JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 103, NO. C4, PAGES 7491–7509, APRIL 15, 1998

An optimized design for a moored instrument array in the tropical Atlantic Ocean

Eric C. Hackert
Hughes STX Corporation, NASA Goddard Space Flight Center, Greenbelt, Maryland

Robert N. Miller
College of Oceanic and Atmospheric Sciences, Oregon State University, Corvallis

Antonio J. Busalacchi
Laboratory for Hydrospheric Processes, NASA Goddard Space Flight Center, Greenbelt, Maryland

Abstract. This paper presents a series of observing system simulation experiments (OSSEs) which are intended as a design study for a proposed array of instrumented moorings in the tropical Atlantic Ocean. Fields of TOPEX/Poseidon sea surface height anomalies are subsampled with the goal being reconstruction of the original fields through the use of reduced-space Kalman filter data assimilation at a restricted number of locations. Our approach differs from typical identical and fraternal twin experiments in that real observed data (i.e., TOPEX/Poseidon data) are subsampled and used in place of synthetic data in all phases of the OSSEs. In this way the question of how closely a particular model-generated data set resembles nature is avoided. Several data assimilation runs are performed in order to optimize the location of a limited number of moorings for the proposed Pilot Research Moored Array in the Tropical Atlantic (PIRATA). Results of experiments in which data are assimilated at 2°N, 2°S and the equator and the longitude is systematically varied by 5° show that the greatest impact of the assimilated data occurs when the observations are taken between 15°W and 30°W. Next, a more systematic technique is presented which allows us to determine optimal points in an objective fashion by applying a least squares regression approach to reconstruct the errors on a dense array of points from the data misfits at any three selected points. The forecast error structure from the Kalman filter is used in a novel way to assess the optimality of mooring locations. From a large sample of triads of points, the optimal mooring locations are found to be along the equator at 35°W, 20°W, and 10°W. Additional experiments are performed to demonstrate the efficacy of the initial and final PIRATA configurations and the added value that can be expected from PIRATA observations beyond existing expendable bathythermograph observations.

1. Introduction

A valuable legacy of the Tropical Ocean-Global Atmosphere (TOGA) program is the extensive ocean observing system developed for the tropical Pacific Ocean. The backbone for this monitoring system is the approximately 70 Tropical Atmosphere-Ocean (TAO) surface moorings [*Hayes et al.*, 1991] that measure surface meteorological variables and upper ocean thermal structure across the full width of the tropical Pacific within 8° of latitude from the equator. In addition, the TAO observations are supplemented with oceanic observations from a variety of other measurement platforms including volunteer observing ships (VOS) expendable bathythermographs (XBT), a network of tide gauge stations, and drifting buoys. At the conclusion of TOGA, it was envisioned [*National Academy of Sciences*, 1994] that the extension of TAO-like arrays to higher latitudes and other tropical oceans would proceed as the need arises and as sufficiently justified in support of long-term monitoring and climate prediction requirements.

Recently, a joint Brazil-France-United States program was proposed to begin the deployment of moored measurement platforms in the tropical Atlantic in order to enhance the existing observational database and subsequent understanding of the processes by which the ocean and atmosphere couple in key regions of the tropical Atlantic Ocean. Empirical studies have suggested that there are strong relationships between tropical Atlantic upper ocean variability, sea surface temperature (SST), ocean-atmosphere coupling, and regional climate variability. During the early 1980s a coordinated set of surface wind, subsurface thermal structure, and subsurface current observations were obtained as part of the United States-France Seasonal Response of the Equatorial Atlantic-Programme Français Océan et Climat dans l'Atlantique Equatorial (SEQUAL-FOCAL) process experiment designed to observe the seasonal response of the tropical Atlantic Ocean to surface forcing. Since that time, however, the observational database for the tropical Atlantic Ocean has been reduced to a few XBT lines and a small collection of tide gauge stations. (For a comparison of Atlantic and Pacific TOGA data coverage, see

Paper number 97JC03206.
0148-0227/98/97JC-03206$09.00

McPhaden et al. 1998). Arguments are now being made that a more comprehensive set of observations, modeling and empirical studies are needed to make progress on understanding the climate variability in regions bordering the tropical Atlantic Ocean.

The proposed PIRATA (Pilot Research Moored Array in the Tropical Atlantic) program will use mooring platforms similar to those of the tropical Pacific TAO [*Hayes et al.*, 1991] array to measure surface fluxes of momentum and heat and the corresponding changes in the upper ocean thermal structure. It is anticipated that the oceanic data from this monitoring array will also be used in a predictive mode for initialization studies of regional coupled climate models. Of particular interest are modes of ocean-atmosphere variability within the tropical Atlantic basin that have significant impacts on the regional climate of the adjacent continents. The zonal mode of variability is similar to the El Niño-Southern Oscillation (ENSO) phenomenon in the Pacific and is situated in the zonal equatorial band [*Zebiak*, 1993]. The warm phase of this equatorial mode is characterized by weak trade winds in the western equatorial Atlantic and an anomalously deep thermocline with higher SST, east along the equator in the Gulf of Guinea. The scenario reverses during the cold phase with an intensification of the equatorial easterlies in the western portion of the basin, and a shoaling thermocline accompanied by anomalously low SST in the east. The meridional mode of variability can be described in terms of a north-south interhemispheric oscillation in SST that has been referred to in the literature as an Atlantic dipole [*Moura and Shukla*, 1981; *Servain*, 1991]. The SST anomalies associated with this oscillating meridional mode (spatially coherent in each hemisphere between 5°N and 25°N and between 5°S and 25°S) are linked with the position and intensity of the Intertropical Convergence Zone (ITCZ). Such displacements in the ITCZ induce precipitation anomalies over northeast Brazil and the African Sahel [*Hastenrath*, 1985].

The purpose of this paper is to describe a series of observing system simulation experiments (OSSEs) to help guide and assess the effectiveness of various configurations of a tropical Atlantic moored array. In the past, most OSSEs which were based on data assimilation were performed after the experimental design had been implemented and the instruments were in the water [e.g., *Miller*, 1990; *Bennett*, 1990]. However, planning for other observational programs has been guided by model studies, and some work has been specifically directed toward the principles of array design.

Perhaps the most important distinction among array design studies is the criterion used to decide which array is best. One obvious criterion is minimization of expected error where simple statistical error estimates can be derived from objective analysis schemes. *Bretherton et al.* [1976] used an objective analysis technique to assess the accuracy of analysis based on the Mid-Ocean Dynamics Experiment (MODE) array. *Bretherton et al.* [1984] used objective analysis to find the most economical sampling strategy for measurement of changes in heat storage in the North Atlantic. In that study, proposed ship tracks and sampling frequencies were varied until the estimated error in the heat storage was within acceptable limits. The error estimates were based on the statistics underlying their objective analysis. They presented a novel Bayesian technique for estimating the sensitivity of their results to their particular choice of error statistics. Along a similar line, *McPhaden et al.* [1984] used ship drift data from the tropical Atlantic Ocean and a stochastic model of drifter trajectories to estimate the accuracy for mapping large-scale seasonal variations of the depth of the 20° isotherm that would result from a number of different deployment strategies. Estimates of errors in the resulting objectively analyzed fields were then formed, based on prior error estimates, and used to help guide the deployment of drifters during FOCAL.

Another possible criterion for optimality of an array is model sensitivity. *Schröter and Wunsch* [1986] proposed sensitivity of model output to changes in forcing, model parameters or measurements of the state of the ocean as a criterion for an optimal array. They used concepts from linear and nonlinear programming to find the type and location of data which would have the greatest impact on the result of a model run. Rather than making assumptions about the moments of the distributions of errors, they imposed limits on data misfits, parameter uncertainties, and the residuals of a set of equations which included the defining equations for the wind-driven circulation in a simple midlatitude gyre. Similarly, *Harrison et al.* [1989] studied the sensitivity of a model of the tropical Pacific Ocean to changes in the forcing fields by comparing runs driven by different wind products. This study, along with that of *Harrison* [1989], was influential in determining the range of latitudes over which wind data would be needed for accurate representation of SST in wind-forced models and helped define where the TAO mooring array should expand. *Harrison* [1989] concluded with recommendations about required coverage of wind measurements, and noted the necessity of including meridional winds in model studies aimed toward simulation of tropical SST.

Yet another criterion one might use is stability of the inversion process within the context of the "inverse methods" calculation. In the process of performing such inversions, researchers wish to obtain the best resolution available without paying too high a price in sensitivity to measurement error. In practice, this amounts to maximizing the amount of independent information available from an array. *Barth* [1992] optimized placement of tomographic array sites in the midlatitude ocean by using simulated annealing and genetic algorithms to choose the locations of sites which give rise to the best conditioned inversion problem. Barth's work was based on statistically optimized interpolation, and involved no explicit dynamical model. *McIntosh* [1987] used a weak constraint variational method to assimilate tide gauge data into a shallow water model of the tides in the Bass Strait. He used the properties of a matrix derived from his method to decide how many independent degrees of freedom were present in the data from the in situ array of 24 tide gauges, and the most efficient way to use a minimum number of measurements. *Bennett* [1990] performed an OSSE of the tropical Pacific in which he examined hydrographic data from ship tracks from a point of view similar to that of McIntosh.

The methods presented in this paper are most like those of *McIntosh* [1987] and *Bennett* [1990]; see also *Bennett* [1992, chap. 6]. Like these studies, the current paper presents a data assimilation scheme which is optimized with respect to a model and a set of prior error estimates. The methods within this paper differ slightly in that a Kalman filter is used instead of weak constraint smoothing schemes. However, all three methods involve similar matrix analyses. The major difference that sets our paper apart from previous work is that it focuses on the errors in the output of the data assimilation system. One reason for this is that since TOPEX/Poseidon data are used, the real sea level signal can be accurately identified. In other

words, well-resolved fields of the exact quantity that should be coming out of the data assimilation system have been observed. The other driving force behind using the data assimilation output error is that the number of moorings in the proposed array will be far too few to fully reproduce the variability of the sea level signal in the tropical Atlantic. Therefore, we must look to the output error to fully optimize the limited number of mooring locations.

In this paper we use a numerical model to determine the influence of different data-gathering strategies; we evaluate the quality of the analyses resulting from specific instrument deployment strategies and make specific recommendations. Our study is distinguished by use of statistically optimized data assimilation, in which the virtues of previous studies are combined and extended, and by the use of TOPEX/Poseidon data for verification. Here, the primary concern is with an optimal sampling strategy for observing low-latitude changes in vertically integrated quantities such as dynamic topography, upper ocean heat content, and sea level which can be considered as proxies for changes to the upper tropical Atlantic Ocean thermal structure and thermocline depth. OSSEs for the surface fluxes and SST are beyond the scope of the present paper, but are envisioned as a necessary next step.

The structure of the paper is as follows: Section 2 contains a description of the data used for assimilation and the forcing fields. The model and Kalman filter scheme are described in section 3 and section 4, respectively. Results of the various data assimilation runs and a description of how the optimal sites for data assimilation are determined are presented in section 5. Section 6 contains information on how data from the PIRATA array will "add value" to other sources of data for the tropical Atlantic. Section 7 summarizes the results of this paper.

2. Data Used for Assimilation

2.1. TOPEX/Poseidon Altimetric Data

The processing of the TOPEX/Poseidon data is similar to that described by *Busalacchi et al.* [1994]. Enhanced TOPEX/Poseidon geophysical data records (GDRs) are first produced by the NASA Ocean Altimeter Pathfinder Project group at Goddard (C. Koblinsky, personal communication, 1996). The 1 Hz data are averaged along track to 6 km resolution and are interpolated in space to the cycle 17 ground track. Each TOPEX and Poseidon track within the area 30°N–30°S and 50°W–10°E is corrected using standard geophysical corrections including the tide model of *Ray et al.* [1994]. For each track, the long-period average is removed to eliminate the error in the geoid. Poseidon data are merged with TOPEX data by assuming a 13.6 cm bias [*Nerem et al.*, 1994]. The anomalies are then collected into 1° × 1° × 10 day bins. The resulting data are optimally interpolated in space onto 1° × 1° grids. These 10 day grids are averaged monthly and then interpolated to the model grid to provide both the verification data and the source for the assimilation data.

The period of the study, October 1992 to December 1995, encompasses enough years to estimate the variability in the tropical Atlantic. A contour map of the standard deviation of the gridded TOPEX/Poseidon surface height anomaly data is presented in Plate 1. Note that there are areas of high variability in the Gulf of Guinea in the far eastern portion of the basin near the equator corresponding to regions where the seasonal cycle has large amplitude and the interannual variability is large. A region of high variability is also found in the region of the North Equatorial Countercurrent ridge system between the equator and 5°N stretching from 45°W out into the Atlantic to about 20°W. These results match those of *Merle* [1980] and *Busalacchi and Picaut* [1983], who found similar patterns in the annual variability when using hydrographic data and model simulations, respectively. The broad regions of low variability (i.e., 1–3 cm) south of the equator are at levels comparable to the errors in the data. The regions of high variability at the equator west of 45°W and the area off the west coast of Africa, near the Cape Verde Islands (20°W–15°W, 10°N–20°N), are artifacts of the tide model which performs poorly in shallow water areas (R. Ray, personal communication, 1997). Intercomparison of all available tide models [*Andersen et al.*, 1995] shows that the Ray et al. tide model performs as well as any other available tide model in these regions.

2.2. Wind Forcing

The wind product of *Servain et al.* [1996] for the period extending from January 1987 to December 1995 is used to provide forcing for the experiments. The data are converted from pseudostress to wind stress by multiplying the pseudostress by the drag coefficient (1.25×10^{-3}) and the air density (1.2 kg/m^3). Since the winds do not cover the model domain completely (i.e., no data south of 20°S), the wind stress climatology of *Hellerman and Rosenstein* [1983] is used to fill the missing areas. First, the two wind products are combined onto a common grid; then this grid is smoothed using a five-point weighted filter. The original Servain et al. and Hellerman and Rosenstein data are replaced north of 17°S and south of 23°S, respectively. The smoothed merged product is left between 17°S and 23°S. The Hellerman and Rosenstein wind stress is reduced by 25% in order to match the drag coefficient chosen for the study. Other investigators have found the Hellerman and Rosenstein wind product to be too strong. *Chelton et al.* [1990] estimated that these winds were 19% stronger than the corresponding winds measured by the Seasat scatterometer. The combined wind stress was interpolated to the model grid. Anomalies are formed using the mean wind stress from October 1992 to December 1995, to match the processing of the TOPEX/Poseidon data.

3. The Model

In the present study, the Kalman filter is incorporated into the linear model of *Cane and Patton* [1984]. The model state variables include normalized height anomaly (h), zonal current (U) and the Kelvin wave amplitude (AK) for each baroclinic mode. The first two baroclinic modes are retained for each of the state variables. The model is set up using roughly realistic coastlines for the tropical Atlantic with the domain extending from 50°W to 10°E and from 30°N to 30°S and a resolution of approximately 1° in longitude and 0.5° latitude. For the experiments the model is forced using the combined Servain et al./Hellerman and Rosenstein monthly wind product interpolated to the model 5 day time step.

The model uses linear shallow water equations on the equatorial β plane subject to the long-wave approximation and finite differences in the horizontal directions to describe the evolution of each baroclinic mode. Linear damping is imposed by Rayleigh friction with decay times of 60 months and 20.5 months for the first and second baroclinic modes, respectively. Kelvin wave speeds correspond to 2.36 and 1.38 m/s, e-folding

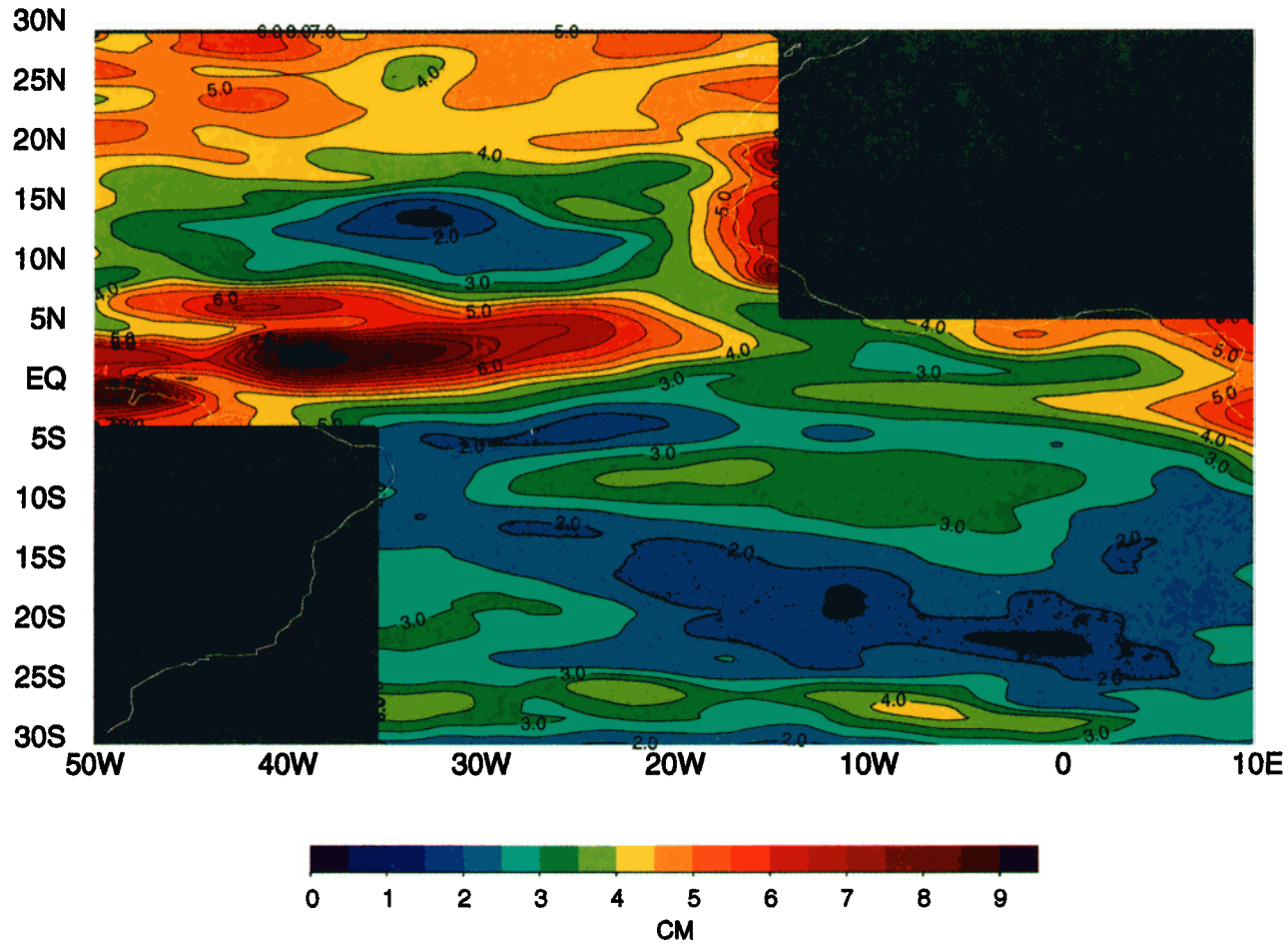

Plate 1. Standard deviation (in centimeters) of the gridded TOPEX/Poseidon sea surface height anomaly product for the period extending from October 1992 until December 1995.

length scales ($L \approx (c/\beta)^{1/2}$) = 2.9° and 2.22°, and the timescales ($T \approx (c\beta)^{-1/2}$) are equal to 1.58 and 2.06 days, for the two baroclinic modes, respectively. Model wave speeds, length scales, and timescales are derived from a density profile created using temperature and salinity profiles from a conductivity-temperature-depth (CTD) cast done on February 10, 1983 at 28°W and the equator. For each experiment the model is spun up over the period January 1987 to October 1992 using the wind stress described in section 2.2. Model runs are carried out for the period extending from October 1992 to December 1995 (i.e., the TOPEX/Poseidon period).

The model geometry is shown in Figure 1. South America and Africa are represented schematically as rectangles removed from the larger domain at the southwest and northeast corners, respectively. The model grid on which the predictions are carried out is shown by the fine mesh.

4. The Kalman Filter

The full Kalman filter [e.g., *Bennett*, 1992; *Ghil and Malanotte-Rizzoli*, 1991] has been applied to this model in the past [*Miller et al.*, 1995], but its demands on computing resources are so severe that its use is restricted to low resolution cases. The reduced-space Kalman filter technique is an efficient tool which allows a high-resolution model to be incorporated into a nearly optimal data assimilation scheme. *Fukumori* [1995] showed that using a reduced-space Kalman filter (his term was quasi-optimal) provides good results for the example of TOPEX data assimilation in the Pacific. *Cohn and Todling* [1996] compared three suboptimal approximations to the Kalman filter for a linearized shallow water model in a periodic channel, including a reduced state space filter, which they

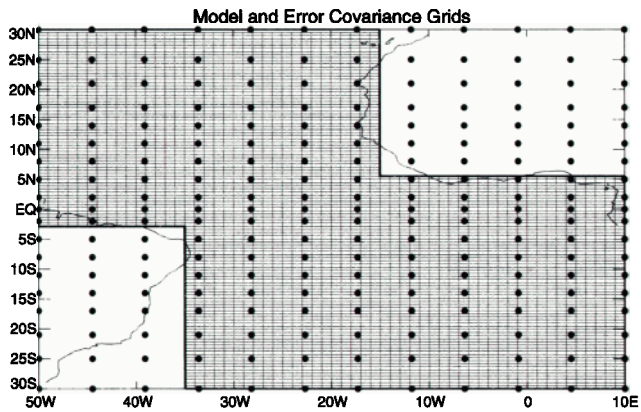

Figure 1. Schematic diagram of the model grid (fine mesh) and error grid (solid circles). Note the higher resolution of the model grid with respect to the error grid.

called a "reduced resolution filter" (RRF), and two others based on representations of the state space in terms of characteristic vectors of the evolution operator and of the covariance matrix. They found that the RRF performed well as long as unstable structures were well resolved, which is not a consideration here. *Cane et al.* [1996] found that the results of a standard Kalman filter could be faithfully reproduced by reduction of the error matrices by the use of multivariate empirical orthogonal functions.

In this study, TOPEX/Poseidon data are assimilated into the linear model using the technique of a reduced-space Kalman filter similar to the method used by *Fukumori* [1995]. This alternative approach is found to reduce the size of the error covariance matrix and thus reduce the number of times the prognostic equations are advanced. (For each row and column of the error covariance matrix the model is advanced to calculate the error covariance). In the current approach, the number of model executions is drastically reduced since the forecast error covariance matrix $\mathbf{P}^f$ and the analysis error covariance matrix $\mathbf{P}^a$ are maintained on a grid which is much coarser than that used in the prognostic model.

A 12 × 19 grid for the error covariance calculations in the Kalman filter is the optimal choice following a series of sensitivity tests. The Kalman filter error grid resolution is 5° longitude, stretched in latitude from 2° at the equator to 5° at the boundaries (matching the meridional resolution of TOGA TAO in Pacific). The density of the error grid is shown as solid circles in Figure 1, and the model resolution is represented as the fine mesh. Although the Kalman filter has reduced resolution with respect to the model grid (the model grid is 61 × 116), interest lies only in the large-scale features of the error field. Therefore this stretched grid is adequate for this data assimilation study.

The initial forecast error covariance ($\mathbf{P}^f$) is approximated using 20 times the system noise covariance $\mathbf{Q}$ [*Chan et al.*, 1996]. In the present case, $\mathbf{Q}$ is estimated by a Monte Carlo technique based on prior assumptions about the errors in the wind-forcing data. $\mathbf{Q}$ is calculated by running the model with a 5 day time step for 150 one-month samples forced by random winds. The random winds are created with the assumed error statistics of the forcing winds, that is, a spatially homogeneous error covariance function of Gaussian form with zonal decorrelation scale of 10° and meridional scale of 4°. Models of covariance statistics of this form have been used successfully in data assimilation studies of the tropical Pacific. While we have less experience with models of the tropical Atlantic, there is strong evidence that error in the wind is the major source of error in model output. As in earlier studies of the tropical Pacific, the error covariance function is used as a device for generating structure for errors due to all sources. The assumption of spatial homogeneity of the system noise is questionable, but experiments with models of the Pacific have shown that relaxation of this assumption has little effect on the analysis or on the calculated estimates of forecast errors. For further discussion of error covariance models for assimilation of data into models of the tropical oceans, see *Miller and Cane* [1996] and references therein.

The assimilation data are the gridded TOPEX/Poseidon anomalies subsampled at model grid locations. Assimilation and update of $\mathbf{P}^f$ takes place on the fifteenth day of each month. When assimilation locations are consistent from month to month, $\mathbf{P}^f$ is stable after 1 year. This result matches the work of *Fukumori et al.* [1993]. In that work, Fukumori et al. applied a specialized algorithm to compute the asymptotic value of $\mathbf{P}^f$ as $t \to \infty$. Application of that algorithm might have saved us some amount of computer resources over the process of simply computing the evolution of $\mathbf{P}^f$ until it stabilized, but we did not feel that the savings realized would have justified the effort involved in implementation.

The observation error variance is assumed to be 9 cm^2 everywhere, with errors at different locations assumed to be uncorrelated. The observation error covariance matrix $\mathbf{R}$ is therefore a multiple of the identity. Although it is understood that altimeters measure sea surface topography and not exactly dynamic height, which is calculated through density profiles and measured by moorings with thermistor chains and salinity sensors, we assume here that the TOPEX/Poseidon data are equivalent to mooring dynamic height. *Picaut et al.* [1995] showed the strong relationship between TOPEX/Poseidon sea level and dynamic height from moorings in the tropical Pacific. By assuming dynamic height and sea level are the same, problems with dynamic height from mooring data such as water mass variability and reference level issues are circumvented.

5. Results

The question of optimal mooring placement is examined by choosing a sample array of mooring sites, assimilating TOPEX/Poseidon data at each site on the sample array, and comparing the result of the assimilation of this restricted data set to the actual TOPEX/Poseidon data. This differs from the usual twin experiment approach, in which a synthetic data set is subsampled according to the plan of the observing system under study, and the sampled information, possibly contaminated with noise, is used to reconstruct the original data set. The advantage of such OSSEs is that the error characteristics of the model are supposedly known better than observations, and the performance of the array can be evaluated in more detail. The difficulty with such studies is the necessity of proving that the synthetic ocean resembles the real ocean. The proposed observing system resulting from the typical class of OSSE must always be interpreted within the context of the limitations of the simulation skill of the synthetic data set. On the other hand, the accuracy and extensive coverage of the TOPEX/Poseidon data set allow us to avoid the question of relevance, which is never fully resolved in studies based on synthetic data sets. The loss of some detail in the ability to assess errors is more than offset by the advantage of knowing that the reference data duplicate observations exactly for our new class of OSSE.

5.1. No Assimilation

The performance of the model without assimilation ("NO ASSIM") is shown in Plate 2. This is the baseline for comparison to the data assimilation results. The top panel of Plate 2 shows the results of the correlation of the model alone versus the TOPEX/Poseidon gridded product. Attention is focused in the region between ±20° of the equator to reduce the influence of spurious boundary effects. This plate shows that the regions of highest correlation correspond to the areas with the highest variability in the TOPEX/Poseidon gridded product (see Plate 1). This is indicated by the high correlation in the region of the South Equatorial trough and Equatorial Countercurrent ridge system in the west (defined by the area 45°W–30°W, 3°S–5°N) and in the Gulf of Guinea, in the east. In the west the model and data are highly correlated, with coefficients sometimes

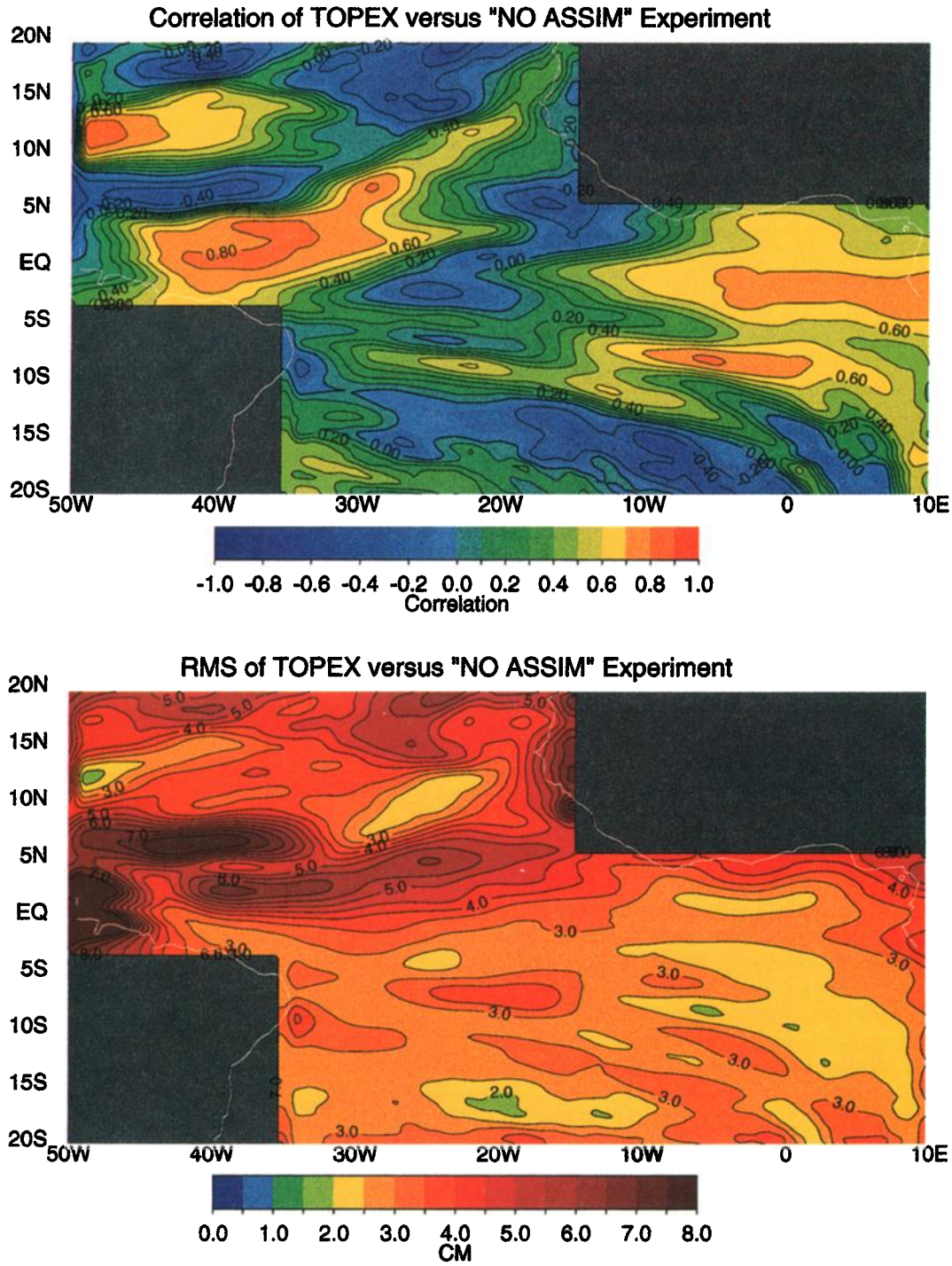

Plate 2. Comparison between model sea level results without assimilation ("NO ASSIM") and the gridded TOPEX/Poseidon product: (top) correlation value in tenths; (bottom) RMS difference in centimeters.

exceeding the +0.8 correlation level. The correlation drops west of 45°W due to deficiencies in the tide model in the TOPEX/Poseidon data. In summary, the model does a good job of reproducing the large temporal signal in the western and eastern parts of the equatorial waveguide, but does poorly in the central equatorial part of the basin between 30°W and 15°W.

Although the relatively high correlations show that the model does a good job in reproducing the temporal signal, the relatively large RMS differences between the model output and the data indicate that the amplitude of the signal is poorly represented by the model. Plate 2 (bottom panel) shows the RMS difference between the sea level from the raw model results ("NO ASSIM") and the gridded TOPEX/Poseidon product. The plate shows that there are large differences in the magnitude of the signal in the region of the Equatorial Countercurrent ridge. These differences often exceed 5 cm RMS. In the eastern part of the basin at the equator and throughout the

waveguide, the RMS differences usually exceed 3 cm. In general, the field of deviations of the model output from the TOPEX/Poseidon data in the northern hemisphere has higher RMS amplitude than that found in the southern hemisphere. This matches the fact that the TOPEX/Poseidon data exhibit higher variability in the north than in the south. Varying the drag coefficient by ±20% and using other wind products (such as the special sensor microwave imager (SSM/I) wind product of *Atlas et al.* [1991]) does not eliminate these basic patterns of high RMS errors. The map of expected RMS errors generated from the evolution of the model error covariance matrix (not shown) indicates that the predicted RMS errors are of similar magnitude to those observed, with similar pattern. The major discrepancy is that the predicted map of RMS errors shows less difference between the RMS errors in the northern and southern hemispheres. In this case the prior estimate of the RMS errors is an overestimate through most of the region south of the equator.

5.2. High-Density Data Assimilation

A high-density assimilation experiment is performed to assess the impact of data assimilation using a very dense observational network consistent with the resolution of the error grid, and also to provide a basis for the statistical experiments which are described in section 5.3.2. Plate 3 shows the results of an experiment in which TOPEX/Poseidon data were assimilated at the 66 locations marked with solid circles for the period extending from October 1992 to December 1995. These data assimilation points are located at 8°N, 5°N, 2°N, 0°, 2°S, 5°S, 8°S in latitude and every 5° in longitude except when these points fall over land. These locations correspond to the same latitudes as the TOGA TAO grid in the Pacific. The zonal resolution corresponds to the resolution of the error grid. This run is referred to as the "HIGH DEN" data assimilation run. It is recognized that such a dense array may not be logistically feasible for a pilot monitoring array. Nonetheless, this is a best case scenario with which to assess the effectiveness of a less dense mooring configuration.

The top panel of Plate 3 shows the correlation between the sea level from the "HIGH DEN" data assimilation run and the TOPEX/Poseidon gridded product. The results show a uniformly high correlation throughout the equatorial waveguide, with values exceeding +0.9 in the region between 5°N and 5°S. Exceptions occur at the western boundary and the eastern boundary and off the northeast coast of South America. The low values at the western and eastern boundaries can be attributed to the fact that the model treats the boundaries in a very simple manner. The negative correlation at the equatorial western boundary may also be associated with the quality of the TOPEX/Poseidon data affected by the poor performance of the tide model in the shallow seas off the Amazon basin. The narrow region of low correlation near 3°S stretching from about 32°W to about 22°W is a region of low signal amplitude (see Plate 1), where the magnitude of the TOPEX/Poseidon sea level variability falls below the observational error. The corresponding region in the lower panel shows that the RMS difference between the output of the assimilation run and the data was less than the observation error in this region.

The bottom panel of Plate 3 shows a map of the RMS differences between sea level from the "HIGH DEN" data assimilation run and the TOPEX/Poseidon gridded product. Between 5°N and 5°S the RMS difference generally is less than 2 cm. It is interesting to note that the observational error for the Kalman filter process was assigned to be 3 cm. If the model results were simply replaced with TOPEX/Poseidon observations, the RMS error would be greater than or equal to 3 cm (the prescribed observational error). The fact that the RMS is less than 2 cm shows that the model is adjusting the dynamics of the system to fit the observations. On the other hand, errors are still high in the region of the Countercurrent ridge system (45°W–30°W, 3°N–7°N). Some of the dynamics of the system cannot be adequately handled given the resolution of the data assimilation scheme or the restrictive assumptions in the error model. Amplitude problems persist in the far eastern and western part of the basin due to failures in the model boundary scheme and problems with the TOPEX/Poseidon data near the Amazon basin.

The next question to be addressed is, What are the consequences of decreasing the number of moorings. To address this question, a series of experiments was performed in which data from a successively smaller number of synthetic mooring lines were assimilated. These mooring lines were equally spaced in the zonal direction. Each mooring line consisted of seven moorings at a fixed longitude. For example, a data assimilation experiment was completed where the basin was divided into 12 equal portions (i.e., the "HIGH DEN" run: 5° zonal resolution and 11 moorings lines). Then the basin was divided into 11 equal parts (i.e., 5.45° zonal resolution) by placing simulated mooring lines at 45°W, 39°W, 34°W, 28°W, 23°W, 17°W, 12°W, 6°W, 1°E, and 5°E. This process continued until only one line was centered at 20°W, the center of the model basin.

For each of these runs, correlation coefficients of the model run with the TOPEX/Poseidon data and RMS differences between the model output and the TOPEX/Poseidon data are calculated. Area-averaged correlation for each of these experiments versus the TOPEX/Poseidon gridded product are presented in the top panel of Figure 2 along with the zero–mooring line case (i.e., the "NO ASSIM" run). The three lines in this figure correspond to the average of all the correlation coefficients for the area encompassing all longitudes between ±3°, ±6° and ±9° for the dotted, dashed and solid curves, respectively. As expected, assimilating more data improves the correlation to a certain extent. For the ±3° area average this improvement rises steeply at first from +0.45 for no lines to +0.63 for one line. This trend increases steadily until the improvement levels off at between five (correlation, +0.79) and nine (correlation, +0.83) mooring lines. The distance between mooring lines corresponds to roughly 10° to 6° separation, which matches one half the expected decorrelation scale at the equator in the Pacific [*Kessler et al.*, 1996; *Meyers et al.*, 1979] as determined from TOGA TAO and XBT observations, respectively.

The bottom panel in Figure 2 is similar to the top panel, except that the quantity is the area-averaged RMS difference between the model runs and the TOPEX/Poseidon product. As the number of mooring lines increases, the RMS difference drops. For the "NO ASSIM" case, the area-averaged RMS for the area encompassed by ±3° is 3.9 cm. This value decreased to a minimum of 2.2 cm for the 11-mooring case (i.e., the "HIGH DEN" run).

5.3. Assimilation of Data From Groups of Three Moorings

The planning process for the PIRATA program envisioned resource and logistical limitations permitting an initial deployment of only five moorings during the first year. Two of these moorings were earmarked to measure the interhemispheric

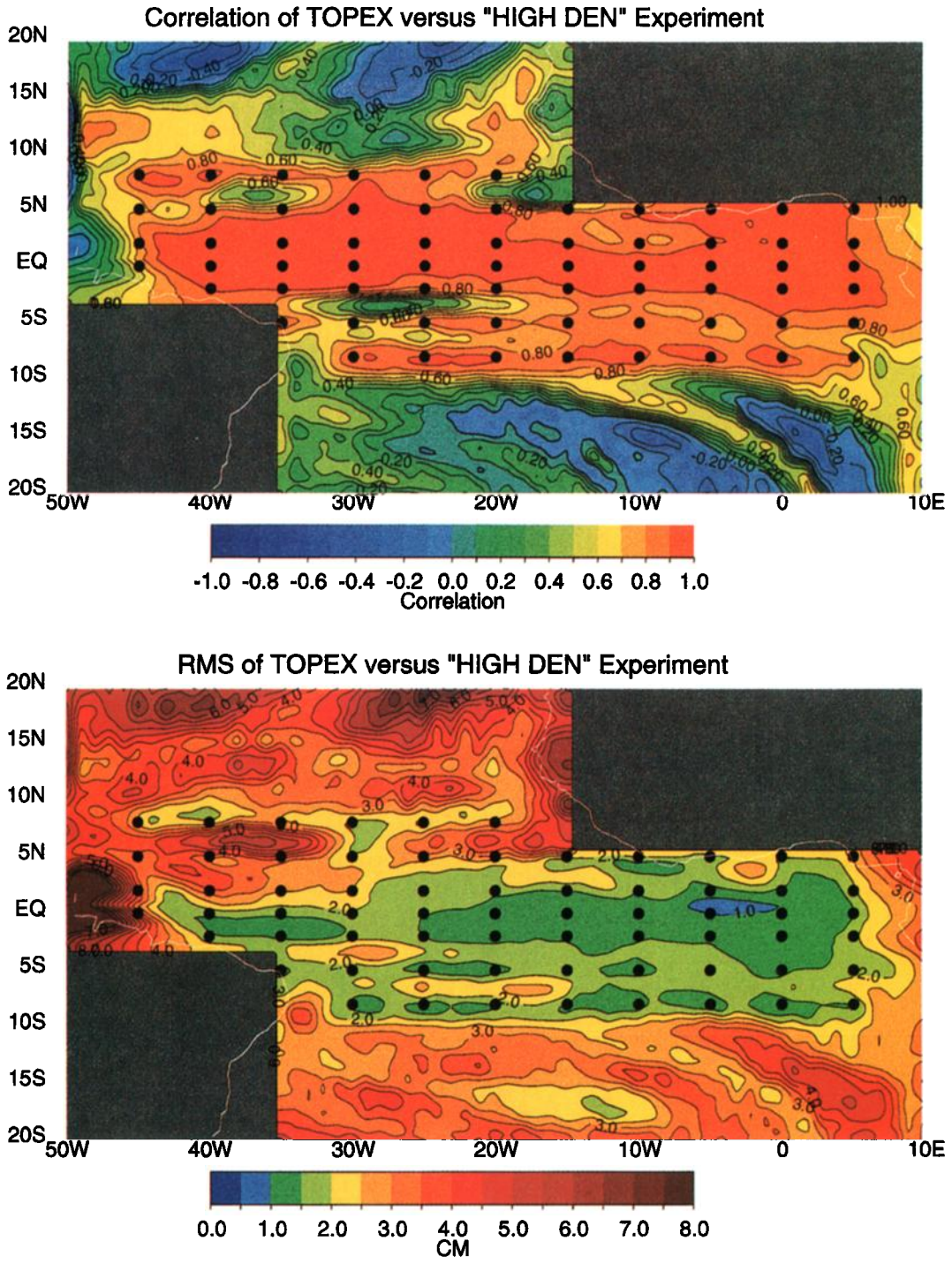

Plate 3. Model sea level from the "HIGH DEN" data assimilation run versus the gridded TOPEX/Poseidon product: (top) correlation coefficient; (bottom) RMS difference in centimeters. Solid circles represent the 66 data assimilation locations.

oscillations in SST well away from the equatorial waveguide. This paper seeks to determine the optimal configuration of the three remaining moorings intended to measure fluctuations in the subsurface thermal field of the equatorial Atlantic Ocean. It is expected that this initial deployment would evolve into a pilot array of 14 moorings in year 3 of the PIRATA plan. Initial emphasis would be placed on monitoring the equatorial waveguide with subsequent enhancements to include additional sections to observe the off-equatorial meridional dipole variability.

5.3.1. Three-mooring deployments along a meridian. A suite of experiments is performed beginning with a triad of simulated moorings in the equatorial Atlantic. This allows an examination of the question of optimal placement by choosing groups of three moorings at a time, assimilating data at each group of three, and comparing the result of the assimilation of

this restricted data set to the actual TOPEX/Poseidon data. In this suite of experiments, three moorings are placed in a north-south oriented line spanning the equatorial waveguide, one on the equator and the others at ±2°. The assumption was made from the beginning that resolving the equatorial meridional structure was desirable, and that logistically it would be easier to deploy a mooring along a single line of longitude. The results of placing a group of three moorings at successively different longitudes from 45°W to 5°E are shown in Table 1. The experiments with no assimilation ("NO ASSIM") and the high-density experiment ("HIGH DEN") are included for comparison. While the differences in correlations and RMS errors between any two choices of longitudes are certainly not significant, the consistent trends are unmistakable. Note that in the three-mooring experiments, the average correlation increases eastward from 45°W to 20°W (or 15°W depending on the domain over which the average is calculated), and then decreases as the array is moved toward the eastern boundary. The correlation for the ±3° area average rises from +0.47 for assimilation at 45°W to +0.66 at 20°W, while RMS error drops from 3.7 cm to 3.4 cm. The correlation drops to +0.50 as the mooring lines are moved to the east to 5°E.

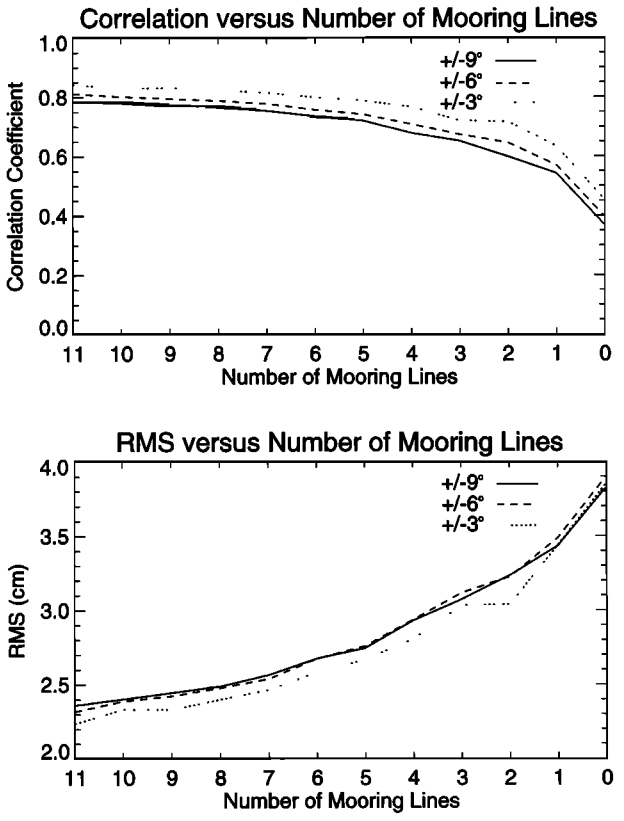

Figure 2. Area-averaged statistics as a function of data density in assimilation experiments. The abscissa corresponds to the number of equidistant meridional mooring lines assimilated across the model basin, decreasing to the right. The ordinate corresponds to (top) the average correlation and (bottom) RMS for the area between the latitudes indicated by the dotted line (±3°), the dashed line (±6°), and the solid line (±9°). The moorings were assimilated along meridians at ±8°, ±5°, ±2° and 0°. The zonal spacing between mooring lines and model boundaries corresponds to 5° for 11 lines ("HIGH DEN" run), 5.45° for 10, 6° for 9, 6.67° for 8, 7.5° for 7, 8.57° for 6, 10° for 5, 12° for 4, 15° for 3, 20° for 2, and 30° for 1. The zero line case corresponds to the results for the "NO ASSIM" run.

Table 1. Area-averaged Correlation and RMS Values for the Gridded TOPEX/Poseidon Product Versus Each of the Suite of Experiments in Which Three Moorings (Located at 2°N, 0°, 2°S) Are Assimilated Along Indicated Longitudes.

Longitude of Assimilation	Correlation (×100)			RMS, cm		
	±10°	±6°	±3°	±10°	±6°	±3°
45°W*	38.	42.	47.	3.7	3.8	3.7
40°W*	40.	45.	53.	3.7	3.8	3.6
35°W*	40.	46.	57.	3.7	3.7	3.5
30°W*	42.	50.	63.	3.6†	3.6†	3.3†
25°W*	43.	51.	66.†	3.6†	3.6†	3.3†
20°W*	44.	52.†	66.†	3.6†	3.6†	3.4
15°W*	45.†	52.†	64.	3.6†	3.6†	3.4
10°W*	43.	50.	58.	3.6†	3.7	3.5
5°W*	42.	47.	54.	3.6†	3.7	3.6
0°W*	42.	47.	54.	3.6†	3.7	3.6
5°E*	41.	46.	50.	3.6†	3.7	3.7
NO ASSIM	36.	40.	45.	3.8	3.9	3.9
HIGH DEN	77.	81.	84.	2.4	2.3	2.2

The results of the model without data assimilation ("NO ASSIM") and the "HIGH DEN" data assimilation runs are included for comparison.

* Assimilation at 2°N, equator, and 2°S.

† Highest correlation/lowest RMS in column (excluding "NO ASSIM" and "HIGH DEN" runs).

Plate 4 shows the comparison of the TOPEX/Poseidon gridded product with the results of the best run in which data are assimilated at 2°N, 0°, 2°S and 20°W. (This run is referred to as the "3@20W" run). The top panel of Plate 4 indicates that correlation is high in the region where the data are assimilated. However, it is interesting to note that the high correlation (areas where correlation exceeds +0.8) extends downstream to the west within the Rossby waveguide. The Kelvin wave mechanism carries information to the east, as is evident by the large correlation directly to the east of the moorings and by the high correlation between 20°W and 5°W. Compare the correlations of the "NO ASSIM" run (Plate 2) with those of the "3@20W" run. The entire region of poor correlation (i.e., correlation less than +0.2 in Plate 2) is eliminated by assimilating just three points at 20°W. The area-averaged correlation for the 3°N to 3°S region improves dramatically from +0.45 for "NO ASSIM" run to +0.66 for "3@20W" run (see Table 1). In other words, the temporal signal within 3° of the equator is correctly adjusted by assimilating at these three points.

A map of the RMS differences between the "3@20W" run and the TOPEX/Poseidon gridded product is shown in the bottom panel of Plate 4. These results imply that the RMS error is greatly reduced at the assimilation site, but it cannot be concluded that significant amplitude information is passed far downstream. Areas of relatively high RMS error (i.e., RMS greater than 2 cm) exist throughout the equatorial band encompassed by ±3° except near the assimilation site. To summarize, assimilation of three simulated moorings at 20°W improves the temporal signal for most longitudes along the equator, but the amplitude of the signal is not significantly improved (3.9 cm for "NO ASSIM"; 3.4 cm for "3@20W") by assimilating at these limited number of sites.

5.3.2. Three-mooring deployment anywhere in the tropical Atlantic. In the previous section an OSSE was performed by assuming that only three moorings would be deployed and that

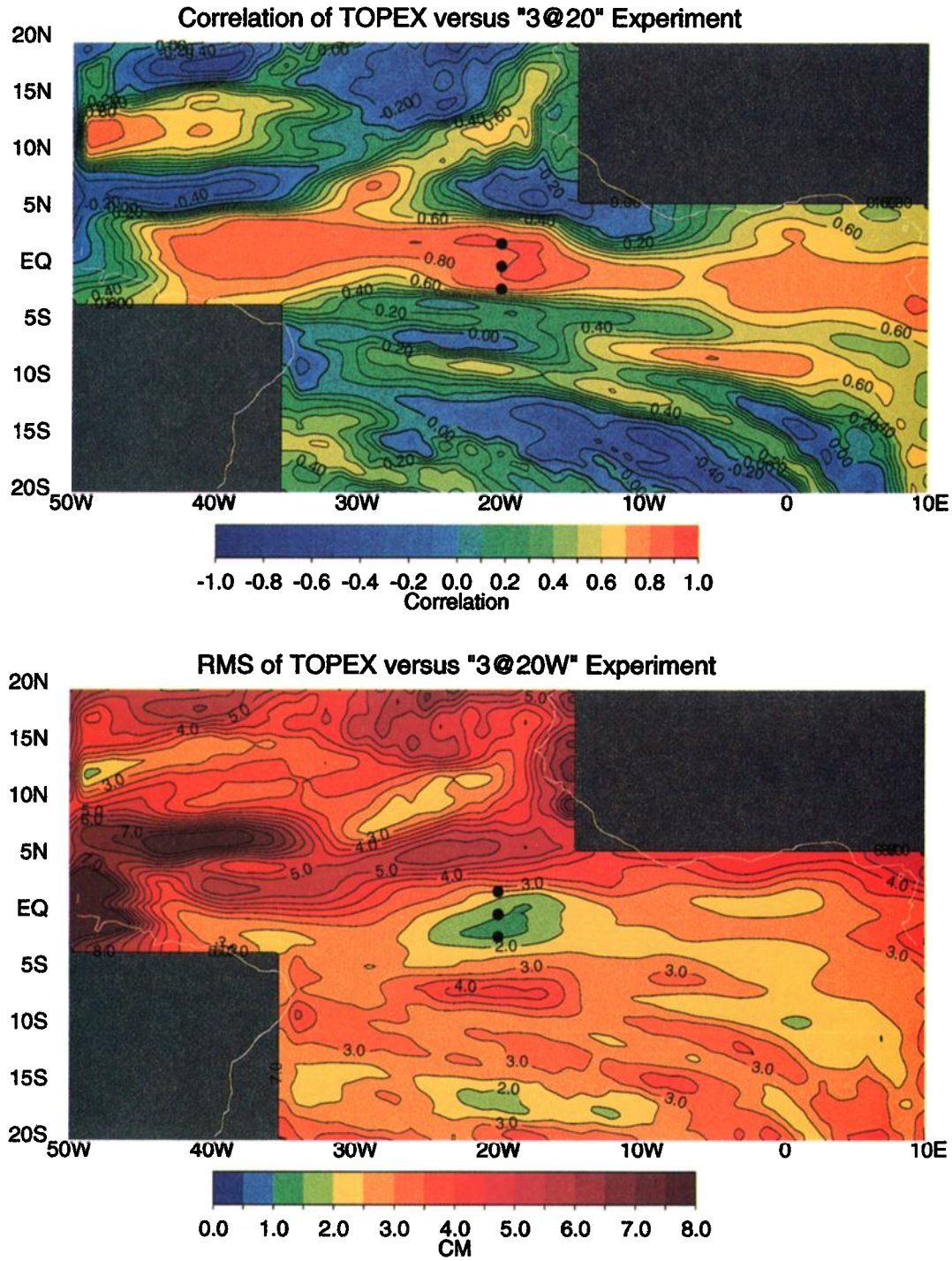

Plate 4. Model sea level from the "3@20W" run (2°N, 0°, 2°S, and 20°W) versus the gridded TOPEX/ Poseidon product: (top) correlation value in tenths; (bottom) RMS difference in centimeters. Solid circles represent data assimilation locations.

all moorings would be along a single meridian. Because of the limited number of experiments (11), each of the array configurations could be tested individually. The OSSE proceeded by cycling through 11 successive placements of the mooring lines at 5° zonal spacings, and the utility of each set was judged through the use of average correlation and RMS values. Now the requirement that all three moorings lie along a single meridian is discarded, and any three points within the set defined by the "HIGH DEN" array can be chosen as the optimal set. Because of the large number of possible combinations (13,244), a theoretical approach to choosing the optimal mooring locations must be sought.

In earlier work, *Bennett* [1990, 1992] addressed a similar problem in the tropical Pacific by attempting to determine the location of optimal ship tracks. Although his method employed a different type of data assimilation technique (namely,

smoother-representer), we attempted to duplicate his methodology to determine the optimal points. First, an attempt was made to eliminate superfluous points by using singular value decomposition (SVD) analysis, and then a perspective plot was created which showed each observation's contribution to each eigenvector [see *Bennett*, 1990, Figure 7]. This technique did not implicate any observation as being superfluous.

For the current problem the SVD analysis showed that there were 17 independent degrees of freedom in the data. This means that only the first 17 of the eigenvalues of the matrix $\mathbf{HP}^f\mathbf{H}^T$ ($\mathbf{H}$ is the linear transformation which relates the state variables to the observed sea level at the 66 "HIGH DEN" sites; see also notation of *Cane et al.* [1996]) indicated model errors with greater variance than the assumed observation noise variance of 9 cm^2. In this case, however, there is no systematic way of choosing 17 observations and discarding the rest from the dense array. *McIntosh* [1987] tried a similar strategy and found that there were 4 independent degrees of freedom out of an array of 24 tide gauges. He chose an array of four instruments by "determining which 4 × 4 matrix comprising four rows and the corresponding columns" of the matrix in question had the smallest condition number. He concluded that there is no guarantee that any four-element subset of the original array exists which contains all of the information in the full array. Perhaps the particular situation addressed in this paper would have allowed a reasonably clear choice of a 17-element subset of the original dense observation array, but since an array of only three measurements is sought, it is unlikely that this exact procedure could distinguish the best three among all of the possibilities. Therefore a linear least squares regression technique was decided upon to choose the best three locations from the candidates which make up the original high-density array ("HIGH DEN").

The assumption is made that part of the initial PIRATA deployment will consist of three moorings to be deployed at any of the 44 points within ±5° and between 40°W and 0°W at the locations defined by the "HIGH DEN" array (known as inner points). In order to choose which three moorings from the original simulation are the best choice, a multivariate least squares regression based on the assumption that the innovations (i.e., the components of the vector of observed minus modeled sea level) are linear functions of the error at any three selected points. From the fundamental Kalman filter equations, innovations are random variables with mean zero and a covariance of $\mathbf{HP}^f\mathbf{H}^T + \mathbf{R}$. Therefore $\mathbf{HP}^f\mathbf{H}^T$ contains all the information necessary for regression analysis. Specifically, the method is performed in the following way: The forecast error covariances for all 66 points that make up the "HIGH DEN" array are saved into a 66 × 66 square matrix. This error matrix is subsampled for three points out of the inner 44 points, leaving a 3 × 3 submatrix of $\mathbf{HP}^f\mathbf{H}^T$. Using this submatrix, we systematically predicted the values for each of the remaining 63 moorings (i.e., 66 total moorings minus three points) using ordinary multivariate least squares liner regression. The coefficients which define the linear functions are saved, and an average value for the explained variance is calculated from all of the 63 realizations. Next, this process is repeated for all possible combinations of three points out of the 44 inner points. This leaves a total of 13,244 values of average explained variance. Once this process is completed, the three points which correspond to the highest average explained variance are the points which are considered to be optimal. The validity of this process depends on the assumption that $\mathbf{HP}^f\mathbf{H}^T$ is a good estimate of the covariance of the difference between the model estimate of the observed quantities and the observed quantities themselves. With this relatively short time series, detailed agreement between the variances of the differences between the model output, the TOPEX observations and the prior estimates of those variances would not be expected, even if all of our statistical models were perfect. In this case the comparisons between the bottom panels of Plates 2 and 3 and the maps generated from the corresponding prior estimates (not shown) indicate a close resemblance.

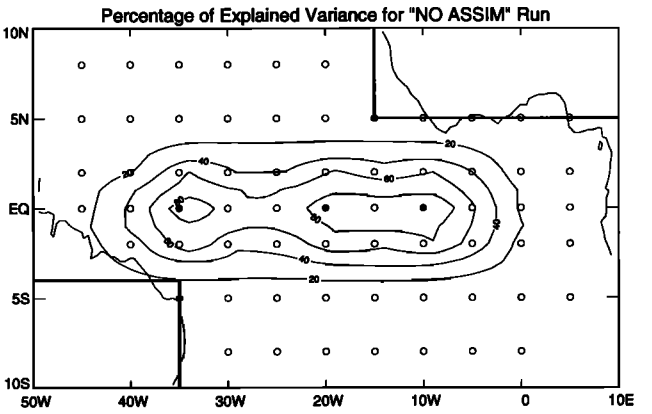

Figure 3. Contour plot of explained variance for "NO ASSIM" run. At each of the locations marked by open circles, the percentage of explained variance is defined by least squares regression using the points at 35°W, 20°W, 10°W, and the equator as a basis (solid circles). Quantity contoured is 100 times the percentage of explained variance. The error covariance is the equilibrium forecast error covariance for the case with no assimilation.

After systematically applying this linear least squares regression approach to all 13,422 combinations of three points out of the inner 44 locations, we found that the best choice is to place three moorings on the equator at 35°W, 20°W, and 10°W. The result that these points are located on the equator is not surprising because of the efficiency of propagation of information by equatorial waves. The zonal spacing between points is also not surprising, since this configuration effectively "spans" the equatorial waveguide. The method of determining optimal points serves the current purpose since the basin-scale features are precisely the phenomena of greatest interest in this study. The optimal points are located at 35°W, 20°W and 10°W at the equator, are chosen in an objective manner, and correspond to the locations that have the highest area-averaged explained variance.

A map of the percentages of the error variance explained by moorings at these three locations is shown in Figure 3. This figure shows that the errors at the three optimal mooring sites explain more than 60% of the error variance over much of the waveguide from 40°W to 5°W, within 2° of the equator. The efficiency of transmission of information along the waveguide is indicated by the zonally oriented isolines of explained variance. Performance outside the waveguide is relatively poor due to the decreased efficiency of transmission of information by equatorial waves, with the consequence of increasing the relative importance of local forcing and hence decreasing of decorrelation scales away from the equator.

A data assimilation experiment is carried out using the three optimal points determined by the linear regression process

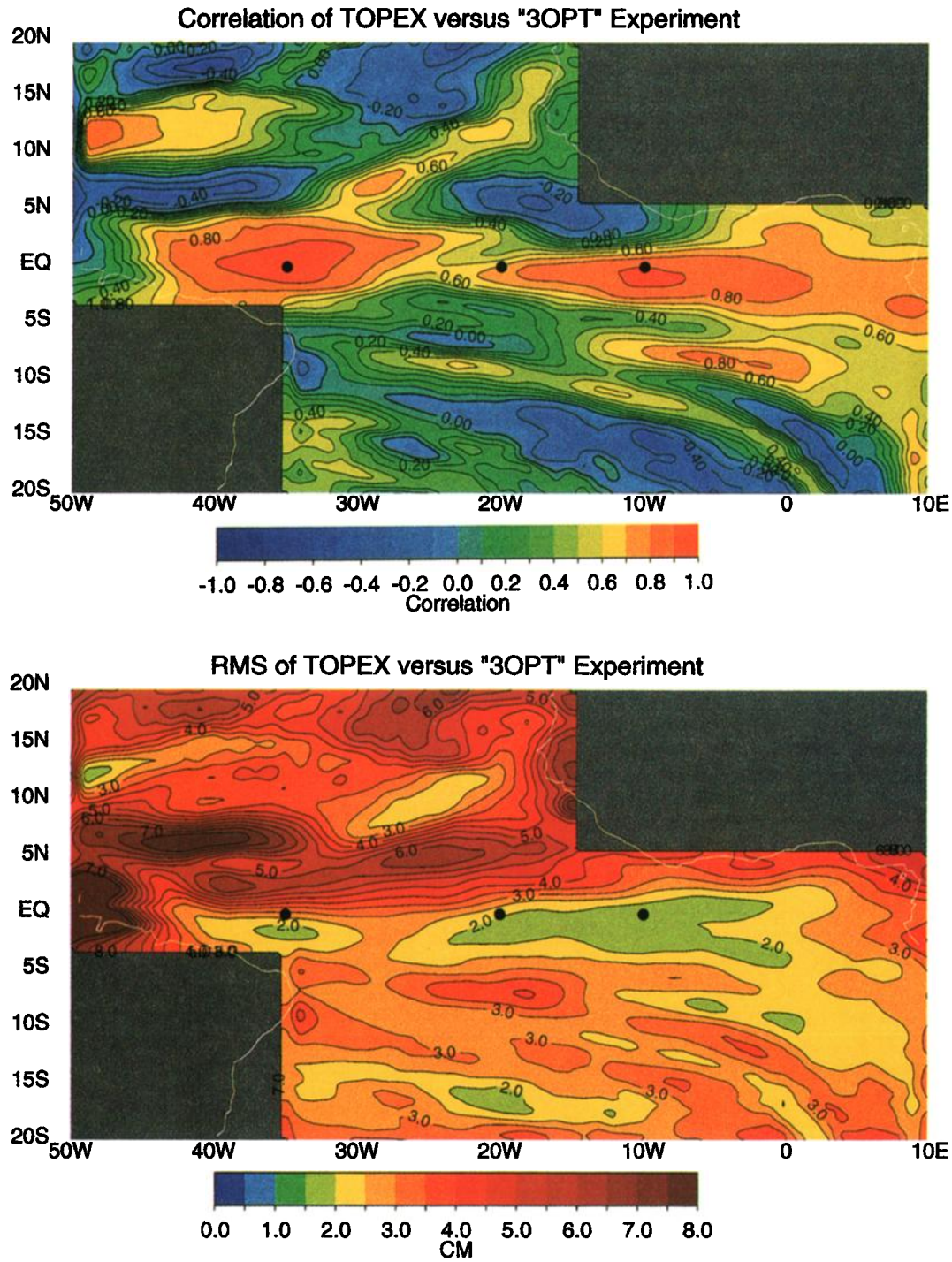

Plate 5. Model sea level results from the "3OPT" run (35°W, 20°W, 10°W, and equator) versus the gridded TOPEX/Poseidon product: (top) correlation value in tenths; (bottom) RMS difference in centimeters. Solid circles represent data assimilation locations. Area averaged correlations correspond to +0.67, +0.53, +0.45 and RMS error averages equal to 3.3 cm, 3.6 cm and 3.6 cm for the areas between ±3°, ±6°, and ±10°, respectively. (These numbers can be directly compared to those found in Table 1.)

described above. This run is designated as the "3OPT" run. The results of the correlation of sea level from this data assimilation run with the TOPEX/Poseidon gridded product is presented in Plate 5. This experiment indicates an overall increased correlation in the equatorial waveguide with respect to the "NO ASSIM" run (see Plate 2). Generally, correlations exceed +0.7 within ±2° of the equator except in the far western part of the basin, where the tide model is known to have problems. Values exceed +0.9 near the assimilation points at 35°W and 10°W. Even though correlations are higher near the 20°W assimilation point than in the "NO ASSIM" run, the meridional influence of assimilation is not as extensive as at the other two points. The off-equatorial signal does not appear to be modified to a great extent, since correlations drop below

+0.4 at points 3° from the equator. When compared with the results of the "3@20W" run (Plate 4), assimilation of two points off the equator (i.e., 2°N and 2°S) causes a clear modification of the signal contained in the Rossby waveguide (i.e., poleward of the range of influence of the Kelvin waves) at 20°W and to the west. In the case of the "3OPT" run, no data are assimilated off the equator, and there is little evidence of Rossby wave modification. Therefore a mooring array would need off-equatorial moorings to establish the Rossby wave signal.

The high correlation along the equatorial waveguide indicates improvement in the temporal signal using the "3OPT" assimilation points. However, unlike the "3@20W" run, the bottom panel of Plate 5 shows a reduction in the RMS difference between the TOPEX data and the reconstructed sea level to the west and east away from the data assimilation points. The RMS error is not only smaller at the assimilation locations (as was the case for the "3@20W" run), but it is smaller throughout the waveguide (not true for the "3@20W" run) since the three equatorial observations and the corresponding decorrelation "box" surrounding each point cover more of the equatorial band. The RMS differences between the reconstructed and observed surface height fields are smaller than 2.5 cm along the equator for most of the basin, excluding only the areas which are tainted by boundary effects. Thus this data assimilation run validates the work done with the linear regression optimization. The "3OPT" run is better than the "3@20W" for the purpose of this paper due to the improvement in the RMS deviation of the reconstructed field versus the observed surface height anomaly field.

5.4. Assimilation of Data at Proposed PIRATA Mooring Sites

With the results of the "3OPT" run in hand, the PIRATA Science Working Group adopted three equatorial sites (30°W, 20°W and 10°W) for part of their proposed initial configuration of the mooring array. Additional off-equatorial sites at 15°N, 30°W and 8°S, 10°W complete the initial array configuration. The off-equatorial points were added in the PIRATA plan to observe the north-south interhemispheric oscillation in SST, which is so important for Brazilian Nordeste rainfall prediction, for example. The experiment which assimilates the sea level data from the five locations of the initial PIRATA array is known as the "PIRATA BEG" run.

Maps of the correlation and RMS error fields from the "PIRATA BEG" run are shown in Plate 6. Close examination of this plate and the corresponding plate for the "3OPT" run (Plate 5) shows little difference between the two. However, some differences are worth noting. The effect of the additional points at 15°N, 30°W and at 8°S, 10°W mostly shows up in the two RMS error plots. Regions of reduced RMS error (less than 2 cm) stretch from the southern point in the PIRATA run, whereas this feature is missing from the "3OPT" run. The differences between the two runs near the northern assimilation point are less evident. This may be due to the weak signal amplitude in the TOPEX/Poseidon data in this region (see Plate 1).

The final proposed configuration of the PIRATA array will include 14 moorings. In addition to the equatorial moorings and the mooring at 15°N, the mooring line along 30°W will be filled in with moorings at 2°S, 2°N, 5°N and 10°N. The mooring line along 10°W will also be expanded from the initial moorings at the equator and at 8°S. Moorings will be placed at 2°S, 2°N, and 5°S. Additional moorings will be located along the equator at 35°W and 0°W to help define the temporal and spatial signals and to measure large-scale ocean dynamics. The point at 35°W corresponds to the westernmost optimal point from the analysis described in section 5.3.2 and also corresponds to a maximum in the interannual variability in the wind stress (J. Servain, personal communication, 1996). The locations of the proposed ending PIRATA array moorings sites are defined by the solid circles in Plate 7.

The results of data assimilation of TOPEX/Poseidon data at the final PIRATA sites are presented in Plate 7. This run is known as the "PIRATA FIN" run. The top panel shows the correlation with the TOPEX/Poseidon data. The improvement over the "PIRATA BEG" (Plate 6) is dramatic. The areas of highest correlation (i.e., areas that exceed +0.90) expand in both the zonal and meridional directions with respect to the "PIRATA BEG" run due to the increased number of data assimilation sites. The average correlation for the region encompassed by ±3° increases from +0.69 to +0.76. Throughout the entire waveguide the correlation exceeds +0.70 (with the continuing exception of the area west of 45°W due to known problems with the tide models). Assimilation at the northern moorings along the 30°W line contributes slightly to increasing the correlation to the west of the mooring sites where Rossby waves carry the temporal signal. The same is true for the southern moorings at 10°W.

The RMS differences between TOPEX/Poseidon and the "PIRATA FIN" run are presented in the bottom panel of Plate 7. Increasing the number of data assimilation points in the "PIRATA FIN" run results in smaller RMS deviations from the observed field than those found in the "PIRATA BEG" run. In particular, the average RMS deviation from observation drops from 3.3 cm to 2.9 cm for the region encompassed by ±3°. Plate 7 shows that the entire area along the equator between the east and west mooring sites has RMS error less than 2 cm. Adding assimilated data at 5°N, 10°N, and 30°W improves the RMS to the west somewhat, but the effect of assimilating data at these mooring locations does little to reduce the high RMS error found between 0°N and 10°N and between 50°W and 35°W.

6. Estimates of Added Value of Data From the Proposed Moored Array Over Existing Observations

Any proper approach to performing an OSSE for a proposed monitoring system should take into account relevant existing observations. In order to assess the contribution of the PIRATA array over and above that obtained from other data sources, two additional data assimilation runs are compared. In the first run, the influence of XBT observations is simulated by assimilating TOPEX/Poseidon data at actual XBT locations. Actual XBT locations are derived using the quality-controlled XBT data set from the Coupled Model Project at the National Centers for Environmental Prediction. The number of actual XBT locations used for assimilation varies from a high of 106 points in January 1995 to a low of 18 points in June 1993. In contrast to other experiments discussed so far in this paper, the TOPEX/Poseidon data are subsampled for the middle 10 day period for a month. This is done to add a higher-frequency signal to the solution and is representative of the time it takes a ship of opportunity to transit the tropical Atlantic Ocean. The XBT locations that are included fall between the tenth and twentieth of each month. This experiment is known as the

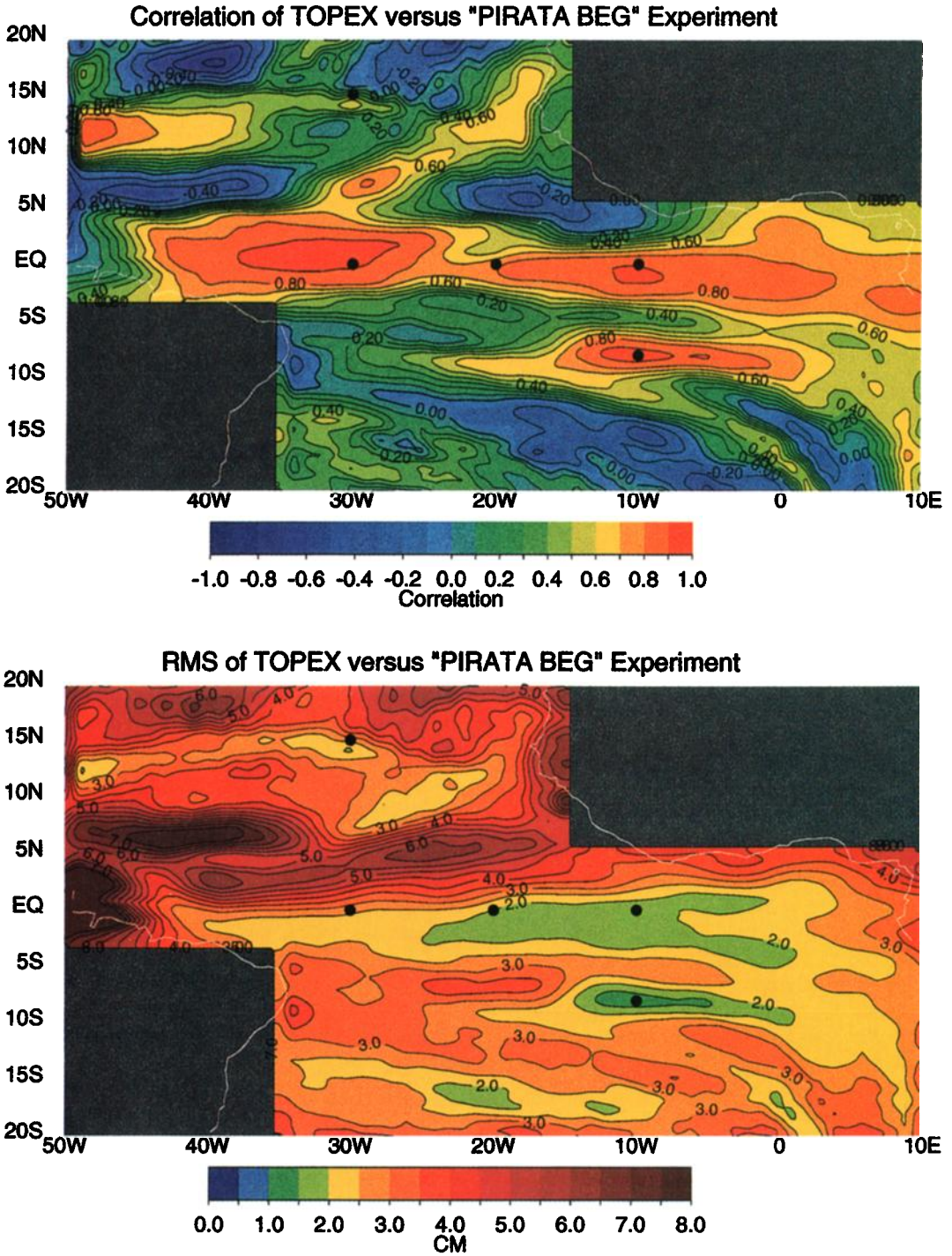

Plate 6. Model sea level results for assimilation at the initial proposed PIRATA array ("PIRATA BEG") versus the gridded TOPEX/Poseidon product: (top) correlation value in tenths; (bottom) RMS difference in centimeters. Solid circles represent data assimilation locations. Area-averaged correlations correspond to +0.69, +0.54, and +0.48, and RMS error averages equal 3.3 cm, 3.6 cm and 3.5 cm for the areas encompassed by ±3°, ±6°, and ±10°, respectively. (These numbers can be directly compared to those found in Table 1.)

"XBT" run. Another difference with these runs is that the forecast error covariance is left to evolve beyond the first year throughout the run. This is necessary because XBT coverage is episodic in nature. In other words, a ship may travel through a region and the data assimilation scheme adjusts the dynamics at that location. However, as soon as the ship leaves the area, the error grows relatively rapidly until it reaches the level of error that would result from having no observations in that region. In this way, the model can be said to "forget" the effect of data assimilation. In the second run, TOPEX/Poseidon data subsampled at XBT locations are assimilated as in the "XBT" run, and, in addition to these data, the monthly mean TOPEX/Poseidon data are assimilated at the locations corresponding to the final PIRATA array locations, as in the "PIRATA FIN"

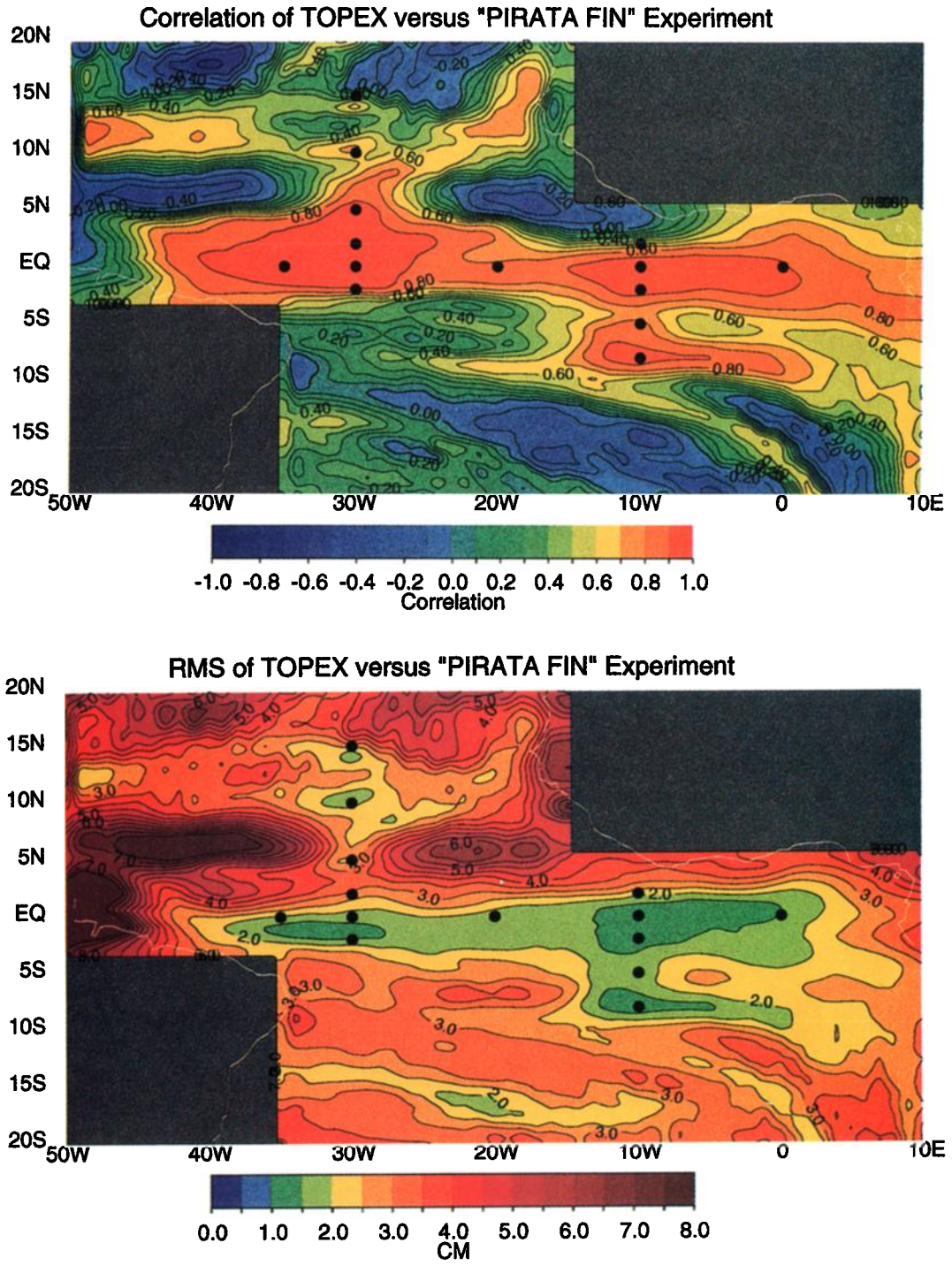

Plate 7. Model sea level results for assimilation at the final proposed PIRATA array ("PIRATA FIN") versus the gridded TOPEX/Poseidon product: (top) correlation value in tenths; (bottom) RMS difference in centimeters. Solid circles represent data assimilation locations. Area-averaged correlations correspond to +0.76, +0.61, and +0.53 and RMS error averages equal to 2.9 cm, 3.3 cm and 3.3 cm for the areas encompassed by ±3°, ±6°, and ±10°, respectively. (These numbers can be directly compared to those found in Table 1.)

run. This run is known as the "XBT+PIRATA FIN" run. Correlation and RMS differences between the "XBT" and "XBT+PIRATA FIN" are presented in Plate 8 in order to show how the PIRATA array might contribute to observing the oceanic structure over and above existing in situ observations.

Any relatively large differences between the "XBT" and the "XBT+PIRATA FIN" runs can be explained by the contribution of the "PIRATA FIN" points. The only difference between these two runs is assimilation of data at the 14 PIRATA mooring locations. Plate 8 shows the relationship between the "XBT" run and the "XBT+PIRATA FIN" run. A correlation of +1.0 would indicate no contribution of the "PIRATA FIN"

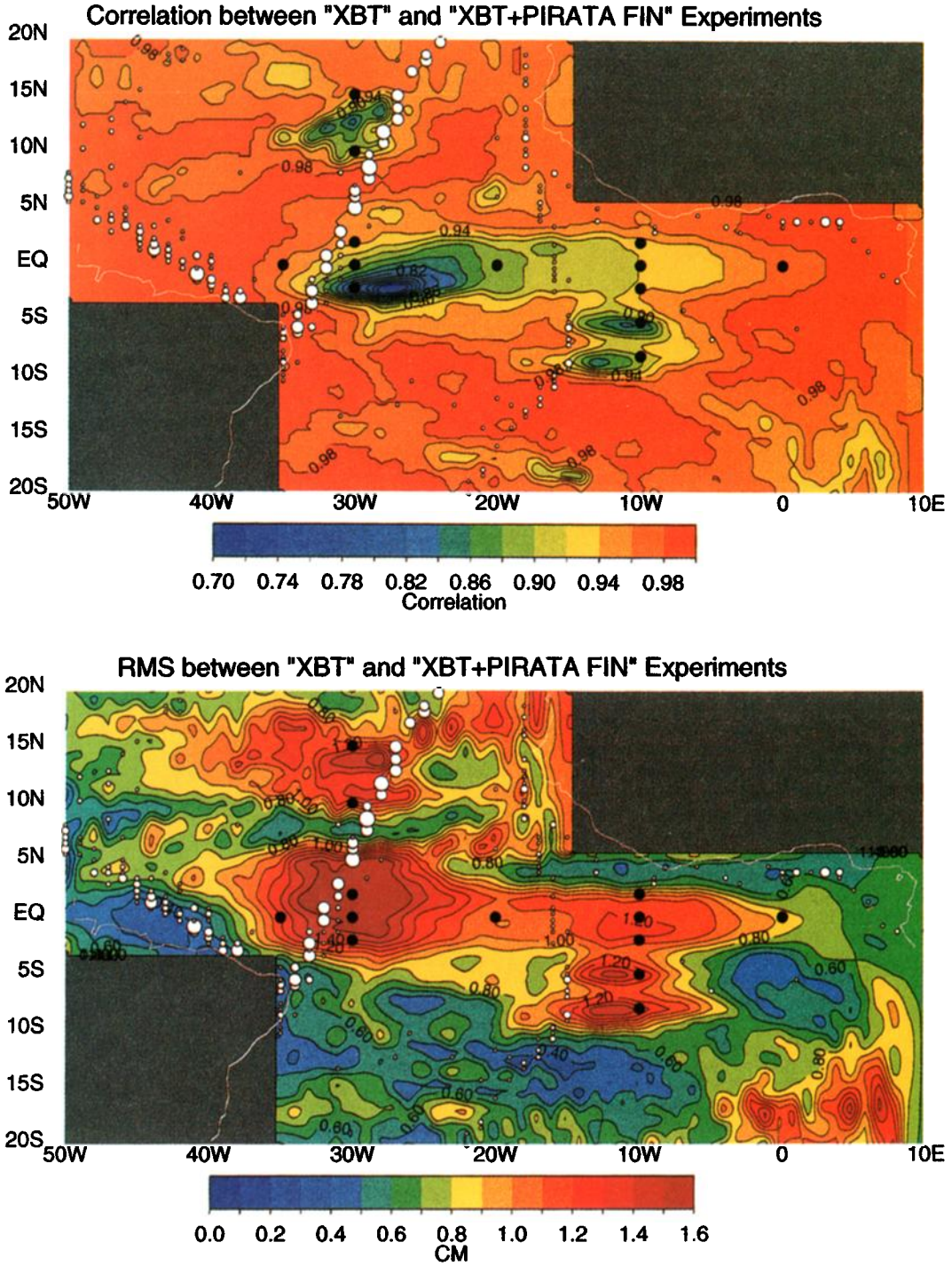

Plate 8. Model sea level results for assimilation at the XBT locations ("XBT") versus the run where both XBT locations and also the final PIRATA array locations are assimilated ("XBT+PIRATA FIN"): (top) correlation value with contours of +0.02; (bottom) RMS difference in tenths of centimeters. Note that the color scales for these plots go from +0.7 to +1.0 for the correlation panel and 0.0 to 1.6 cm for the RMS panel, unlike previous plots. Black circles represent data assimilation locations for the PIRATA final array configuration. Size and location of the white dots define XBT data coverage. Smallest white dot denotes a model grid point which assimilates data for 4 months; the largest white dot corresponds to a point where 13 months of data are assimilated. Area-averaged correlations correspond to +0.94, +0.95, and +0.96 and RMS error averages equal 0.9 cm, 0.9 cm, and 0.8 cm for the areas encompassed by ±3°, ±6°, and ±10°, respectively.

points (i.e., the "XBT" run equals the "XBT+PIRATA FIN" run). Therefore the lower the correlation, the bigger the impact of the PIRATA points. The top panel of Plate 8 shows the correlation between these two runs and serves to highlight the contribution of the PIRATA data assimilation points. Here the range for the correlation coefficient has been changed to +0.7 to +1.0. Areas where the correlation drops below +0.94 indicate regions where assimilation at the PIRATA moorings has the biggest impact. This plate shows low correlation can be found within ±2° of the equator between 35°W and 5°W and in regions west of the off-equatorial PIRATA locations at 30°W and at 10°W. The biggest effect of assimilation of the TOPEX/Poseidon data at the PIRATA locations can be seen between 30°W, 2°S and 20°W, equator where the correlation drops below +0.82. The conclusion of section 5.3.1 and the results of Table 1 also point out the importance of the mooring sites along the equator between 20°W and 30°W. Relatively high (+0.96) correlation near the moorings at 2°N, 5°N, and 30°W can be explained by the fact that ships frequent this area when traveling the Europe to Brazil line. Large white circles in Plate 8 stretching from the Brazilian coast toward Europe define the World Ocean Circulation Experiment (WOCE) designated XBT line, AX11. Correlation coefficients to the west of 35°W and to the east of Greenwich meridian generally exceed the +0.98 level.

In the bottom panel of Plate 8, the RMS differences between sea level from the "XBT" run and the "XBT+PIRATA FIN" run are presented. A similar argument can be made for RMS as for the correlation. If the "PIRATA FIN" points had no impact, the "XBT" run would equal the "XBT+PIRATA FIN" run and the RMS difference would be zero. Therefore areas of high RMS difference correspond to regions where the "PIRATA FIN" points have the most influence. Note that the scale differs for this plot and ranges from 0.0 cm to 1.6 cm. Relatively large differences (RMS values exceeding 1.1 cm) can be seen throughout the equatorial waveguide between 35°W and Greenwich meridian where the PIRATA moorings are located. Another area exceeding 1.1 cm can be found to the west of the 10°N, 15°N, and 30°W PIRATA mooring locations. Assimilation at 5°N, 30°W and at 10°N, 30°W has little influence in the zonal band between these two latitudes. High seasonal variability and frequent sampling by the XBT data in this region serve to limit the incremental effectiveness of the PIRATA mooring assimilation. Outside of the waveguide encompassed by the PIRATA mooring sites, RMS deviations are small (values less than 0.6 cm) due to the remoteness of the assimilation sites relative to local decorrelation scales. Exceptions occur at the northeast and southeast boundaries of the grid, where high RMS values are caused by spurious boundary effects related to the assimilation procedure.

XBT observations in the tropical Atlantic are sporadic, and their numbers are highly variable. Vast areas of the ocean remain unsampled by volunteer observing ships, particularly near the equator, where wave dynamics play an important role in interannual and seasonal variability. The continuous data assimilation afforded by the regular series of observations at the PIRATA moorings allows the model to be anchored to the observed signal, especially in the equatorial waveguide. This shows the importance of having a permanent mooring array providing consistent time series located in the tropical Atlantic Ocean.

7. Summary

In this paper a series of OSSEs are presented in which optimal locations for a proposed monitoring array in the tropical Atlantic Ocean are identified. A reduced-space Kalman filter technique is used to combine data with a simple linear model for the tropical Atlantic. One important aspect of this study is that these OSSEs assimilate TOPEX/Poseidon data at a restricted number of locations in an attempt to reconstruct the complete field of observed sea level from the subsampled altimetry data. Most OSSEs take an identical twin approach: A synthetic data set is subsampled according to an observing plan, and the subsampled data are then used to reconstruct the full synthetic data set. The current study uses real TOPEX/Poseidon data because the unprecedented accuracy and extensive coverage of this data set allows us to avoid the question of how well a model-generated synthetic data set reproduces the real ocean.

Another important aspect of this study is the way the forecast error covariance is utilized to determine the optimal locations of moorings in the tropical Atlantic. Within the context of the Kalman filter, the forecast error covariance ($\mathbf{HP}^f\mathbf{H}^T$, where $\mathbf{P}^f$ is the model error covariance and $\mathbf{H}$ is the matrix which maps the model state vector to the sea surface height at the 66 "HIGH DEN" locations) contains the optimal filter for the solution of the placement problem. Operating under the assumption that the three points that do the best job of explaining the differences between prediction and observation at the other 63 points of the "HIGH DEN" array should be optimal points for assimilation, an experiment was performed in which all groups of three moorings from the "HIGH DEN" mooring locations are systematically tested. Performance was evaluated according to the criterion that three points are optimal if they explain the highest average variance for the other 63 mooring locations. This is similar in spirit to the approach of *McIntosh* [1987] and *Barth* [1992] in which locations were chosen to provide the greatest amount of independent data. Results of this experiment show that the best choice is to place three moorings on the equator at 35°W, 20°W, and 10°W. The process described in this study can be simply applied to any data assimilation technique that keeps track of the error structure.

Using the results of the current study, the PIRATA Science Working Group adopted three equatorial sites (30°W, 20°W, and 10°W) for their proposed initial configuration of the mooring array. This marks the first time that OSSEs which incorporate dynamical models have been used to optimize exact mooring locations prior to deployment of a major mooring array in the tropics. The technique employed in this study has implications for future OSSEs because it is general enough to apply to any experiment design.

Much information has been gained on the dynamics of the tropical Atlantic through analysis of recently available altimetric data [*Katz et al.*, 1995; *Carton.*, 1989; *Arnault and Cheney*, 1994]. However, there exist very few in situ observations for the tropical Atlantic. For example, during a typical month, only eight tide gauge stations report sea level and 54 XBTs are dropped to observe subsurface temperature for the entire tropical Atlantic between 30°N and 30°S. Very few of these observations are taken within the equatorial waveguide where changes in the upper ocean thermal structure signal large-scale variability which is similar to the ENSO phenomenon in the Pacific. Therefore it is important to establish an observation system which can continuously monitor changes in the large-

scale equatorial signal without relying on altimeter data. The proposed PIRATA array is intended to supplement the presently sparse in situ observing system and altimetry in the tropical Atlantic by monitoring the changes in the subsurface thermal structure along the equator. Not only will the array monitor the large-scale dynamic topography changes across the Atlantic, but it also has been designed to measure variables critical for the understanding of the coupling of the ocean and atmosphere, namely, surface winds, SST, and surface heat fluxes which are important for studying climate fluctuations in the region and are not measured directly by altimeters, per se. The present study has focused on a monitoring strategy for the basin-scale dynamics responding to surface wind forcing. Although analysis of SST was beyond the scope of this paper, subsequent design experiments will utilize a nonlinear primitive equation model that reasonably simulates tropical Atlantic SST. This will permit follow-on OSSEs to be performed that will consider the optimal sampling of the thermodynamic forcings.

Acknowledgments. We would like to thank A. Bennett, M. McPhaden, M. Rienecker, J. Servain, and the anonymous reviewers for their helpful comments during the writing and revision of this manuscript. We would also like to thank S. Blue for programming support, B. Blumenthal for help with the model, C. Koblinsky and B. Beckley for providing the TOPEX/Poseidon data, D. Behringer for XBT locations, and R. Ray for helpful discussions regarding tide models. This research is supported in part by NASA RTOP 578-21-03, TOPEX grant 666-55-24, and NOAA Office of Global Programs Grant NA-36-GP0123-01.

References

Andersen, O. B., P. L. Woodworth, and R. A. Flather, Intercomparison of recent ocean tide models, *J. Geophys. Res.*, *100*, 25,261–25,282, 1995.

Arnault, S., and R. E. Cheney, Tropical Atlantic sea level variability from GEOSAT (1985–1989), *J. Geophys. Res.*, *99*, 18,207–18,223, 1994.

Atlas, R., S. C. Bloom, R. N. Hoffman, J. Ardizzone, and G. Brin, Space-based surface wind vectors to aid understanding of air-sea interactions, *Eos Trans. AGU*, *72*, 201–208, 1991.

Barth, N. H., Oceanographic experiment design, II, Genetic algorithms, *J. Atmos. Oceanic Technol.*, *9*, 434–443, 1992.

Bennett, A. F., Inverse methods in assessing ship-of-opportunity networks and estimating circulation and winds from tropical expendable bathythermograph data, *J. Geophys. Res.*, *95*, 16,111–16,148, 1990.

Bennett, A. F., *Inverse Methods in Physical Oceanography*, 346 pp., Cambridge Univ. Press, New York, 1992.

Bretherton, F. P., R. E. Davis, and C. B. Fandry, A technique for objective analysis and design of oceanographic experiments applied to MODE-73, *Deep Sea Res.*, *23*, 559–582, 1976.

Bretherton, F. P., M. J. McPhaden, and E. B. Kraus, Design studies for climatological measurements of heat storage, *J. Phys. Oceanogr.*, *14*, 318–337, 1984.

Busalacchi, A. J., and J. Picaut, Seasonal variability from a model of the tropical Atlantic Ocean, *J. Phys. Oceanogr.*, *13*, 1564–1587, 1983.

Busalacchi, A. J., M. J. McPhaden, and J. Picaut, Variability in equatorial Pacific sea surface topography during the verification phase of the TOPEX/POSEIDON mission, *J. Geophys. Res.*, *99*, 24,725–24,738, 1994.

Cane, M. A., and R. J. Patton, A numerical model for low-frequency equatorial dynamics, *J. Phys. Oceanogr.*, *14*, 1853–1863, 1984.

Cane, M. A., A. Kaplan, R. N. Miller, B. Tang, E. C. Hackert, and A. J. Busalacchi, Mapping tropical Pacific sea level: Data assimilation via a reduced state space Kalman filter, *J. Geophys. Res.*, *101*, 22,599–22,617, 1996.

Carton, J. A., Estimates of sea level in the tropical Atlantic Ocean using Geosat altimetry, *J. Geophys. Res.*, *94*, 8029–8039, 1989.

Chan, N. H., J. B. Kadane, R. N. Miller, and W. Palma, Estimation of tropical sea level anomaly by an improved Kalman filter, *J. Phys. Oceanogr.*, *26*, 1286–1303, 1996.

Chelton, D. B., A. M. Mestas-Nuñez, and M. H. Freilich, Global wind stress and Sverdrup circulation from the Seasat scatterometer, *J. Phys. Oceanogr.*, *20*, 1175–1205, 1990.

Cohn, S. E., and R. Todling, Approximate data assimilation schemes for stable and unstable dynamics, *J. Meteorol. Soc. Jpn.*, *74*, 63–75, 1996.

Fukumori, I., Assimilation of TOPEX sea level measurements with a reduced-gravity shallow water model of the tropical Pacific Ocean, *J. Geophys. Res.*, *100*, 25,027–25,039, 1995.

Fukumori, I., J. Benveniste, C. Wunsch, and D. B. Haidvogel, Assimilation of sea surface topography into an ocean circulation model using a steady-state smoother, *J. Phys. Oceanogr.*, *23*, 1831–1855, 1993.

Ghil, M., and P. Malanotte-Rizzoli, Data assimilation in meteorology and oceanography. *Adv. Geophys.*, *33*, 141–266, 1991.

Harrison, D. E., Local and remote forcing of ENSO ocean waveguide response, *J. Phys. Oceanogr.*, *19*, 691–695, 1989.

Harrison, D. E., W. S. Kessler, And B. J. Giese, Ocean circulation and model hindcasts of the 1982–83 El Niño: Thermal variability along the ship-of-opportunity tracks, *J. Phys. Oceanogr.*, *19*, 397–418, 1989.

Hastenrath, S., *Climate and Circulation of the Tropics*, 455 pp., D. Reidel, Norwell, Mass., 1985.

Hayes, S. P., L. J. Mangum, J. Picaut, A. Sumi, and K. Takeuchi, TOGA-TAO: A moored array for real-time measurements in the tropical Pacific Ocean, *Bull. Am. Meteorol. Soc.*, *72*, 339–347, 1991.

Hellerman, S., and M. Rosenstein, Normal monthly wind stress over the world ocean with error estimates, *J. Phys. Oceanogr.*, *13*, 1093–1104, 1983.

Katz, E. J., J. A. Carton, and A. Chakraborty, Dynamics of the equatorial Atlantic from altimetry, *J. Geophys. Res.*, *100*, 25,061–25,067, 1995.

Kessler, W. S., M. C. Spillane, M. J. McPhaden, and D. E. Harrison, Scales of variability in the equatorial Pacific inferred from the Tropical Atmosphere-Ocean (TAO) buoy array, *J. Clim.*, *9*, 2999–3024, 1996.

McIntosh, P. C., Systematic design of observational arrays, *J. Phys. Oceanogr.*, *17*, 885–902, 1987.

McPhaden, M. J., G. Reverdin, J. Merle, Y. Du Penhoat, and A. Kartavtseff, Objective analysis of simulated equatorial Atlantic Ocean data on seasonal time scales, *Deep Sea Res.*, *31*, 551–569, 1984.

McPhaden, M. J., et al., Tropical Ocean Global Atmosphere observing system: A decade of progress, *J. Geophys. Res.*, in press, 1998.

Merle, J., Seasonal heat budget in the equatorial Atlantic Ocean, *J. Phys. Oceanogr.*, *10*, 464–469, 1980.

Meyers, G., H. Phillips, N. Smith, and J. Sprintall, Space and time scales for optimal interpolation of temperature—Tropical Pacific Ocean, *Oceanogrpahy*, *28*, 189–218, 1979.

Miller, R. N., Tropical data assimilation experiments with simulated data: The impact of the Tropical Ocean and Global Atmosphere Thermal Array for the Ocean, *J. Geophys. Res.*, *95*, 11,461–11,482, 1990.

Miller, R. N., and M. A. Cane, Tropical data assimilation: Theoretical aspects, in *Modern Approaches to Data Assimilation in Ocean Modeling*, edited by P. Malanotte-Rizzoli, pp. 207–234, Elsevier Sci., New York, 1996.

Miller, R. N., A. J. Busalacchi, and E. C. Hackert, Sea surface topography fields of the tropical Pacific from data assimilation, *J. Geophys. Res.*, *100*, 13,389–13,425, 1995.

Moura, A., and J. Shukla, On the dynamics of droughts in northeast Brazil: Observations, theory, and numerical experiments with a general circulation model. *J. Atmos. Sci.*, *38*, 2653–2675, 1981.

National Academy of Sciences, *Ocean-Atmosphere Observations Supporting Short-Term Climate Predictions*, 51 pp., Natl. Acad. Press, Washington, D.C., 1994.

Nerem, R. S., E. J. Schrama, C. J. Koblinsky, and B. D. Beckley, A preliminary evaluation of ocean topography from the TOPEX/POSEIDON mission, *J. Geophys. Res.*, *99*, 24,565–24,583, 1994.

Picaut, J., A. J. Busalacchi, M. J. McPhaden, L. Gourdeau, F. I. Gonzalez, and E. C. Hackert, Open-ocean validation of TOPEX/Poseidon sea level in the western equatorial Pacific, *J. Geophys. Res.*, *100*, 25,109–25,127, 1995.

Ray, R. D., B. Sanchez, and D. E. Cartwright, Some extensions to the response method applied to TOPEX/Poseidon (abstract), *Eos Trans. AGU*, *75*, Spring Meeting Suppl., 108, 1994.

Schröter, J., and C. Wunsch, Solution of nonlinear finite-difference

ocean models by optimization methods with sensitivity and observational strategy analysis. *J. Phys. Oceanogr.*, *16*, 1855–1874, 1986.
Servain, J., Simple climatic indices for the tropical Atlantic Ocean and some applications, *J. Geophys. Res.*, *96*, 15,137–15,146, 1991.
Servain, J., J. N. Stricherz, and D. L. Legler, *TOGA Atlantic pseudo-stress Atlas: 1985–1994*, 160 pp., Inst. Fr. de Rech. pour l'Exploit. de la Mer, Cent. ORSTOM de Brest, Plouzanè, France, 1996.
Zebiak, S. E., Air-sea interaction in the equatorial Atlantic region, *J. Clim.*, *6*, 1567–1586, 1993.

A. J. Busalacchi and E. C. Hackert, Laboratory for Hydrospheric Processes, NASA Goddard Space Flight Center, Greenbelt, MD 20771. (e-mail eric@maya.gsfc.nasa.gov)
R. N. Miller, College of Oceanic and Atmospheric Sciences, Oregon State University, Corvallis, OR 97331.

(Received September 19, 1996; revised September 23, 1997; accepted November 4, 1997.)