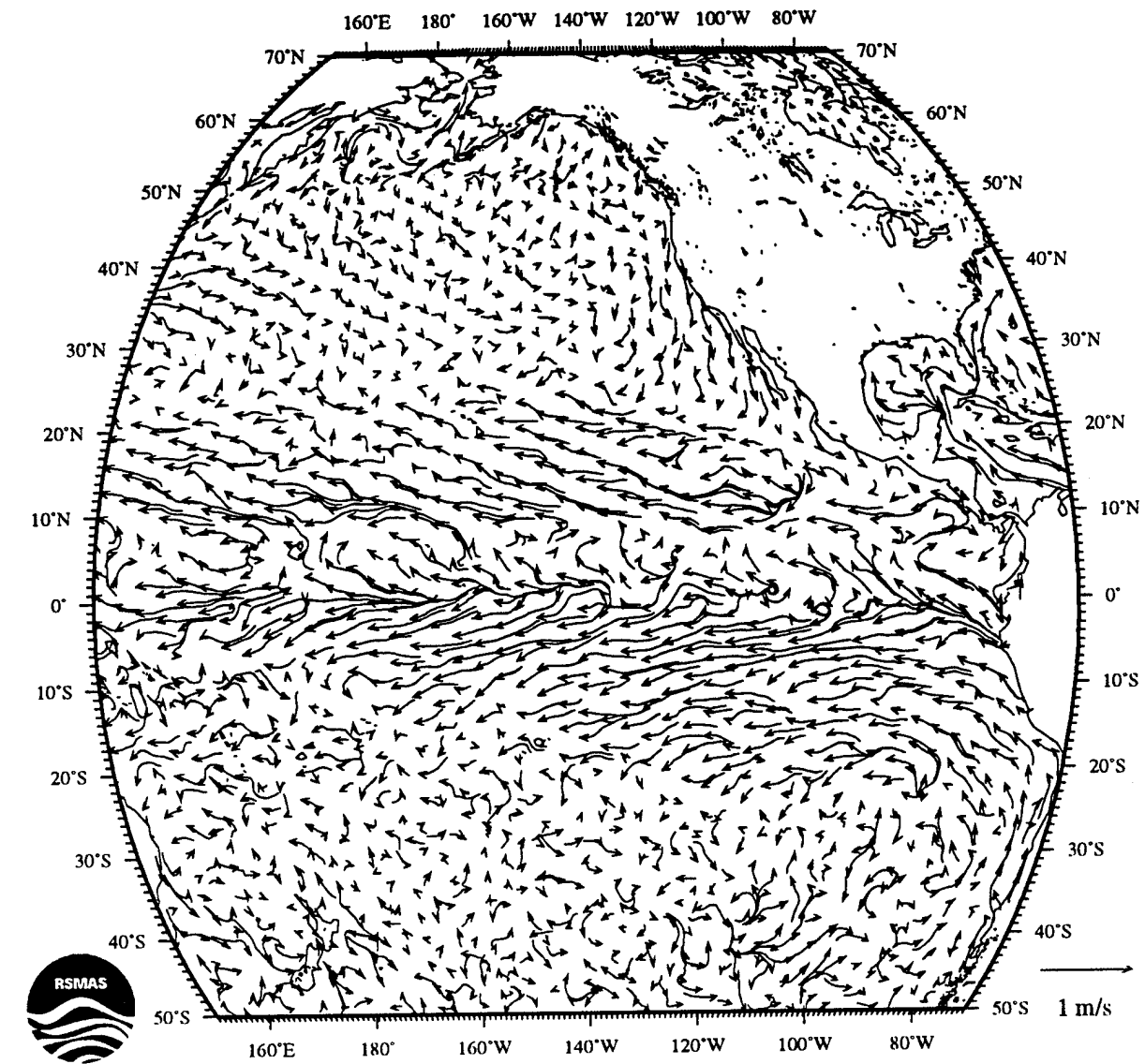
Spring Average



6(a)

Fig. 6 (a-d) The weighted seasonal average of the monthly surface velocity estimates for the months of March, April, and May.

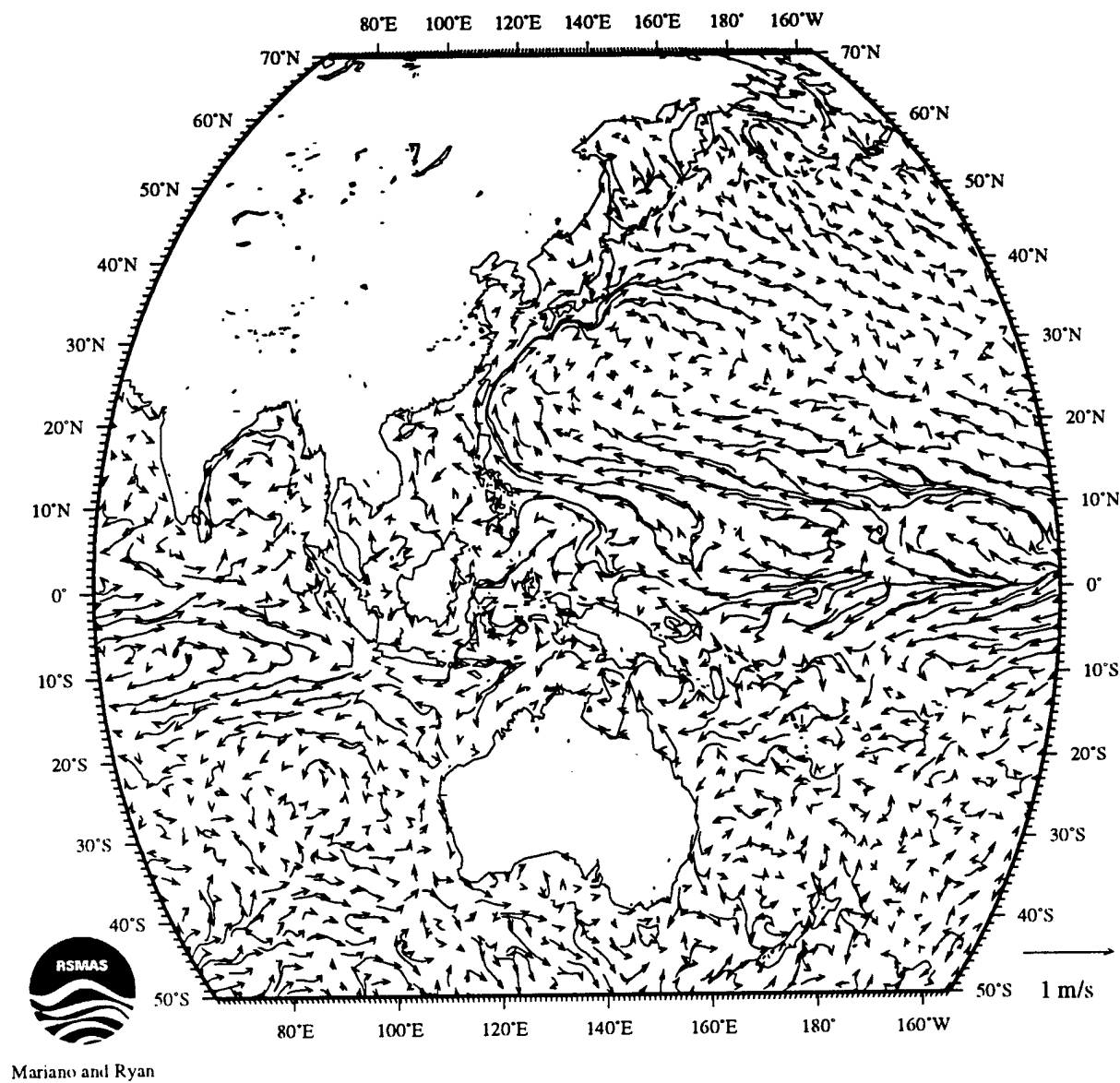
Spring Average



Mariano and Ryan

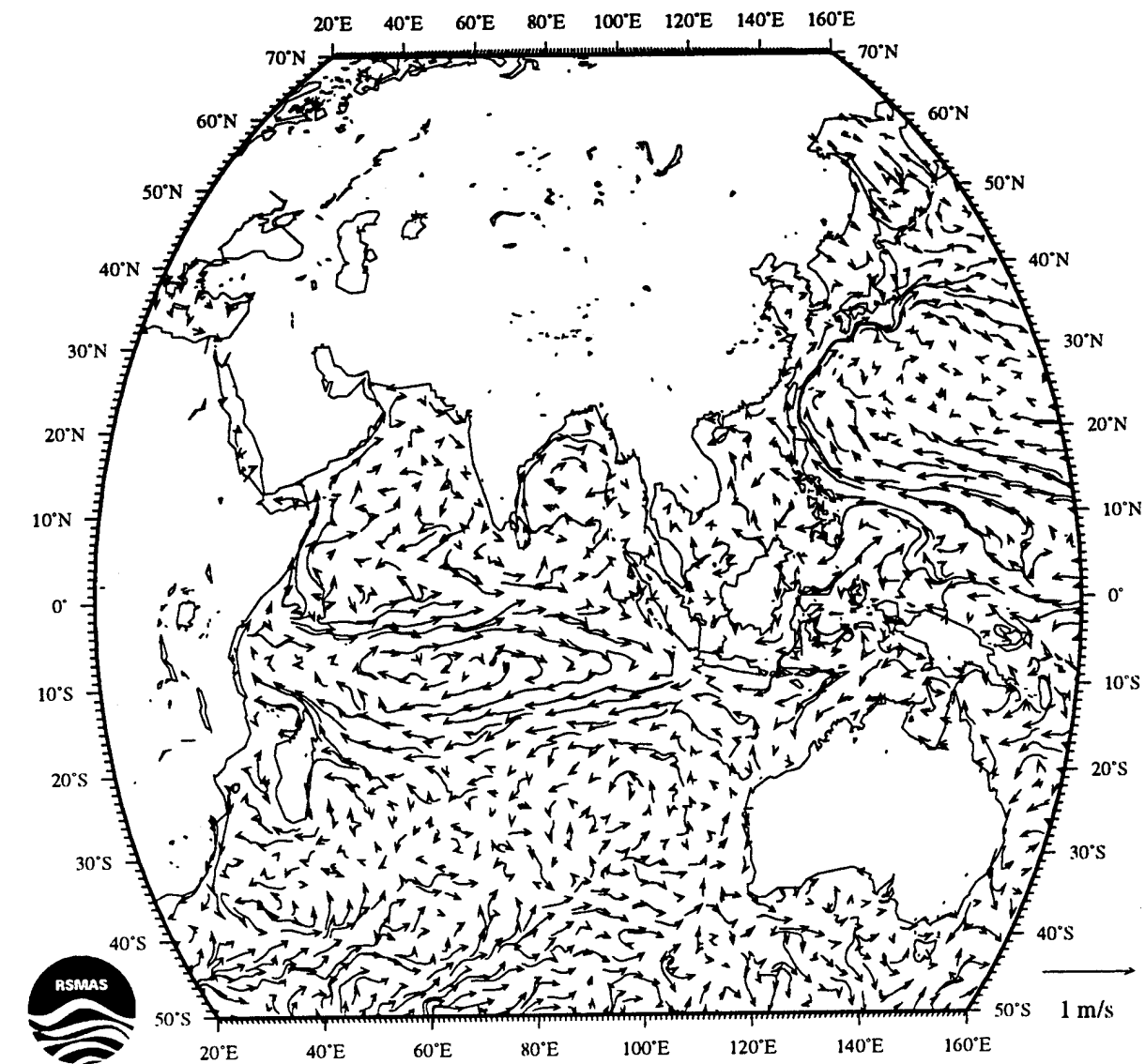
6(b)

Spring Average



6(c)

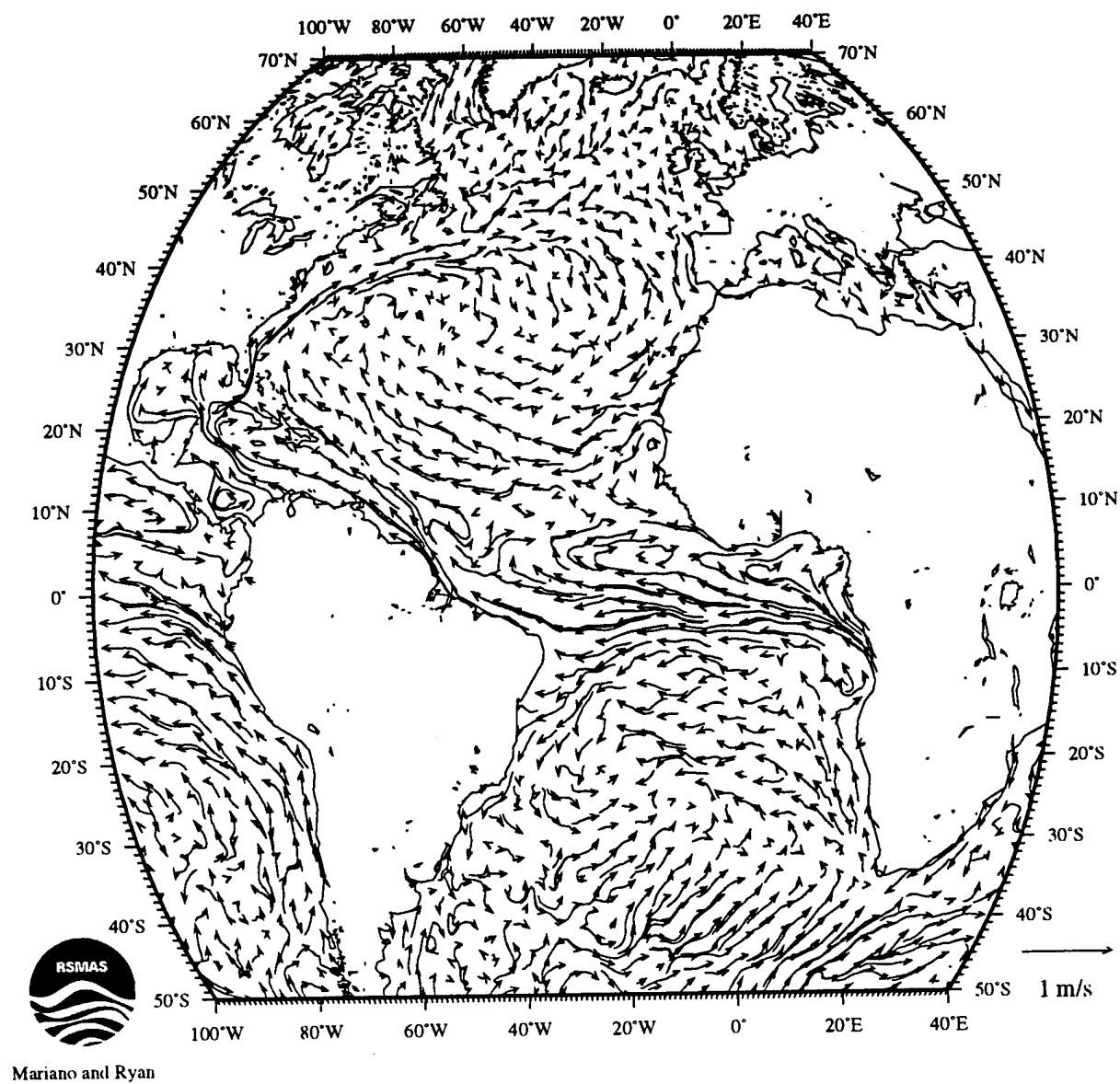
Spring Average



Mariano and Ryan

6(d)

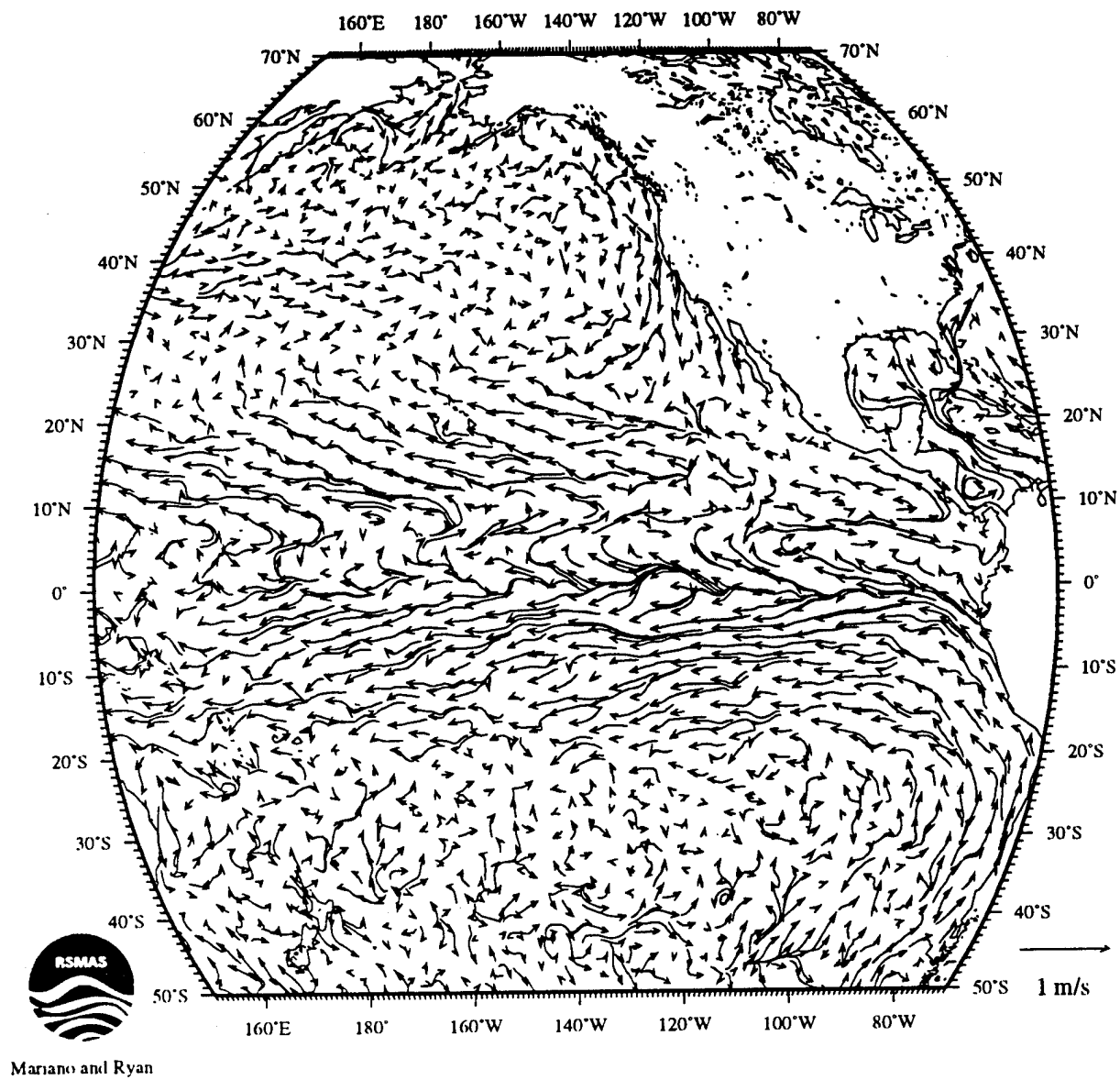
Summer Average



7(a)

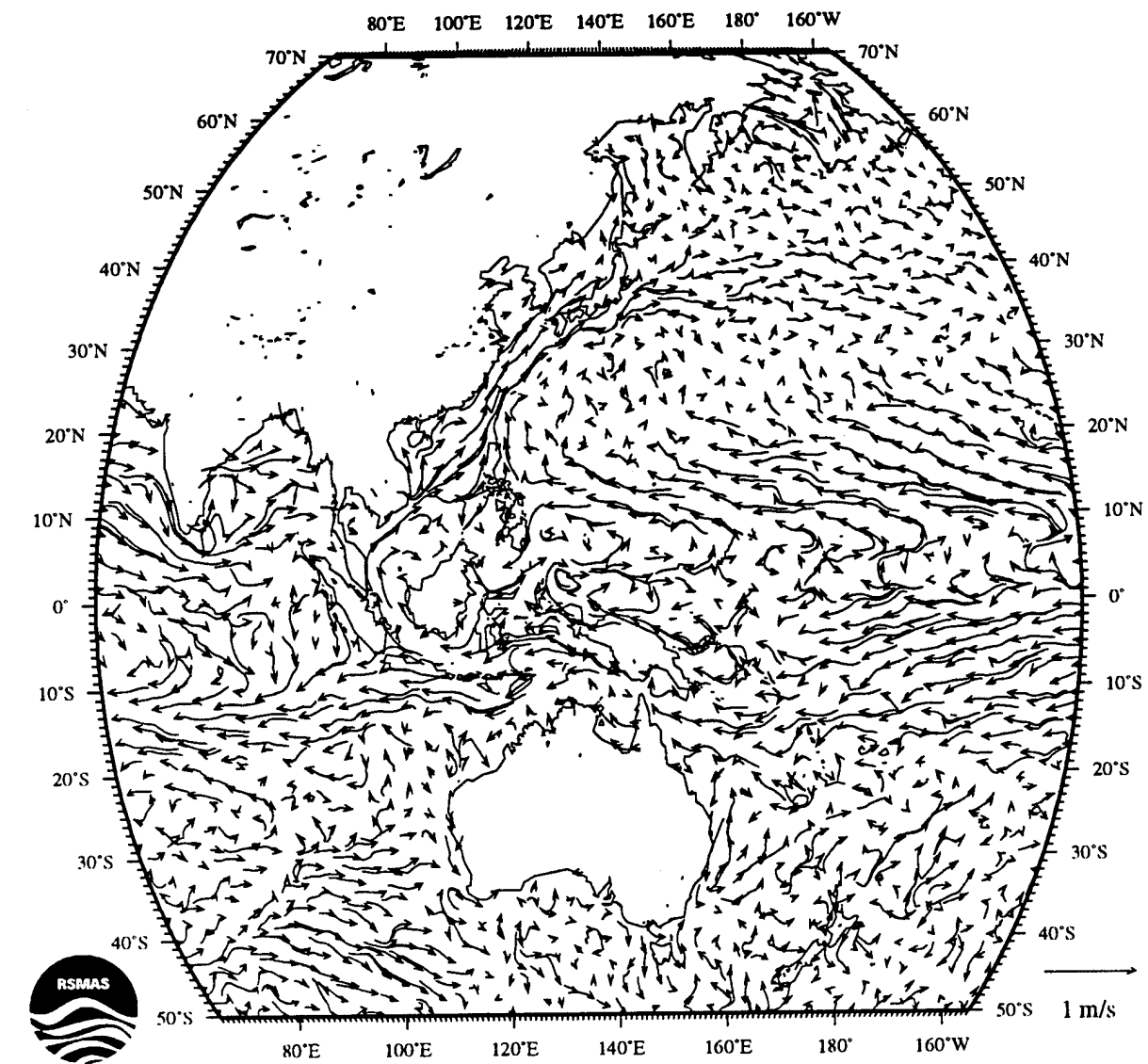
Fig. 7 (a-d) The weighted seasonal average of the monthly surface velocity estimates for the months of June, July, and August.

Summer Average



7(b)

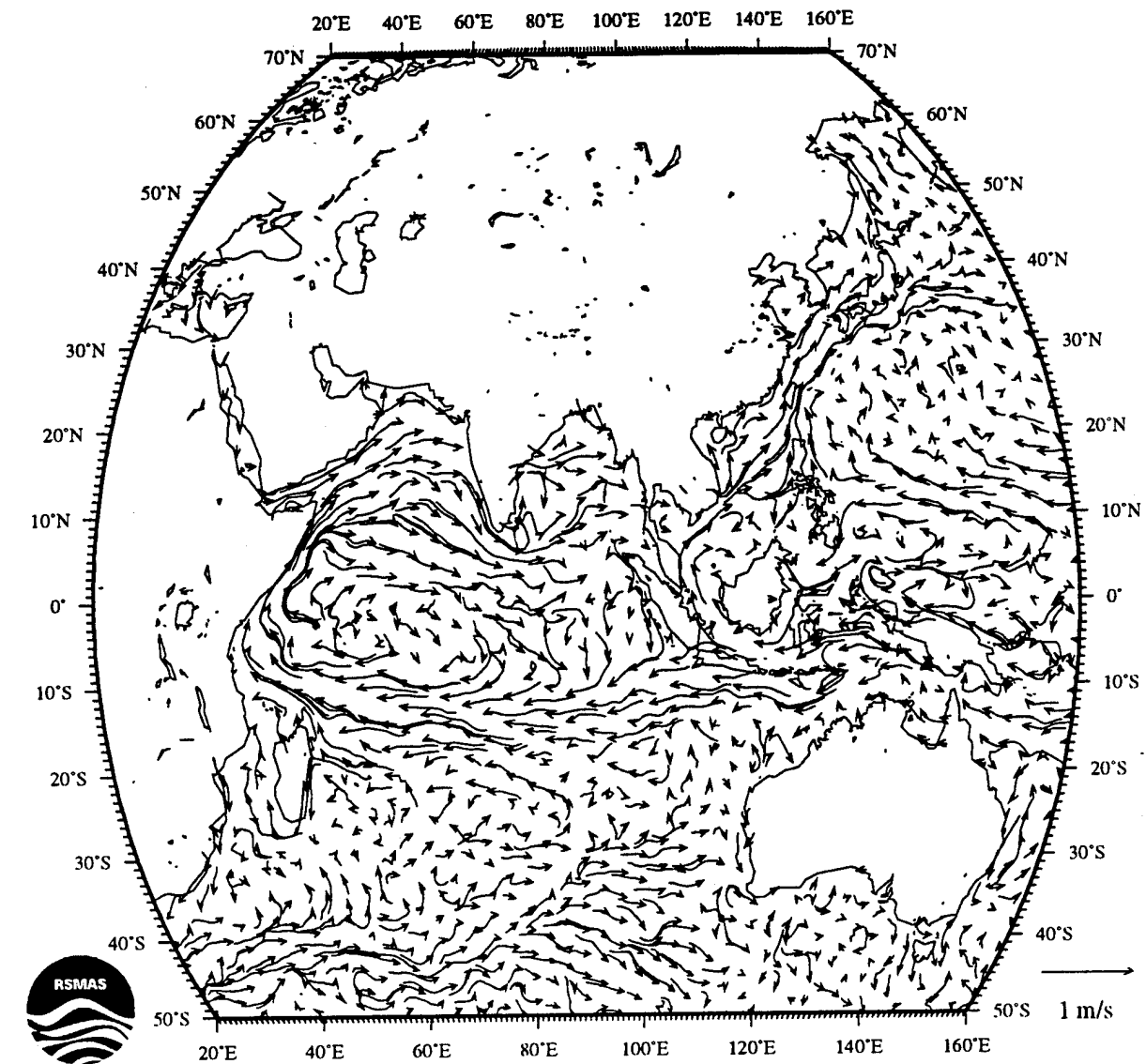
Summer Average



Mariano and Ryan

7(c)

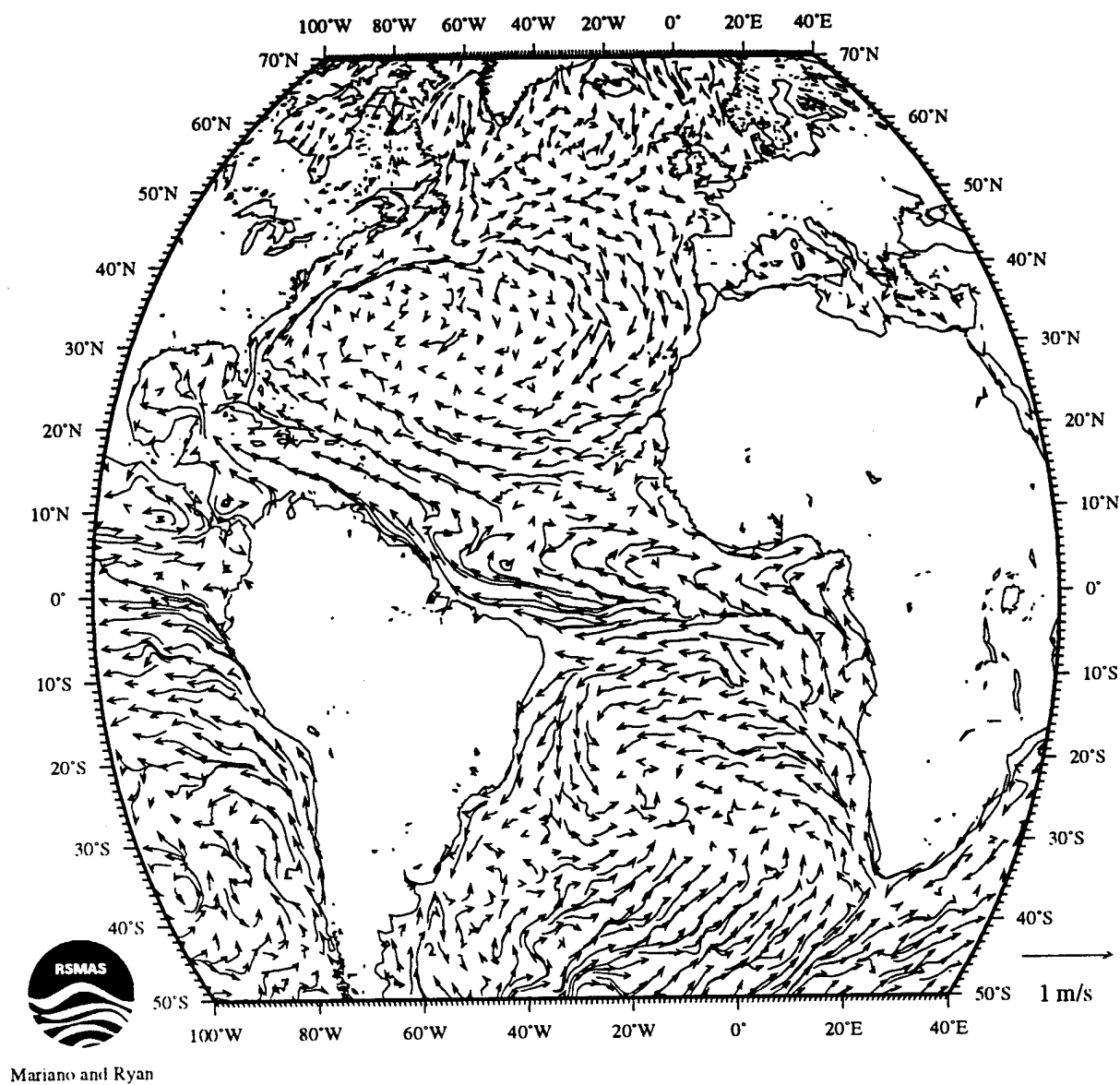
Summer Average



Mariano and Ryan

7(d)

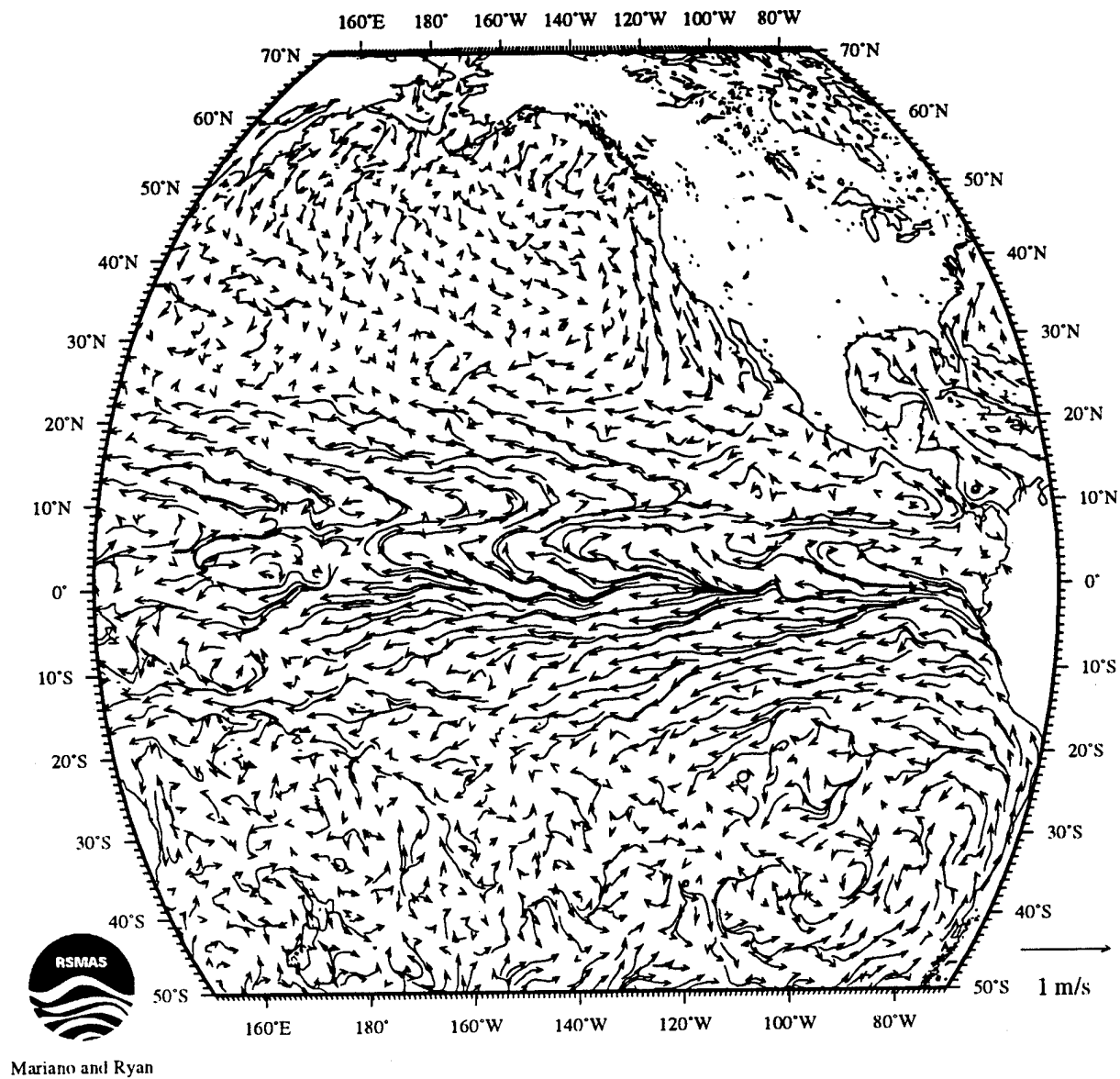
Fall Average



8(a)

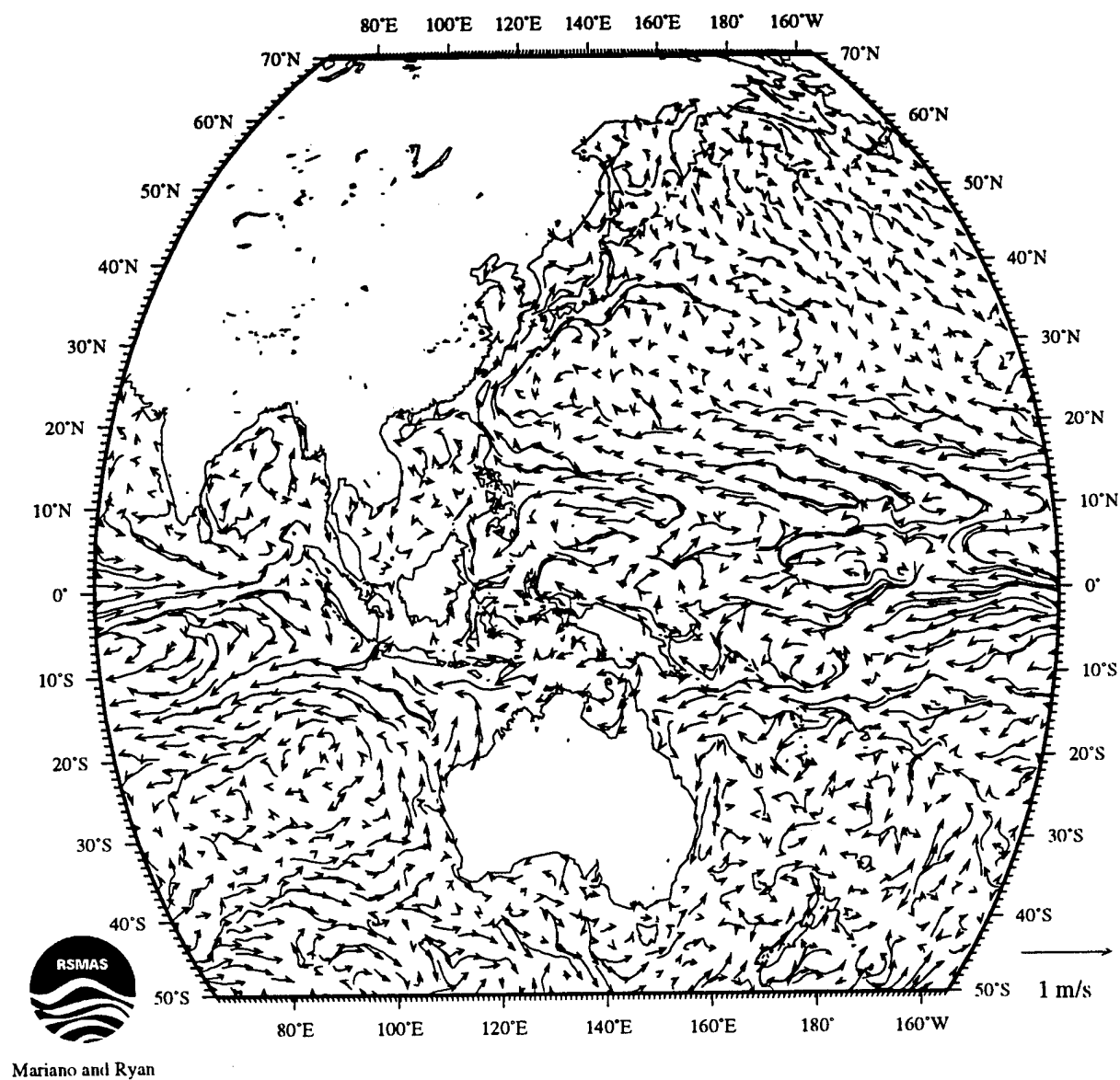
Fig. 8 (a-d) The weighted seasonal average of the monthly surface velocity estimates for the months of September, October, and November.

Fall Average



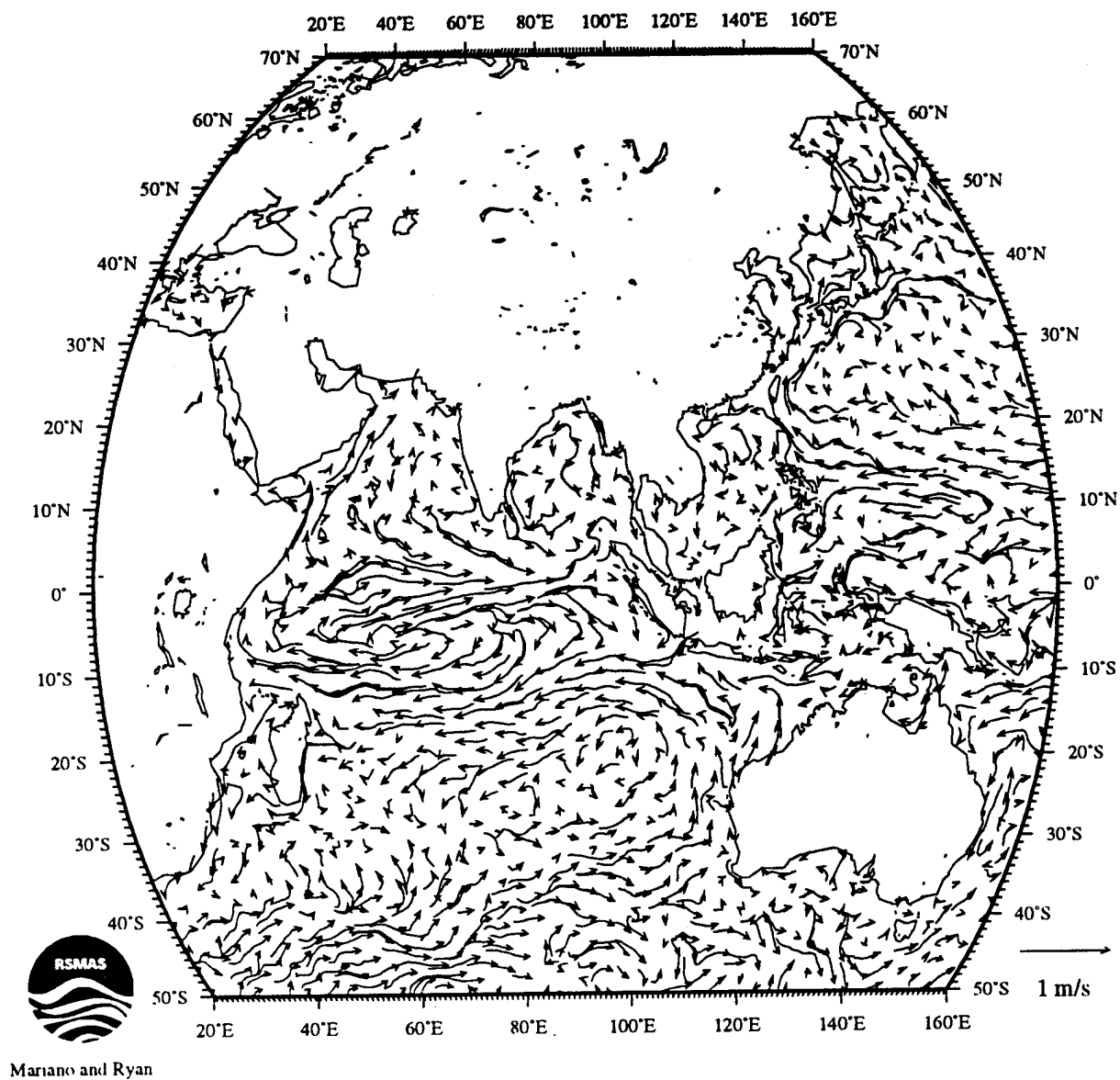
8(b)

Fall Average



8(c)

Fall Average



8(d)

Atlantic - Indian

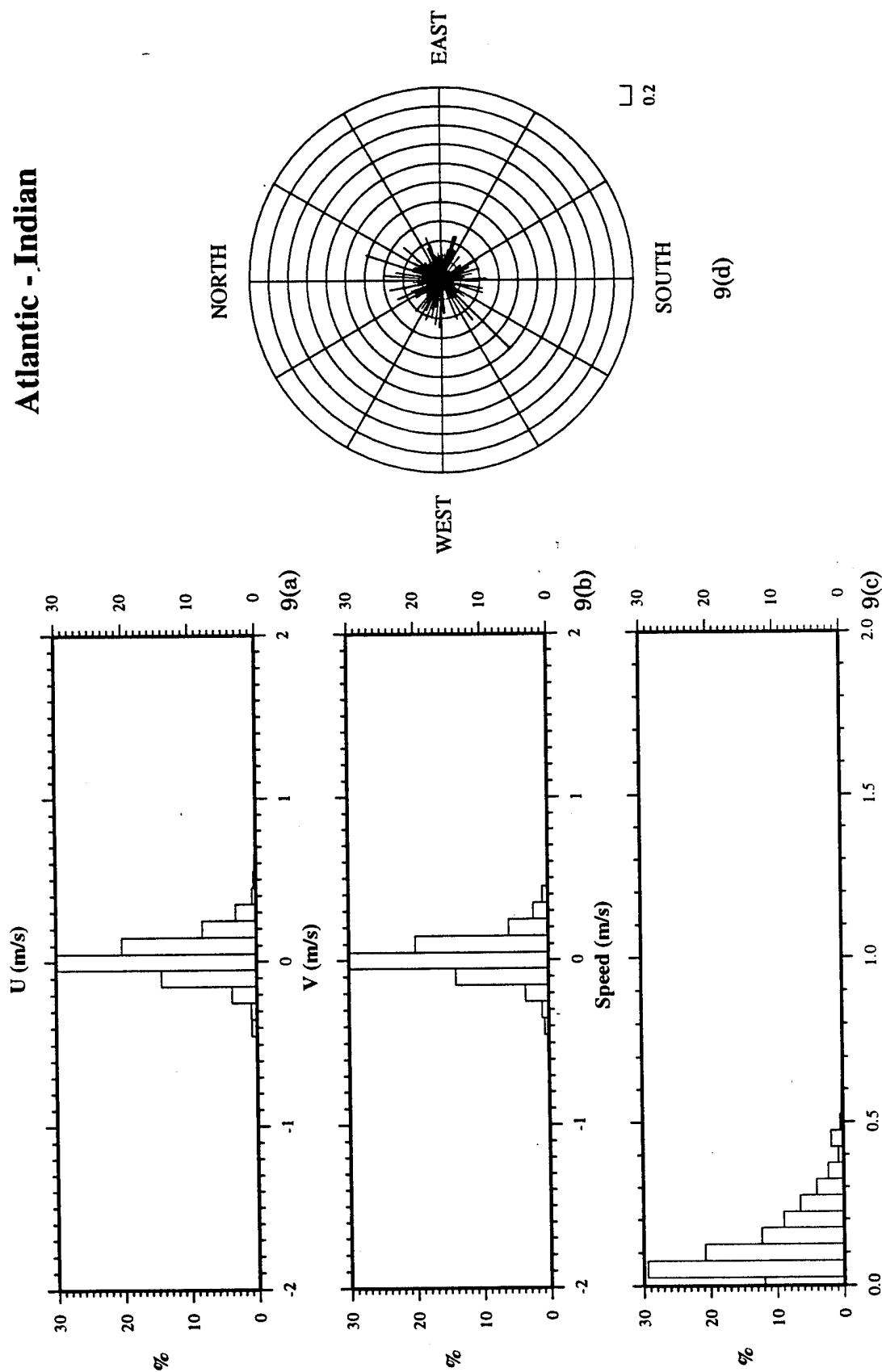


Fig. 9 (a) The histogram of the east-west velocity component, u , differences between the Atlantic minus the Indian basins. (b) Same as (a) but for v . (c) A histogram of the speed differences, $\sqrt{(u_{Atlantic} - u_{Indian})^2 + (v_{Atlantic} - v_{Indian})^2}$. (d) A polar plot of the vector differences between the Atlantic and Indian basin velocities.

with u_1 . Let u_2 and e_2 be the corresponding variables for the second region. Then the final estimate, u_f , in the data overlap regions should be given by

$$u_f = \frac{(e_2 u_1 + e_1 u_2)}{(e_1 + e_2)}.$$

A similar expression is used for v . The final blending of the data in the overlap regions is the responsibility of the USCG R&D Ctr.

In addition to the velocity estimates, the corresponding uncertainties, σ_u and σ_v , were also estimated. Since the OA scheme usually calculates estimation errors for the problem of estimating fields at a given estimation time, and since our estimates here represent historical averages, the typical error formula (e.g., eqn. (3.8)) was modified. For each bin, month, and velocity component, the maximum normalized error level, given by eqn. (3.8), was calculated, c_{max} . At each grid point i , the standard deviation of the original data, σ_i , was calculated from the spline fits to the standard deviation fields. Let the normalized error level given by eqn. (3.8.1) be denoted by c . Then the error of our estimates for each component was estimated by

$$c(\sigma_i / c_{max}).$$

The final velocity fields were smoothed by a first-order Shapiro filter (Shapiro, 1970) to remove one-point grid noise for mapping purposes.

5. Limitations and Recommendations

The MGSVA was constructed on a volunteer basis from a noisy, unedited, but very large data set using existing software developed for another project. Consequently, these fields can be improved in a number of ways by generalizing our application of the parameter matrix objective analysis algorithm and the data processing steps. This suggestion is discussed after the discussion of the inherent limitations of Eulerian-based field estimates for USCG operational use.

Ocean current systems meander or change their positions on a variety of time scales. Coherent circulation features, such as Gulf Stream warm-core rings, frontal eddies, planetary waves, etc., also propagate over time. It is a well-known fact that Eulerian-based estimates will smear circulation features and that consequently, these estimates are too slow (Mariano, 1990; Mariano and Chin, 1995). Halkin and Rossby (1985) demonstrated that peak Gulf Stream velocities in an Eulerian-based average of 16 PEGASUS velocity transects were significantly lower than peak velocities in a stream-based average of the same data, 100 *cm/s* versus 170 *cm/s*. Another example of smearing in Eulerian averaging is shown in Fig. 10, where two estimates of the Gulf Stream front, a warm- and cold-core ring, are averaged using a simple arithmetic average with equal weights of 1/2. This is the simplest example that is relevant to this OA method. Note that the field values are reduced by a factor of 1/3 between the two data realizations. The contour-based averaging approach of Mariano (1990) shows a more pleasing estimate. The construction of contour-based software for averaging the MMSDD would be a daunting task because of the sophisticated pattern recognition that is needed for contour-based averaging and is not recommended at this time.

A far more economical approach is to compare actual drift velocities calculated in the field and to compare these to our estimates and then tabulate the amount of underestimation. A smooth regional map of a correction factor, which may be a function of wind speed, could then be applied to the OA estimates to correct some of the underestimation problems of least-square Eulerian-based estimates.

Based on the lead author's experience with feature-based estimation, it is suggested that, in Computer Aided Search Planning (CASP) 2.0 or its successor, a feature-based combination of real-time and historical data would lead to a major improvement in SAR activities because many ocean features have a fairly "rigid" cross-feature structure. For example, in the Gulf Stream, two nearby drifters yield two estimates of velocities and an estimate of a spatial derivative. This would tell you what side of the Gulf Stream you are on and, by translating the historical Gulf Stream cross-stream velocity structure so that it matches the *in situ* data, would provide a much better estimate of the true velocity field than either an OA map of the drifter velocity or the MGSVA data by itself or any Eulerian-based scheme. As Fig. 10 illustrates, a simple average of the two will lead to undesirable results, and a feature-based approach should be tried.

As demonstrated by Robinson and collaborators (Robinson *et al.*, 1989; Spall and Robinson, 1990), feature models can be used with satellite data for reliable

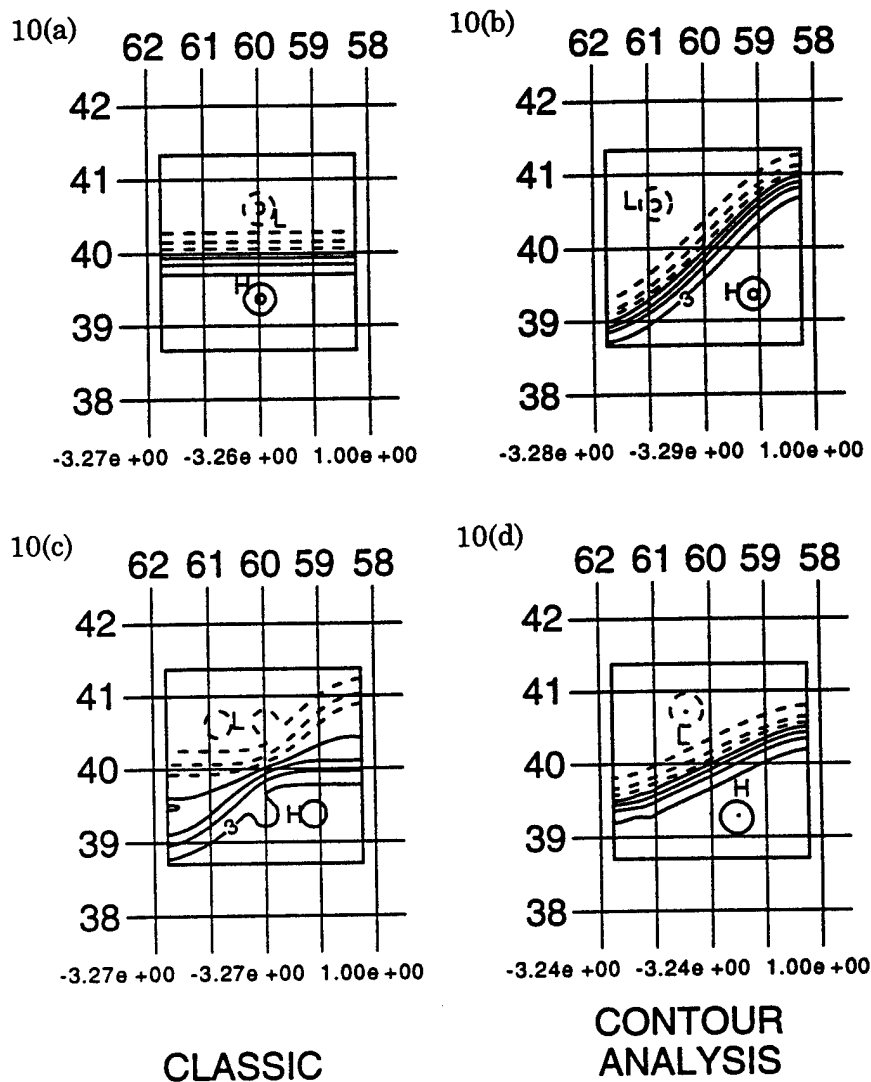


Fig. 10 (a) An estimate of the streamfunction field for the Gulf Stream region. (b) Same as (a) but another field estimate. (c) Using the classic optimal estimator for combining fields from (a) and (b). (d) Using contour analysis.

prediction of ocean currents. For example, Advanced Very High Resolution Radiometry (AVHRR) data in the Gulf Stream can convey the position of the stream and the rings. Analytical three-dimensional models of the circulation associated with these features can be combined with other data.

In addition to the incorporation of feature-based analysis, a number of possible generalizations in the OA of these data can be tried. These can be broadly characterized into two separate research avenues: improvements to the input data sets and improvements to the OA algorithm. The input data set can be improved by assigning error bars to the individual drifts since our OA algorithm allows this. For instance, earlier estimates would be given larger bars based on progress in navigation technology. The input data can also be improved by incorporating the newest global drifting buoy data set from the Tropical Ocean Global Atmosphere (TOGA) experiment, the World Ocean Circulation Experiment (WOCE), and other major ocean field experiments, and finding the missing ship-drift estimates. This should be done as quickly as possible, especially in the data-sparse southern oceans. Given the planned experiments in the southern ocean, new data would certainly lead to major improvements in surface velocity estimates south of 50° S.

Another potential research avenue that may lead to major improvement in this product for SAR activities would involve first binning the data by wind speeds and directions. Velocity would then be estimated as a function of longitude, latitude, and wind regime. Visual inspection of surface drifters in a Mineral Management Service video (courtesy of Walter Johnson) show a strong correlation between wind direction and drifter movement. It is very evident from drifter trajectories in this video and, given our estimates for this region, that the inherent variability between different wind regimes will lead to very sub-optimal SAR drift predictions during anomalous wind conditions. This is unavoidable with our historical analysis because of the strongly stochastic nature of both the ocean and the atmosphere. It is highly recommended that this research avenue be given high priority.

The most obvious extension to our OA algorithm would be the use of a multi-variate OA scheme that incorporates the cross-correlations between u and v . Carter and Robinson (1987) showed that their multi-variate approach improved velocity estimates in the thermocline waters in the NW Atlantic over scalar OA techniques using the POLYMODE data set. This improvement was expected for subsurface thermocline flow where geostrophic dynamics lead to a significant cross-correlation between u and v . Our analysis indicates significant horizontal divergences and convergences in some areas that would cause smaller correlations between u and v . Thus, the calculation of regional and monthly cross-correlations between u and v should be given high priority so that the issue of scalar versus multi-variate OA scheme for surface velocity estimates can be settled.

OA also requires the specification of a mean and covariance function at each of the estimation points. This specification is the most subjective component of an OA scheme. A tense, smooth, two-dimensional bicubic spline was used to represent the mean field for each variable u and v . The spline was fitted to the monthly data sets in all of the basins except for the Antarctic, where data-sparsity dictated just one annual analysis. These splines have been used by the lead author extensively

and have performed well for global SST estimation (MB92). Further experiments with the adjustable spline parameters is needed, but is not crucial. The choice of other mean surfaces would probably not improve the velocity estimates.

Because only a finite number of data points are used in estimating the correlation functions, correlation function estimates are the convolution of the true correlation function with a sinc function. Thus, with limited data, it is hard to determine the correct form of the correlation function. Work by Thiébaux (1975, 1976) suggests that if a fairly general correlation function is used and fitted to the data, it would be hard to improve on it unless more data become available. So the choice of another correlation function would probably not improve the velocity estimates, but using the full power of the parameter matrix may. This would require estimation of the nine correlation parameters in each $5^\circ \times 5^\circ$ bin for each basin for each month. This computer-intensive component should be first tested in a region that is both of high priority to the USCG and that contains numerous good quality data sets for comparisons. The latter requirement for a large-scale region would certainly suggest the tropical Pacific, seeded with numerous drifters since the 1980s and with a large increase in data-density starting in the early 1990s. A small-scale region, also of interest to the International Ice Patrol, is the NW Atlantic Ocean.

In the near future, assimilation of either the edited data or the mapped fields into numerical circulation models forced by real-time winds is recommended for the long-term goal of accurate and reliable sea surface velocity estimates that are needed for USCG activities. Robinson and his research group at Harvard University have shown that this is feasible, given the necessary resources. A group at the National Weather Center is working with Mellor (Princeton) on such a system for the East Coast. Rosenstiel School of Marine and Atmospheric Science Ocean Pollution Research Center at the University of Miami is experimenting with a very high resolution numerical circulation model in the Florida Straits area and has been supplying forecasts to the Brother to the Rescue for finding Cuban rafters lost at sea. In the long run, although this is more costly, it is the best approach. In the interim, we hope that the use of the MGSVA in USCG SAR operations will lead to more optimal searching strategies, save lives, and reduce fuel consumption through more efficient SAR flights.

Acknowledgements

Special thanks to the US Naval Oceanographic Office for the ship-drift data. Comments by Ms. Louise M. Jacques Mariano and Ms. Corinne Weber improved this report in many ways. We would like to thank Ms. Mariano for this and her assistance in the calculations. Mr. Steven Ricard deserves special praise for logistical and technical support. The curly arrow plotting code used here was based on the code developed by Mr. Alan J. Wallcraft and was supplied to us by Dr. Harley Hurlburt of the Naval Research Lab.

Bibliography

- Arnault, S., 1987. Tropical Atlantic geostrophic currents and ship drifts. *J. Geophys. Res.*, **92**, 5076-5088.
- Bergman, K.D. 1978. Role of observational errors in optimum interpolation analysis. *Bull. Am. Meteor. Soc.*, **59**, 1603-1611.
- Bowditch, N., 1984. *American Practical Navigator*. Defense Mapping Agency.
- Bretherton, F.P., R.E. Davis and C.B. Fandry, 1976. A technique for objective analysis and design of oceanographic experiments applied to MODE-73. *Deep Sea Res.*, **23**, 559-582.
- Carter, E.F., 1983. Statistics and dynamics of ocean eddies. Ph.D. thesis Harvard (18 in *Reports in Meteorology and Oceanography*, Harvard U.).
- Carter, E.F. and A.R. Robinson, 1987. Analysis models for the estimation of oceanic fields. *J. of Atmos. and Ocean. Techn.*, **4**(1), 49-74.
- Chelton, D., 1983. Effects of sampling errors in statistical estimation. *Deep-Sea Res.*, **30**, 1083-1103.
- Clancy, R.M., 1983. Effect of observational error correlations on objective analysis of ocean thermal structure. *Deep Sea Res.*, **30** (9A), 985-1002.
- Davis, R.E., 1985. Objective Mapping by Least Squares Fitting. *J. Geophys. Res.*, **90** (C3), 4773-4777.
- Daley, R., 1991. *Atmospheric Data Analysis*, Cambridge University Press. Cambridge, U.K.
- Denman, K.L. and H. J. Freeland, 1985. Correlation scales, objective mapping and a statistical test of geostrophy over the continental shelf. *J. Mar. Res.*, **43**, 517-539.
- Freeland, H.J. and W.J. Gould, 1976. Objective analysis of mesoscale ocean circulation features. *Deep Sea Res.*, **23**, 915-923.
- Gandin, L.S., 1963. *Objective Analysis of Meteorological Fields*. Translated from the Russian by Israel Program for Scientific Translations, 1965, 242pp. [NTIS TT65-50007.]
- Ghil, M.S., S. Cohn, J. Travantzis, K. Bube and E. Issacson, 1981. Application of Estimation Theory to Numerical Weather Prediction. In: *Dynamic Meteorology. Data Assimilation methods* (L. Bengtsson, M. Ghil and E. Kallen, editors), Springer-Verlag, New York, 139-284.
- Halkin, D. and H. T. Rossby, 1985. The structure and transport of the Gulf Stream at 78° W. *J. Phys. Oceanogr.*, **15**, 1439-14562.
- Hansen, D.V. and A. Herman, 1989. Temporal Sampling Requirements for Surface Drifting Buoys in the Tropical Pacific. *J. of Atmos. and Oceanog. Tech.*, **6** (4), 599-607.
- Inoue, H., 1986. A Least-Square Smooth Fitting for Irregularly Spaced Data. Finite-Element Approach Using the Cubic B-Spline Basis. *Geophysics*, **51** (11), 2051-2066.
- Journel, A. G. and C. S. Huijbregts, 1978. *Mining Geostatistics*, Academic Press, 600 pp.

- Julian, P.R. and H.J. Thiébaux, 1975. On some properties of correlation functions used in optimum interpolation schemes. *Mon Wea Rev.*, **103**, 605-616.
- Mariano, A.J., 1990. Contour Analysis. A new approach for melding geophysical fields. *J. of Atm. and Ocean. Techn.*, **7**(2), 285-295.
- Mariano, A.J. and O. Brown, 1992. Efficient objective analysis of dynamically heterogeneous and nonstationary fields via the parameter matrix. *Deep-Sea-Res.*, **39** (7/8), 1255-1271.
- Mariano, A.J. and T.M. Chin, 1995. Feature and Contour based data analysis and assimilation in physical oceanography. *Stochastic Modelling in Physical Oceanography*, (Alder, Muller, and Rozovskii, eds.), Birkhauser, (In press).
- McPhaden, M.J., D.V. Hansen, P.L. Richardson, 1991. A comparison of ship drift, drifting buoy, and current meter mooring velocities in the Pacific South Equatorial Current. *J. Geophys. Res.*, **96** (C1), 775-781.
- McWilliams, J. C., E. D. Brown, H. L. Bryden, C. C. Ebbesmeyer, B. A. Elliott, R. H. Heinmiller, B. Lien Hua, K. D. Leaman, E. J. Lindstrom, J. R. Luyten, S. E. McDowell, W. Breckner Owens, H. Perkins, J. F. Price, L. Regier, S. C. Riser, H. T. Rossby, T. B. Sanford, C. Y. Shen, B. A. Taft, and J. C. Van Leer, 1983. 5. The local dynamics of eddies in the Western North Atlantic. In: *Eddies in Marine Science*, A. R. Robinson, ed., Springer-Verlag, Berlin, Heidelberg.
- McWilliams, J.C., W.B. Owens, and L.B. Hua, 1986. An objective analysis of the POLYMODE Local Dynamics Experiment. Part I. General formalism and statistical model selection. *J. Phys. Ocean.*, **16**, 483-504.
- Meehl, G.A., 1982. Characteristics of surface current flow inferred from a global ocean current data set. *J. Phys. Oceanogr.*, **12**(6), 538-555.
- Perkins, B.D. and R.L. Marsee, 1994. Florida Gulf Stream Data (FLDA) current study. *CG-D-94*.
- Richardson, P.L. and T.K. McKee, 1984. Average seasonal variation of the Atlantic equatorial currents from historical ship drifts. *J. Phys. Oceanogr.*, **14**, 1226-1238.
- Richardson, P.L. and S.G.H. Philander, 1987. The seasonal variations of surface currents in the tropical Atlantic Ocean: A comparison of ship drift data with results from a general circulation model. *J. Geophys. Res.*, **92**, 715-724.
- Richardson, P.L. and D. Walsh, 1986. Mapping climatological seasonal variations of surface currents in the tropical Atlantic using ship drifts. *J. Geophys. Res.*, **91**, 10,537-10,550.
- Richardson, P.L. and G. Reverdin, 1987. Seasonal cycle of velocity in the Atlantic North Equatorial Counter Current as measured by surface drifters, current meters and ship drifts. *J. Geophys. Res.* **92**, 3691-3708.
- Richardson, P.L., 1989. Worldwide ship drift distributions identify missing data. *J. Geophys. Res.* **94** (C5), 6169-6176.

- Robinson, A.R., A. Hecht, N. Pinardi, J. Bishop, W.G. Leslie, Z. Rosentroub, A.J. Mariano and S. Brenner, 1987a. Small synoptic/mesoscale eddies and energetic variability of the eastern Levantine basin. *Nature*, **327**, 131-134.
- Robinson, A.R., Spall, M.A., Walstad, L.J., and Leslie, W.G., 1989. Data assimilation and dynamical interpolation in Gulfcast experiments. *Dyn. Atmos. Oceans*, **13** (3-4), 269-300.
- Seaman, R.S., 1977. Absolute and differential accuracy of analyses achievable with specified observational network characteristics. *Mon Wea. Rev.*, **105** (10), 1211-1222.
- Shapiro, 1970. Smoothing, Filtering and Boundary Effects. *Reviews of Geophysics and Space Physics*, **8** (2), 359-387. 1147-1160.
- Spall, M.A. and A. R. Robinson, 1990. Regional primitive equation studies of the Gulf Stream meander and ring formation region. *J. Phys. Oceanogr.*, **20**, 985-1016.
- Stidd, C.K., 1975. Meridional profiles of ship drift components. *J. Geophys. Res.*, **80** (12), 1679-1682.
- Thièbaux, J.J. 1974. Estimation of covariances of meteorological parameters using local-time averages. *J. Appl. Meteor.*, **13**, 592-600.
- Thièbaux, J.J. 1975. Experiments with correlation representations for objective analysis. *Mon. Wea. Rev.*, **103**, 617-627.
- Thièbaux, J.J. 1976. Anisotropic correlation functions for objective analysis. *Mon. Wea. Rev.*, **104**, 994-1002.
- Thièbaux, J.J. and M.A. Pedder, 1987. *Spatial Objective Analysis with applications in atmospheric science*. Academic Press, 299 pp.
- U.S. Naval Oceanographic Office, 1955. Atlas of Pilot Charts, *Central American Waters and South Atlantic Ocean*. Publ. 106, Defense Mapping Agency Hydrographic Center, Washington, DC, 16 pp.
- U.S. Naval Oceanographic Office, 1966. Atlas of Pilot Charts, *South Pacific and Indian Oceans*. Publ. 107, Defense Mapping Agency Hydrographic Center, Washington, DC, 16 pp.
- U.S. Naval Oceanographic Office, 1979a. Atlas of Pilot Charts, *Indian Ocean*. Publ. 109, Defense Mapping Agency Hydrographic Center, Washington, DC, 12 charts.
- U.S. Naval Oceanographic Office, 1979b. Atlas of Pilot Charts, *North Pacific Ocean*. 55 Series, Defense Mapping Agency Hydrographic Center, Washington, DC, 12 charts.
- U.S. Naval Oceanographic Office, 1979c. Atlas of Pilot Charts, *North Atlantic Ocean*. 16 Series, Defense Mapping Agency Hydrographic Center, Washington, DC, 12 charts.
- Viekman, B.E. and T.L. Jordan, 1991. Evaluation of the Levitus Sea Surface Current Climatology for use in Coast Guard Computer Assisted Search Planning. *Tech. Rep. 5-91*. Center for Advanced Studies, USCG Academy, New London, Ct.

- Watts, D.R., K.L. Tracey and A.I. Friedlander, 1989. Producing accurate maps of Gulf Stream thermal front using objective analysis. *J. Geophys. Res.*, **94** (c6), 8040-8052.
- Wyrski, K., L. Magaard, and J. Hager, 1976. Eddy energy in the oceans. *J. Geophys. Res.*, **81** (15), 2641-2646.

ABBREVIATIONS AND ACRONYMS

AVHRR	Advanced Very High Resolution Radiometry
EKE	Eddy Kinetic Energy
MGSVA	Mariano Global Surface Velocity Analysis
MMSDD	Monthly Maury Ship Drift Data base
MSDD	Maury Ship Drift Data base
NAVOCEANO	U.S. Naval Oceanographic Office
NEC	North Equatorial Current
NECC	North Equatorial Counter Current
OA	Objective Analysis
RSMAS	Rosenstiel School of Marine and Atmospheric Science
SAR	Search and Rescue
SEC	South Equatorial Current
SST	Sea Surface Temperature
USCG	U.S. Coast Guard