Signal and noise in tropical Pacific sea level height analyses*

A. Kaplan†, M. A. Cane, D. Chen, D. L. Witter‡
Lamont-Doherty Earth Observatory of Columbia University,
Palisades, NY 10964

R.E. Cheney
NOAA/NESDIS Laboratory for Satellite Altimetry,
Silver Spring, MD 20910

December 2002

Abstract

Monthly interannual anomalies of tropical Pacific sea level height from Topex/Poseidon altimetry are compared with simulation and assimilation products from a variety of models, ranging from a simple linear long wave approximation to ocean general circulation models. Major spatial similarities in the error patterns are identified. These include zonally elongated maxima in the northwest and southwest tropical Pacific Ocean, a narrow band of high values near 10^oN which is slightly inclined towards the equator from the Central American coast, and low values on the equator and in the southeastern tropical Pacific. These features are also present in the pattern of small-scale variability of sea level height. Spatial and temporal components of this small-scale variability are analyzed for predominant variability types. Monte Carlo experiments identify the areas where high small-scale sea level height variability is wind-driven, caused by a similar pattern of variability in the wind stress. Model products systematically underestimate signal variance in such areas. Variability in other areas is due to the instability of ocean currents. The major component of uncertainty in the gridded satellite altimeter analyses is due to sampling error, for which estimates are developed and verified.

Keywords: Tropical sea level height, small-scale variability, ocean modeling.

*Submitted to *Journal of Geophysical Research – Oceans*
†To whom correspondence should be addressed: alexeyk@ldeo.columbia.edu
‡Presently at Kent State University, Department of Geology, Kent, OH 44242

1 Introduction

The use of observations in climate research normally requires data records substantially longer than most currently available sets of satellite data. But pre-satellite in situ ocean data sets are too sparse to describe complete fields. Climate researchers hope to fill in these fields using data assimilation. It is essential to learn as much statistical and dynamical information as possible from the satellite data sets in order to create a more accurate extension of such data sets back into the pre-satellite era.

Detailed analyses of the global surface ocean are available for the period after 1992 because of the high-quality and spatially expansive data coverage of the TOPEX/Poseidon (T/P) altimetry [*Behringer et al.*, 1998; *Carton et al.*, 2000ab; *Fukumori et al.*, 1999], but existing analyses of the earlier period [*Smith*, 2000; *Cane et al.*, 1996; *Reverdin et al.*, 1996] are less well validated and arguably of lower quality. In the case of the tropical Pacific Ocean, the lack of long high quality data sets limits our ability to improve El Niño–Southern Oscillation (ENSO) prediction techniques [*Chen et al.* 1998,2000].

Large-scale month-to-month variability of sea level height in the tropical Pacific Ocean is important for many reasons, perhaps the most prominent one being a close connection of this variability to ENSO. Tropical sea level height, due to its dynamical connection to the thermocline depth, arguably carries the most potent predictive signal for ENSO forecasts. Perhaps because of that, a simple intermediate model, like LDEO4 [*Zebiak and Cane*, 1987; *Chen et al.*, 2000] still can predict ENSO about as well as state-of-the-art coupled general circulation models (GCMs). Assimilation of observed sea level heights into a model for the tropical Pacific Ocean plays an important role in ENSO prediction. LDEO4 [*Chen et al.*, 2000] uses assimilations of sea level observations into its ocean component at the initialization step (in addition to surface winds and sea surface temperatures). State-of-the-art GCM-based schemes of ENSO prediction [*Schneider et al.*, 2001, 2002] initialize the ocean component with the states obtained by the assimilation of observed temperature profiles into a surface flux driven ocean GCM.

It is important to bring the wind-driven ocean model component into a realistic state by means of data assimilation. For the goal of climate prediction, weekly to monthly temporal resolution of ocean anomalies will suffice. At first sight, given the expansive and relatively precise satellite altimetry data from the T/P and ERS-2 missions plus the availability of weekly and monthly gridded products with well verified skill [*Cheney et al.*, 1994; *Ducet et al.*, 2000], the task doesn't look very daunting. However, it is necessary to distinguish between errors in the observed sea surface height and contributions to the signal by processes not simulated in the model, as well as to take into account the model internal error and the error stemming from the incorrect surface flux fields, most notably, wind stress. Variability in nature exists on all spatial and temporal scales, including those smaller than the resolution of model and observational data sets. Imperfect parameterization of this variability in models and incomplete sampling of it by observational systems creates model and observational error on the resolved scales of variability. All of these considerations make the optimal assimilation of the altimetry data, and the proper attribution of errors to the model or to the altimetry data a very involved procedure [e.g. *Fukumori et al.*, 1999; *Tokmakian and Challenor*, 2000]. The chaotic nature of the climate system puts a particularly high demand on the accuracy

of the initial state and its consistency with the system dynamics, and small changes in the initial condition can have a tremendous effect on the forecast skill [*Chen et al.*, 2000; *Cañizares et al.*, 2001].

In this paper we focus on the tropical Pacific sea level height anomalies. After a description of data and methods (Section 2), we first present in Section 3 analyses of the tropical Pacific sea level anomaly fields which combine a wind-driven linear ocean model by *Cane and Patton* [1984] with a few different sets of sea level observations using a reduced space optimal smoother [*Kaplan et al.*, 1997], and compare results to the monthly T/P altimetry fields from *Cheney et al.* [1994], which have a spatial resolution of 4^o longitude and 1^o latitude. We discuss our ability to constrain sea level height anomalies on these spatial and temporal scales and refer to the corresponding gridded values as "signal". A persistent pattern of seemingly irreducible error variance which emerges from these comparisons is shown to dominate the differences with the T/P altimetry for a wide variety of products from ocean models and data assimilation systems (section 3). We show that this pattern can be traced to the small-scale sea level height variability in the ocean, or "noise", which we define as the variability on scales smaller and shorter than those of the "signal" (Section 4). Incomplete sampling of this variability is responsible for a large part of the error in the gridded altimetry product. Thus we obtain an error estimate for the T/P altimetry product of *Cheney et al.* [1994], and verify it with the altimetry product of *Ducet et al.* [2000] and tide gauge data. In Section 5 we calculate the pattern of small-scale variability in surface winds, and use Monte Carlo experiments with a linear ocean model to imply that most of the observed small-scale variability in sea level heights is driven by that in surface winds. Areas where high energy small-scale variability cannot be modeled as wind-driven are associated with eastward ocean currents, particularly North Equatorial Countercurrent (Section 6). We show that the small-scale variability can be separated into its temporal and spatial components whose ratio has an almost uniform meridional structure, determined by the dispersion relations of the ocean waves (Section 7). In the discussion of Section 8 we connect small-scale variability with a well-known pattern of the mesoscale eddy energy and demonstrate that even high resolution models heavily underestimate this variability, even if driven by momentum flux fields with adequately represented short-term and small-scale variability. Thus the underestimation of the small-scale variability translates into an inhomogeneous pattern of systematic underestimation of variance in the ocean models. Consequences of this error for data assimilation systems and ways to obtain more adequate estimates for model and altimetry data error are discussed in Section 8 as well. Conclusions are presented in Section 9.

2 Data and methods

Altimetry data. In this paper we discuss various analyses of interannual variability of sea level height in the tropical Pacific Ocean. As a benchmark for comparisons we use the gridded T/P altimetry product of *Cheney et al.* [1994]. This product is produced by a straightforward gridding technique: on a uniform 4^o longitude by 1^o latitude grid the monthly value of each point is computed as the average of altimetric values from all 1^o track segments that cross the given grid

cell during that month. Depending on the location, there are usually from 6 to 12 crossings every month. All model products compared here are averaged into monthly values on the same $4^o \times 1^o$ grid as that of *Cheney et al.* [1994].

Definition of signal and noise variability. In nature there is variability of physical variables on all spatial and temporal scales. When these variables are put into a gridded form for use in numerical models or for presentation in observational data sets, the size of grid boxes is chosen, and all data is reduced to the space–time averages over these gridboxes (which we will also call space-time bins). Scientists usually pay attention to the variability of the bin averages (which for the purpose of this paper we will call "signal"), although the variability inside the bins (which we will call "noise") is important for many practical purposes as well.

Note, that this division of the total variability of a physical variable into signal and noise depends on the space and time dimensions of a chosen bin. A common attitude is to try to make the bins as small as possible, provided there are enough data or computer time and memory to represent their averages. Then the subgrid variability characterizes the noise, i.e. the effect of the natural processes that are responsible for sampling uncertainty in the estimates of bin averages.

In this paper we compute both signal and noise variability for a variety of model and data sea level height products, using *Cheney et al.*'s [1994] $4^o \times 1^o \times 1$ month bins. One can say that the "signal" here is virtually all of the spatial variability that can be constrained by the T/P data on a monthly basis. To make the presentation of our results unambiguous, we introduce a slightly unconventional but quite convenient notation. It can be viewed as an extension of the notation introduced originally by *Starr and White* [1954] and widely used since [e.g. *Peixoto and Oort*, 1995]. Unlike them, we will use square brackets not only for zonal averages, but for averages over any space-time regions. The region of averaging will be specified by a subscript, to avoid confusion. Similarly, a functional σ will denote the standard deviation of its argument over the region indicated in its subscript.

Suppose $s(x, y, t)$ is a continuous physical variable within a spatial gridbox G and a time interval T. Then the mean sea level anomaly for the space-time bin $G \times T$ ("signal") is

$$[s]_{G\times T} = \frac{1}{|G\times T|} \int_T dt \iint_G sAdxdy, \tag{1}$$

where $Adxdy$ is the element of the ocean surface, and $|G\times T| = |G||T|$, where $|G| = \iint_G Adxdy$ and $|T| = \int_T dt$ are weights for spatial and temporal averaging respectively.

The variability $\sigma_{G\times T}(s)$ inside the bin (i.e. "noise" variability) is the root-mean-squared (RMS) deviation of s from its mean $[s]_{G\times T}$:

$$\sigma^2_{G\times T}(s) = \frac{1}{|G\times T|} \int_T dt \iint_G (s - [s]_{G\times T})^2 Adxdy. \tag{2}$$

If some averaging is done before the variability is evaluated, the σ operation applies to averages, e.g. the intrabin temporal variability of spatial averages

$$[s]_G = \frac{1}{|G|} \iint_G sAdxdy$$

can be written as

$$\sigma_T^2([s]_G) = \frac{1}{|T|}\int_T ([s]_G - [s]_{G\times T})^2 dt. \tag{3}$$

Or additional averaging may be done after the variability is estimated:

$$[\sigma_G^2(s)]_T = \frac{1}{|T|}\int_T \sigma_G^2(s)dt = \frac{1}{|G||T|}\int_T dt \iint_G (s - [s]_G)^2 A dx dy.$$

It is easy to show that the last two variability values can be used for separating total noise variability into its spatial and temporal components:

$$\sigma_{G\times T}^2(s) = [\sigma_G^2(s)]_T + \sigma_T^2([s]_G). \tag{4}$$

Angle brackets will denote statistical ensemble averaging, which, under the ergodicity assumption, in all practical calculations we replace by averaging over a temporal sample. For example, $\langle \sigma_{G\times T}^2(s)\rangle_{\text{all } T}$ is the variance inside $G\times T$ bins averaged over all time periods T available for the spatial gridbox G.

Anomaly comparison. All comparisons are done for the period October 1992 – March 2001 when the altimetry product is available or for the period of its longest overlap with a model product period. Since it is the interannual variability that is being studied here, when a pair of data sets is being compared, their mean monthly seasonal cycles for the longest common time period are estimated and subtracted from each of them. The comparison of the resulting monthly anomalies is then presented as a field of RMS differences between them.

3 Performance of model products

The linear wind-driven model used here is the same as that used by *Cane et al.* [1996] and *Reverdin et al.* [1996]. The model has 2^o zonal by 0.5^o meridional resolution and uses the *Cane and Patton* [1984] numerical algorithm for solving the linear long-wave approximation to the shallow water equations on the equatorial β-plane. The model assumes a uniform density profile and uses the first two baroclinic modes to construct the solution.

This linear model has no thermodynamics or salinity; it is forced by the smoothed and detrended monthly pseudostress anomalies derived from the Florida State University (FSU) analysis [*Stricherz et al.*, 1997]. Figure 1 (upper left panel) shows by comparison with the T/P altimetry [*Cheney et al.*, 1994] that even such a simple model, with no sea level height data assimilated, is not totally without skill. Model simulation of monthly anomalies averaged into $4^o\times 1^o$ grid boxes are within 5-8 cm of the altimetry anomalies, and there are narrow strips in which the wind-driven model achieves 2-4 cm accuracy.

For assimilating observations into this model we employ the reduced space optimal smoother (RSOS) [*Kaplan et al.*, 1997] which uses all available data (past, present, and future) to estimate the model state at each assimilation step. In the RSOS we employ the same covariance settings and assumptions on model and data error used in *Cane et al.* [1996] and *Reverdin et al.* [1996]. Assimilation is implemented in a reduced space of 32 degrees of freedom. The patterns spanning

the reduced space are found as leading empirical orthogonal functions (EOFs) of the wind-driven model run for the period 1964–1993 without data assimilation. Sea level height observations are assimilated once a month.

Our in situ data are either measurements from 34 tide gauge stations in the tropical Pacific as in [*Cane et al.,* 1996] or dynamic heights estimates from a few hundred temperature profiles (mostly expendable bathythermographs and TAO moorings) as in [*Reverdin et al.*, 1996]. Following *Miller and Cane* [1989] and *Cane et al.* [1996] we assume that the system noise (error generated by the model at every time step) is dominated by the error in wind forcing, and we use their scheme for obtaining a system noise covariance estimate from a Monte Carlo model run. In this run the model is forced by a random multivariate Gaussian error $e(x, y, t)$ in the zonal kinematic windstress with covariance

$$\langle e(x_1, y_1, t_1)e(x_2, y_2, t_2)\rangle = W\delta(t_1 - t_2)e^{-(x_1-x_2)^2/L_x^2-(y_1-y_2)^2/L_y^2}. \quad (5)$$

It depends on the amplitude W and decorrelation length scales L_x and L_y. Wind errors at different months are taken to be uncorrelated. Assimilations of Figure 1 follow *Cane et al.* [1996] using decorrelation scales $L_x = 10^o$ and $L_y = 4^o$, and a uniform

$$W = ((\rho_a/\rho_w)C_D\sigma_{p\tau_x})^2. \quad (6)$$

Here ρ_a, ρ_w are densities of air and water respectively, C_D is a drag coefficient, and the standard error in the zonal pseudostress is assumed $\sigma_{p\tau_x}$ =12.8m^2/s^2.

Assimilation of just 34 tide gauge records for the entire tropical Pacific area results in impressive improvement of the hindcast product (Figure 1, upper right panel). Assimilation of a few hundred temperature profiles produces a similar overall improvement with a somewhat different spatial structure which reflects the main patterns of data availability (Figure 1, lower left panel).

In the case of the full-grid assimilation system, assimilating altimetry for every ocean gridpoint can bring the analysis results as close to the assimilated data as one deems necessary: the level of the system's "belief" in the data is determined by the assumed model-to-data noise variance ratio. In the case of the reduced space data assimilation, however, we can correct the model only by the fields from a low-dimensional linear subspace. All patterns which are orthogonal to this subspace belong to "effective" (for this system) observational error; the analysis fields cannot possibly get any closer to the assimilated data than to within the size of this error [*Cane et al.*, 1996].

As Figure 1 (lower right panel) demonstrates, a drastic improvement due to T/P altimetry assimilation happens only on the equator and in the southeastern part of the domain. While assimilation of the altimetry brings the solution closer to it virtually everywhere (to within 4.0 cm on average), more than a half of the improvement over a purely wind-driven solution (whose average RMS difference from the T/P data is 5.9 cm) is captured by the assimilation of the data from the sparse in situ networks: average RMS difference from T/P is 5.0 cm for the assimilation of tide gauges and 4.9 cm for the assimilation of temperature profiles. Assimilation of the altimetry further improves estimates but only to a point: it has limited influence on the high RMS differences in the zonally elongated areas in the northwestern and southwestern parts of the domain and in a band near 10^{o}N in the eastern part of the basin. High values in these areas are characteristic of the standard deviation pattern for the portion of T/P altimetry fields which does not project on our set

of corrections, i.e. this portion cannot be represented through the patterns of variability produced by the linear model under the FSU wind forcing.

It appears to be most likely that it is our data assimilation methodology (a restricted low-dimensional set of corrections) and the simplicity of the model itself which are responsible for the inability of the assimilation procedure to approach the altimetry product more closely. Testing this idea, we compare T/P altimetry with simulation and assimilation products from a few state-of-the-art GCMs.

The OGCM comparisons are presented in Figure 2. The NCEP product of *Behringer et al.* [1998] uses a Pacific basin version of MOM1 and the *Derber and Rosati* [1989] scheme to assimilate temperature profiles and altimetry. The results are no closer to the T/P data than the assimilation into the linear model, spatially averaged RMS difference being 4.0 cm as well (Figure 1, lower right panel).

Fukumori et al. [1999] developed a sophisticated assimilation system for a global domain MOM1 model. Their simulation uses forcing by daily winds and monthly climatological heat fluxes from the NCEP analysis. The assimilated version combines the model with along-track T/P altimetry values binned with 2.5^o latitudinal resolution. This product is available for 3 years (1993-1995). The comparison shows 3-4 cm RMS difference for the simulation and 1.5-3 cm difference for the assimilation in most of the domain (spatial averages are 4.4 cm and 3.4 cm respectively). The pattern that we identified earlier can be seen in these plots as well.

The high-resolution POCM 4C model [*Tokmakian and Challenor,* 1999; *Semtner and Chervin,* 1992] driven by daily fluxes from the ECMWF (reanalysis before 1994 and the operational analysis starting from 1994) also shows this pattern, although embedded into a more complicated pattern of other model errors (average RMS difference is 5.26 cm).

The *Carton et al.* [2000ab] global ocean analysis uses the *Derber and Rosati* [1989] scheme to assimilate all available in situ and altimetry data. Comparisons of the T/P altimetry with two of their products, one which assimilates only in situ data, and another which uses the altimetry as well, both show the same pattern of maximum differences, though the altimetry assimilation is very close to the T/P altimetry data (spatially averaged RMS differences are 4.4 cm and 2.9 cm respectively).

Therefore the estimates of sea level height anomaly fields produced by a variety of ocean models and data assimilation systems differing dramatically in their degree of sophistication all show spatially similar features in their pattern of differences from the T/P altimetry: zonally elongated maxima in the northwest and southwest tropical Pacific Ocean, a narrow band of high variance near 10^oN slightly inclined towards the equator from the Central American coast, and low values on the equator and in the southeastern tropical Pacific. Either there must be some similarity in the problems these models have with simulating sea level height variability or the problem lies with the altimetry fields.

4 Small-scale variability and error of gridded altimetry products

Figure 3 (top left panel) shows that the error pattern identified in the previous section is similar to the pattern of "noise" variance in the ocean, which we define here as the variability inside $4^o \times 1^o$ monthly bins.

In a practical calculation with the T/P altimetry, instead of a continuous distribution $s(x, y, t)$ inside a bin we only have values from those 6 to 12 T/P tracks that pass through a $4^o \times 1^o$ grid box G during a given month T: $\{s_i, \quad i = 1, \cdots, N_{G\times T}\}$. Equation (1) has to be replaced by a sample estimate for $[s]_{G\times T}$ [cf. *Cheney et al.*, 1994]:

$$\widehat{[s]}_{G\times T} = \frac{1}{N_{G\times T}} \sum_{i=1}^{N} s_i. \tag{7}$$

Equation (2) is similarly replaced by

$$\widehat{\sigma}^2_{G\times T}(s) = \frac{1}{N_{G\times T} - 1} \sum_{i=1}^{N_{G\times T}} (s_i - \widehat{[s]}_{G\times T})^2 \tag{8}$$

to obtain an unbiased sample estimate of the noise variability inside a $G \times T$ bin (e.g. *Mardia et al* [1979]). To get a more reliable estimate of this variability the values computed by equation (8) for every month are then averaged over all months in the period from October 1992 to February 1999, and square roots $\sqrt{\langle \widehat{\sigma}^2_{G\times T}(s) \rangle_{\text{all } T}}$ are presented in a spatial pattern of the small-scale variability in Figure 3 (top left panel).

The influence of this variability on the grid box average adds a sampling error to the error of individual altimetry track values. Since an estimate (7) involves averaging of $N_{G\times T}$ observations, it is affected by a sample error whose standard deviation can be estimated as $r_{G\times T} = \hat{\sigma}_{G\times T}/\sqrt{N_{G\times T}}$, assuming optimistically that all $N_{G\times T}$ altimeter passes in the month T sample uncorrelated deviations of the sea surface height from its gridbox mean $s_{G\times T}$. By averaging $r^2_{G\times T}$ over all months from October 1992 to February 1999, we are able to present here a crude estimate $\sqrt{\langle r^2_{G\times T} \rangle_{\text{all } T}}$ of the sampling error contribution to the total error of the *Cheney et al.* [1994] gridded altimetry product (Figure 3, lower left).

Formally, the values $r_{G\times T}$ provide a lower bound for the gridded altimetry product, because they account for only one of the possible error sources (sampling error), and in computing the latter we assumed that errors of individual track values are uncorrelated. This optimistic estimate may require inflating by some factor $\alpha > 1$ to become more realistic.

To estimate the size of α, we compare T/P data with tide gauge measurements from the Sea Level Data Center at the University of Hawaii. We use monthly sea level height anomalies s_g (tidal variations removed and corrections for the inverse barometer effect applied) at 31 stations which have more than 8 years of data during the T/P period and are located inside $4^o \times 1^o$ *Cheney et al.* [1994] grid boxes with more than 6 track crossings per month on average. Note that in their comparison with tide gauges, *Cheney et al.* [1994] often choose a grid box which is not the closest

to the station, because the presense of the land in the grid box adversely affects the sampling and error of the altimetric measurements. Being primarily interested in estimates of noise, we do not use their approach here.

For a given station g the squared difference $d_g^2 = \langle([s_g]_T - [s]_{G\times T})^2\rangle_{\text{all } T}$ is expected to be

$$d_g^2 = \alpha^2 r_{G\times T}^2 + r_g^2, \tag{9}$$

where r_g is the RMS error of tide gauge values as estimators of the grid box mean, i.e. it includes the tide gauge error and the physical deviation of a pointwise monthly tide gauge value from that of $4^o\times 1^o$ resolution. Figure 4 (top panel) illustrates for Christmas Island the contrast between monthly sea level height anomalies from either source (T/P or tide gauge records) and the scatter of individual altimetry passes.

Figure 4 (lower left panel) shows d_g versus $r_{G\times T}$ for all 31 stations. As expected from (9) with $\alpha > 1$, the former exceeds the latter: all circles are higher than the solid line. With d_g and $r_{G\times T}$ computed, (9) gives a constraint on α and r_g. Elliptical arcs of corresponsing curves on the (r_g, α) plane are drawn for all 31 stations in Figure 4 (lower right panel). For most of the stations α is less than 2 for all values of r_g. Out of the four stations which allow $\alpha > 3$ (Johnston, Rarotonga, Hilo, and Nauru) the former three are located at around 20^o latitude where the comparison is known to be problematic [*Mitchum*, 1994]: sea level height variability is dominated by planetary waves that are slow enough to create significant phase shifts between the tide gauge and altimetric values. Assuming that α has a common value for all stations, we form a common constraint by taking the median over all stations of both sides in Equation (9). Resulting curve gives $\alpha = 1.5$ for median r_g equal 1 cm, and smaller values for larger tide gauge errors. The line corresponding to $\alpha = 1.5$ is shown in Figure 4 (lower left panel). Even for a precisely correct value of α a number of circles can be expected to fall below the error line, because all our error estimates themselves are subjects to the sampling error. The actual number of circles below depends on the tide gauge error r_g: if it is large, only clear outliers can fall below the line, if it is close to zero, the half of the circles can be below (cf. Equation (9)). The Monte Carlo experiment with normally distributed tide gauge and altimetric errors showed that the number of circles below the $\alpha = 1.5$ line (10) in Figure 4 (lower left panel) is consistent with r_g equal 1 cm or slightly smaller values.

To further verify the spatial structure of our error estimates we compute RMS differences between the *Cheney et al.* [1994] altimetry product and sea level anomaly estimates by *Ducet et al.* [2000] that merge T/P and ERS-1,2 measurements via an advanced global objective analysis (Figure 3, lower right panel). We expect the error of the *Ducet et al.* [2000] to be significantly smaller than that of *Cheney et al.* [1994], since they use more data. We also expect the error of the two analyses to be positively correlated, since T/P altimetry data is used by both. Consistent with that, the pattern of the RMS difference field is similar to the the pattern of error estimates, the former values exceeding the latter on average by only 6%.

Since our estimates of the T/P error do not require inflation by a factor higher than α =1.5, their relatively small size suggests that there is much more to the model–altimetry differences than just the altimetry error. We suspect that, like a gridded altimetry error, the GCMs' error has a spatial variance distribution similar to that of the small-scale sea level height variability (Figure 3, upper left panel). This variability is not being adequately captured even by the high resolution models, as

a similar computation (by a discrete version of formula (2)) for POCM 4C shows (Figure 3, upper right panel). To understand the nature of this variability and the reason for the models' failure, we start by interrogating the usual suspect: wind data.

5 Small-scale variability in surface winds and sea level height response

The top panel of Figure 5 shows the small-scale, short-term variability of the surface zonal wind calibrated to a height of 10 m above the ocean surface for neutral stability conditions (Level 3.0 NSCAT data [*JPL*, 1998]). The calculation is similar to that done for the sea level height variability in the previous section, except that for the wind we use $4^o \times 4^o$ monthly bins. This wind variability pattern is impressively more complete than that computed on the basis of ship data from COADS [*Woodruff et al.*, 1998] (Figure 5, bottom panel): places with poor data coverage show no variability. The meridional wind exhibits a similar spatial pattern of small-scale variability (not shown).

Figure 5 suggests that any gridded wind stress analysis produced from observations with less than total temporal and spatial coverage will have an error with a similar spatial variance and rather short spatial and temporal decorrelation scales. This error corresponds to the small-scale variability in the wind removed by the wind stress gridding procedure.

We estimate the standard deviation of the corresponding small-scale variability in the zonal pseudostress by multiplying the pattern of Figure 5 (top panel) by the climatological windspeed pattern evaluated from daily data of the NCEP-NCAR reanalysis [*Kalnay et al.*, 1996]. Maxima of this pattern seem to roughly correspond to the maxima in the small-scale sea level height variability (Figure 6, top panel), suggesting that the former is causing the latter.

To explore the causality hypothesis further, we generated random fields intended to represent the small-scale variability in the pseudostress and forced the *Cane and Patton* [1984] linear ocean model ($2^o \times 0.5^o \times 0.25$ month resolution) with it. The random field generation was done as described by equations (5) and (6), except that $\sigma_{p\tau_x}$ was the spatially variable standard deviation of small-scale zonal pseudostress variability (contours in Figure 5, top panel) rather than a constant. An independent pseudostress field (both zonal and meridional components) was generated at each model time step (a quarter of a month). We didn't perform a formal evaluation of spatial decorrelation scales for the small scale wind variability, but since we expect them to be shorter than the wind gridscale, we assumed $L_x = L_y = 1^o$ in equation (5).

Figure 7 presents (upper left panel) the sea level height response of the ocean model in terms of the variability inside $4^o \times 1^o$ monthly bins $\sqrt{\langle \sigma_{4^o \times 1^o \times 1 \text{ month}}(s)^2 \rangle_{\text{months}}}$ (small-scale variability averaged over the entire Monte Carlo run of 2000 months), and (lower left panel) the RMS of the means of these bins $\sigma_{\text{months}}([s]_{4^o \times 1^o \times 1 \text{ month}}) = \sqrt{\langle [s]^2_{4^o \times 1^o \times 1 \text{ month}} \rangle_{\text{months}}}$ (signal variability). The simulated spatial pattern of the small-scale variability (Figure 7, upper left panel) contains the major features of the observed pattern, but underestimates its amplitude by a factor of 2, due to the relative coarseness of the model grid and the omission of all variability mechanisms but the long planetary waves. Simultaneously with the small-scale variability, this random forcing with

the small (1^o) decorrelation scale causes a variability at signal resolution with a similar spatial structure and only slightly weaker magnitude (Figure 7, lower left panel).

The spatial scale of random wind variability has a large impact on the structure of the sea level height response. Right hand panels of Figure 7 illustrate sea level height variability obtained in response to the random wind fields with the same spatial pattern of W as was used to obtain left hand panels, but with rather long spatial decorrelation scales: $L_x = 20^o,\ L_y = 10^o$. Small-scale and signal scale variability patterns are quite dissimilar in this case, with the latter resembling the variability of the response to the realistic wind field. Both, however, are very different from the variance pattern for the response to the spatially uncorrelated wind shown in the left hand panels.

Overlaying the contours of the small-scale sea level height variability in responce to the random wind with the 1^o decorrelation scale on the pattern evaluated from the T/P data (Figure 6, bottom panel) identifies the most prominent feature which is omitted from the simulation: a zonally elongated area of high sea level height variability between 170^oW and 110^oW at around 5-8^oN. Another maximum of variability, located to the northeast of this area, near the Central American coast, is simulated very weakly. Since these areas are associated with the North Equatorial Countercurrent, we compare of the intensity of small-scale sea level height variability with the intensity of the eastward currents.

6 Influence of eastward currents

Since the eastward currents of the tropical Pacific, e.g. North Equatorial Countercurrent (NECC), manifest a prominent seasonality, we overlay them with the patterns of the small-scale sea level height variability on a seasonal basis (Figure 8). The latter is estimated from the T/P data in the same way the estimation of the Figure 3 (upper left panel) was done, except the temporal averaging is done selectively over the seasons: December-January-February (DJF) $\sqrt{\langle \hat{\sigma}^2_{G\times T}\rangle_{\mathrm{DJF}}}$, March-April-May (MAM) $\sqrt{\langle \hat{\sigma}^2_{G\times T}\rangle_{\mathrm{MAM}}}$, etc. For the currents we use the zonal velocity data (7.5m depth) from the *Carton et al.* [2000ab] ocean reanalysis for the period from October 1992 to February 1999, and show only positive (eastward) seasonal means.

Figure 8 shows that the small-scale sea level height variability areas located between 5^oN and 8^oN and omitted in our wind-driven Monte Carlo simulations (Figure 6) belongs to the area of the NECC. They also manifest a similar timing within the seasonal cycle: NECC and small-scale variability in its area both are very weak in the spring (aside from the westernmost part), then pick up strength during the summer, reach their maximum intensity in the fall, and stay strong throughout the winter. Since the areas of high sea surface height variability are near the southern edge of the NECC, this variability might be caused by the horizontal shear between the NECC and the westward South Equatorial Current immediately to the south [*Giese et al.*, 1994].

7 Decomposition of small-scale sea level height variability

The small-scale variability estimate defined by (1) involves both temporal and spatial variability. However, these two can be separated; e.g. by equation (4). We took advantage of the fine resolution ($0.25^o \times 0.25^o \times 10$ days) of the *Ducet et al.* [2000] fields to compute all three terms appearing in (4). The *Ducet et al.* [2000] analysis combines the T/P and ERS-1,2 data and was used for the period from October 1992 to August 2001 within which the data from December 24, 1993 to March 24, 1995 is missing due to the gap in the ERS-1 data. Since there are only 3 temporal samples per month, the temporal variability estimates computed by a discrete version of formula (3) with linear interpolation between samples are multiplied by 3/2 to produce an unbiased estimate [*Mardia et al.*, 1979].

The pattern of $\sigma_{4^o \times 1^o \times 1\,\mathrm{month}}$ computed from *Ducet et al.* [2000] values for each $4^o \times 1^o \times 1\,\mathrm{month}$ bin (Figure 9, upper left panel) is quite similar to the one computed from T/P track data (Figure 3, upper left panel), providing an additional evidence in support of the latter pattern, which was computed from just T/P track values. Comparison of temporal and spatial variability patterns indicates that the variability maxima in the northwestern and southwestern corners of the domain are dominated by spatial variability. In contrast, the eastern Pacific variability band near 10^oN is divided into two parts: the northeastern part near the Central American coast is dominated by the spatial variability, while the band at 5–8^oN between 170^oW and 110^oW predominantly involves temporal variability of the bin spatial means. This division is consistent with the current view of two different variability mechanisms (wind bursts and instability waves) in these two adjacent areas [*Giese et al.*, 1994].

The ratio of temporal to spatial variability (Figure 9, lower right panel) manifests a remarkably zonally uniform meridional variability structure, the main exceptions being a prominent maximum in the mentioned above area (5–8^oN and 170–110^oW) and slightly higher values near coasts. The reason for this structure is that high temporal-to-spatial variability ratio is only produced by long waves with relatively short periods, and such waves propagate predominantly at the low latitudes. To investigate the ratio distribution in a more quantitative fashion, we computed analytically this ratio for a harmonic wave with a period T and a length L (see formula (A1)). Its values $\gamma(T/T_b, L/L_b) = \sigma_{T_b}([s]_{L_b})/\sqrt{[\sigma^2_{L_b}(s)]_{T_b}}$ depend only on relative sizes of T and L with respect to the bin dimensions L_b and T_b. This dependence is shown by colors on the panels of Figure 10.

To attribute parameters of the actual ocean waves to the points of the diagrams in Figure 10, we let $L_b = 4^o$ and $T_b = 1$ month. Only zonally propagating waves are taken into account here, since the meridional propagation is rarely observed [*Killworth et al.*, 1997]. We evaluate the propagation of the first baroclinic mode and assume its velocity $c = 2.5$m/s.

The dispersion relation for Rossby waves zonally propagating at a latitude y can be written through L and T as

$$T = \frac{f(y)^2 L^2/c^2 + 4\pi^2}{\beta(y) L} \tag{10}$$

[*Gill, 1982*]. (Unlike the wavenumber, we take the period L to be positive for waves propagating either to the west or east). These relationships for different latitudes are shown in Figure 10 by thin

solid lines. The points corresponding to the minimum allowable period at every latitude

$$T_{\min}(y) = 4\pi f/\beta c, \quad L_{\min}(y) = 2\pi c/f$$

are marked by black dots with indicated latitudes.

Low latitudes are dominated by equatorially trapped modes with turning latitudes

$$y_T = 180^o(\pi R_e)^{-1}\sqrt{(2n+1)c/\beta},$$

where n indexes the meridional structure of the mode, and R_e is the radius of the Earth. Formula (10) is applicable to trapped modes as well if f is evaluated at the turning latitude y_T and β has the equatorial value (which is very close to $\beta(y_T)$ for low number modes) [*Cane and Sarachik*, 1976]. Thick white lines in Figure 10 show dispersion relations for the first three trapped Rossby modes ($y_T \approx 5.2^o, 6.7^o, 7.9^o$).

Most of other planetary waves, i.e. Poisson (inertia-gravity), Kelvin, and Yanai modes have their dispersion curves in the short-period part of the dispersion diagram in Figure 10 (except for short-length Yanai waves). As in Rossby wave case, the trapped Poisson modes obey the dispersion relationship for the off-equatorial Poisson modes if its coefficients are evaluated at the turning latitudes. Hence the dispersion curves for the first few trapped Poisson modes (not shown) are located in between the two off-equatorial Poisson curves shown for the latitudes of 5^o and 29^o.

Therefore we have two basic types of a dispersion relationship to account for: $T = T_R(L)$ for Rossby waves, defined by (10), and $T = T_{\rm sp} \approx 0$ for all other waves, which are concentrated in the short-period part of the diagram. We use formula (A2) to estimate γ for these wave types. The integrals are evaluated numerically. Based on the study by *Stammer* [1997] we take the power spectral density of tropical sea level heights of the form

$$F(k,\omega) \sim k^{-2.5}\omega^{-0.5}$$

for waves longer than 100 km with periods between 30 and 200 days. For the arguments outside these limits, the wavenumber- and frequency-dependent factors in $F(k,\omega)$ are extended as constants. Since this spectral density strongly favors the long waves, at each latitude we integrate Rossby waves up from $L_{\min}(y)$ along their dispersion curves. Equatorward of the turning latitude of the first equatorial mode (5.2^o) we assume the same dispersion relation as holds at this latitude, thus γ_R is constant between 5.2^oS and 5.2^oN. In integrals of short period waves, 100 km is taken as the lower integration limit, results not being very sensitive to this choice. However, all integrals and their ratios $\gamma(L_{\min}, L_{\max})$ depend strongly on the assumed $L_{\max}$.

This theory is found to fit well observed values of γ if $L_{\max}$ is taken to be $1.9{\cdot}10^3$km within 15^o of the equator , $1.3{\cdot}10^3$km poleward of 25^o, and changing linearly with latitude in between. Resulting dependencies of γ_R and $\gamma_{\rm sp}$ on the lattitude are shown in Figure 11 alongside a latitudinal profile of zonally averaged γ evaluated from observations. The Rossby wave ratio γ_R captures the transition from $\gamma < 1$ (spatial variability dominates) to $\gamma > 1$ (temporal variability dominates) approximately correctly but manifests too steep a change with the latitude. Accounting for the effect of short-period waves makes the latitudinal change more realistic.

If the total energy ratio between Rossby and short-period waves is α, then

$$\gamma = \sqrt{\frac{\alpha^2\gamma_R^2(\gamma_{\rm sp}^2+1)+\gamma_{\rm sp}^2(\gamma_R^2+1)}{\alpha^2(\gamma_{\rm sp}^2+1)+\gamma_R^2+1}} \tag{11}$$

(cf. equation (A3)). Comparison of the variability in pentad (five day periods) averages of $4^o\times4^o$ box means with the day-to-day variability inside these boxes for the reanalysis wind [*Kalnay et al.*, 1996] showed a nearly equal amount of energy in these two types of variability. Because of this one can assume a roughly similar amount of energy in Rossby and short-period waves. Hence we take $\alpha = 1$ in formula (11) and obtain a reasonably good fit to the observed ratios (Figure 11).

The observed peak at around 5^oN is due to the area of instability waves associated with the NECC. These waves are characterized by the longer wavelengths than the waves propagating in the stable conditions have at these latitudes. Similarly, values of γ right on the equator and in the most of the coastal areas outside of the equatorial region (cf. Figure 9, lower left panel) have higher values of γ than their surroundings since the equatorial and coastal Kelvin waves are more likely to have larger lengths than Rossby waves at the same latitude.

8 Discussion

The results presented above gain us some insights into the problems that are common to a wide range of the ocean model simulations and assimilations. They also give us practical constraints on the error in the model and observational sea surface height data sets. This is crucial information for setting up optimal data assimilation procedures.

Small-scale sea level height variability and eddy kinetic energy. The pattern of the T/P small-scale variability (Figure 3, top left panel) is similar to the tropical Pacific portion of what has long been known as a global pattern of mesoscale variability, or eddy energy. With progress in satellite altimetry, techniques for estimating this pattern have been getting more elaborated, and the pattern estimates themselves more detailed. *Cheney et al.* [1983] computed the global mesoscale sea level variability for just one month of SEASAT data, but reproduced the general shape of this pattern (cf. their Fig. 3, p.4345). *Fu et al.* [1988] used the data from one year of the Geosat mission and estimated annual mean eddy energy by computing the standard deviation of the altimetry values made along repeat tracks. The tropical Pacific part of their map (also reproduced by *Tomczak and Godfrey* [1994] in their Figure 4.8) is quite detailed and similar to our small-scale variability map (Figure 3, upper left panel). In the T/P period an explicit computation of the geostrophic eddy kinetic energy (EKE) became the most common way to map the mesoscale variability. Initially the computation was done only for along-track derivatives of altimetric values, under an assumption of isotropy of the turbulent eddy field [*Stammer*, 1997; *Stammer and Wunsch*, 1999]. Later, when merging T/P and ERS-1,2 data allowed a fine gridding of global altimetry (0.25^o spatial and 10 day temporal resolution) , the computation of EKE via both components of the geostrophic velocity has become possible [*Ducet et al.*, 2000]. Comparison with Plates 8 and 10 of *Ducet et al.* [2000] shows that aside from the equator, where the geostrophic velocities tend to infinity, our pattern

of small-scale sea level height variability resembles the structure of the variance in their seasonal cycle of the eddy kinetic energy.

The similarity between a pattern of small-scale variability and of the eddy kinetic energy pattern is interesting. The former is a basic statistical characteristic of a continuous sea level height field, of prime importance for our error analysis; the latter is a basic dynamical characteristic of the ocean, used for describing its mesoscale variability. They are computed differently, but they both reflect the intensity with which nearby values can differ from each other.

Causes for the small-scale variability and its modeling. Connection between the small-scale variability and the EKE pattern suggests that the latter indicates not only *bona fide* eddies, but the variability on all scales and all levels of strength, including linear ocean waves, i.e. perturbations weak enough not to cause non-linear effects to become significant. (Rossby waves are sometimes called "weak" eddies). Maxima in the northwestern and southwestern parts of the tropical Pacific domain are usually viewed as areas of intense eddy activities due to the shear between equatorial currents and subtropical countercurrents. While we do not dismiss this general perception, our Monte Carlo experiments with the linear model suggest that linear waves generated by the month-to-month uncertainty in winds with short spatial decorrelation scale can partially explain (Figure 6, left panels) variability in these regions. At least half of the variability RMS can be obtained as a direct response to the local small-scale variability in the windstress, and this ratio can only increase with the refinement in the model's spatial and temporal resolution. However, a close inspection of the northwestern area (Figure 6, bottom panel) reveals a notable difference between observations and a simulation: a variability maximum at 20^oN is not reproduced by the model, whose variability increases uniformly towards the northwest, as the variance carried by the Rossby waves is getting "piled up" in that direction. The maximum variability which in fact takes place at 20^oN is of seasonal and non-linear nature: instabilities intensify in June and July when the Subtropical Countercurrent strengthens while its density contrast with the surrounding water masses diminishes [*Qui*, 1999; *Ducet et al.*, 2000].

The interpretation of the eastern Pacific variability band is more complicated. In fact, this "band" consists of two distinct pieces (Figure 3): an area to the north of 10^oN from 110^oW to the South American cost, and a narrow band at around 5-8^oN between 165^oW and 110^oW. *Giese et al.* [1994] identified the first piece as the area of strong anticyclonic eddies generated by intense wind bursts across the Gulfs of Tehuantepec and Papagayo. The variability in the second was attributed to the activity of the tropical instability waves which form just south of the NECC. The latter point is explicitly confirmed by our Figure 8: seasonality and positioning of the variability maximum at 5-8^oN corresponds to the areas where horizontal shear between NECC and SEC is likely to be high. This maximum is completely missed in our Monte Carlo experiments, emphasizing the importance of the instability mechanisms compared to the local wind forcing in this area. This area is also identified by a particularly high ratio of temporal over spatial variability (Figure 9, lower right) pointing towards the predominance of very long waves in this area.

Simulation of small-scale variability due to currents' instabilities needs a high resolution non-linear model. Figure 12 overlays contours of geostrophic zonal velocity with the fields of small-scale variability for the 0.25^o resolution POCM 4C model [*Tokmakian and Challenor*, 2001]. All maxima in small-scale variability produced by this model seem to be either in the areas of eastward

currents or near islands. The POCM 4C model successfully produced variability maxima in the areas of NECC and STC. However, the major variability maxima due to the wind in the northwestern and southwestern corners of the domain are almost completely missing. At the same time, some variability is seasonally produced in the area of NECC termination, north of 10^oN from 110^oW to the South American cost. In our Monte Carlo experiments (Figure 6) this variability maximum is modeled very weakly. One can expect this if it is the intense eddy-generating windbursts that are responsible for the most of variability. On the other hand, the POCM 4C simulation suggests that the instability mechanisms could be contributing to this variability as well. If true, this would mean that both possible mechanisms discussed by *Giese et al.* [1994] (windburst-caused eddies and instability waves) along with a smaller amplitude wind noise which causes a linear wave activity contribute to the small-scale sea level height variability in this area. Applicability of a few different mechanisms also helps to maintain this variability at some level of intensity all year around (Figure 8), while individual mechanisms have a substantial seasonality.

Error in model products. A large part of the variability responding to small-scale wind forcing can be modeled linearly (Figure 6), even though an eddy resolving model would be needed to simulate the response in its complete intensity. Simulation of the current instabilities requires a non-linear model. POCM 4C (Figure 12) and other ocean GCMs seem to be dealing quite well with this type of variability, e.g. note a complete lack of error maximum there in the NCEP assimilation run (Figure 2).

Small-scale variability in the model sea level – and in nature – depends on the smoothness of the wind field (cf. left and right top panels of Figure 7). High resolution models underestimate the small-scale sea level height variability, even when driven by winds with adequate representations of short-term and small-scale variability. For example, POCM 4C run driven by daily fluxes from the ECMWF almost completely omits areas of wind-driven small-scale variability in the northwestern and southwestern tropical Pacific – only variability due to currents' instability gets simulated (Figure 12). This underestimation is most likely due to the friction parameterization in ocean GCMs, which is made too large in order to achieve numerical stability.

Underestimation of the small-scale variability in the areas where it is high results in underestimation of interannual anomalies in grid size of $4^o \times 1^o$ as well, as Figure 13 demonstrates. Areas of high variability poleward of 20^o in the tropical western Pacific in the T/P panel of Figure 13 are missing in its other panels. Not producing variability at all, of course, results in the error, and this error type is particularly difficult to correct by the data assimilation procedures: depending on the assimilation technique, data corrections in such areas either will just attenuate in time, or won't be allowed at all.

In other words, a pattern of the error in an assimilation product includes all areas where the signal variability is undersimulated or misrepresented because of the errors in the forcing or inadequacy of the parameterization of subgrid processes. These areas include most of the places where the small-scale variability in the ocean is high. Simulation error patterns additionally include areas where the signal variability is high, since such is the pattern of the model response to error in the forcing with long (signal-type) decorrelation scale, which are often present in the wind products as well. For example, simulation error for the linear model (Figure 1, upper left) contains the features of the error patterns responding to the random wind fields with both short and long decorrelation

scales (Figure 7). *Fukumori et al.* [1999] assumed that the wind error had a spatial covariance proportional to the covariance of the observed winds and a short temporal decorrelation scale. The model error obtained in response to such a wind (their plate 1) has some features, e.g. equatorial maximum in the tropical Pacific, which resemble the signal variance (cf. Figure 13, botom right).

Error in the gridded altimetry fields. The error map for the altimetry product by *Cheney et al.* [1994] (Figure 3, bottom left panel) was developed here on the basis of the small-scale variability estimates and only accounts for the sampling error in gridbox averages. Yet the comparison with another satellite product, which uses more data [*Ducet et al.*, 2000] (Figure 3, bottom right panel), and with in situ (tide gauges) data (Figure 4, lower left panel) suggests that the underestimation is not too severe, no larger than 50% of the estimated error RMS. In other words, sampling error is a dominant term in this altimetry product for interannual anomalies. Gridded altimetry error thus have many features of the small-scale variability pattern, the same pattern which influences the model error as well.

Our error map resembles the tropical Pacific part of the T/P error map produced in a different way by *Fukumori et al.* [1999] (their plate 1a). It is also somewhat similar to the T/P error field computed by *Tokmakian and Challenor* [1999], but is less noisy and by construction has no negative variance estimates.

9 Conclusions

T/P altimetry data verify that analyses of the tropical Pacific sea level height anomaly based on assimilation of in situ observations into a simple linear model are reasonably accurate over most of the domain from 20^oS to 20^oN. The exceptions are the northwestern and southwestern corners of the domain, as well as a band around 10^oN in the eastern part of the domain. A surprisingly similar pattern of differences was found in the comparison of a wide variety of the GCM simulation and assimilation products with the T/P altimetry fields.

This pattern was identified as the pattern of high small-scale variability in the tropical sea level height. Accordingly, the gridding error in the T/P altimetry product we used has a similar structure. However, since it is smaller in size, we conclude that the model assimilation error must have the similar structure as well. The explanation is that (1) most of the features of this pattern of variability can be caused by the small-scale variability (and error) in wind, and (2) most GCMs underestimate small-scale variability even if driven by the wind forcing with enough of such variability (possibly because of dissipation schemes that overdamp small scales) and, as a result, underestimate variability on assimilated scales as well; and (3) areas of high error which are not associated with local small-scale wind variability are those of high shear and current instabilities in the ocean, and have their own host of simulation problems.

Systematic underestimation of variance by the ocean models should be taken into account by data assimilation systems. Comparison of the top panels of Figure 13 with the left panels of Figure 7 suggest a possible solution: to evaluate model error covariance from the run forced by the observed wind which is perturbed by a specially designed short-scale noise. Covariance estimated that way will allow large data-driven corrections in the areas where noise forcing can drive variance

high.

The structure of the noise used for error modeling matters a great deal. Figure 7 shows two types of error patterns: one, the response to the small-scale noise, is typical of the analyzed fields and another plagues a model response to the wind observations affected by a large-scale error.

It is of great use for El Niño predictions to be able to improve the hindcasts of the tropical Pacific sea level heights anomalies [*Cañizares et al.* 2001]. This goal can be achieved via optimal tuning of data assimilation system for the tropical Pacific Ocean on the basis of the satellite period, when detailed information on the small-scale behavior of the tropical Pacific ocean-atmosphere system is available (e.g. Figures 3 and 5).

Acknowledgments. This work was supported by the NOAA through grants NA86GP0515, NA06GP0414, SIO #10130, by the NASA through the IDS grant NAG5-4058 and its renewal, and by NOAA/NASA Enhanced Data Set Project grant NA06GP0567. AK is grateful to Jim Carton for his insights into the sources of the eastern Pacific sea level height variability, and to Gilles Reverdin, Bob Miller, Detlef Stammer, Carl Wunsch, Warren White, Peter Challenor, Jochem Marotzke, and Dudley Chelton for helpful discussions at various stages of this work. The hospitality of the Institute for Mathematics and its Applications in University of Minnesota during AK's April 2002 visit and discussions with participants of their data assimilation workshop are gratefully acknowledged. We are indebted to Robin Tokmakian, Ichiro Fukumori, Pierre-Yves Le Traon, Jim Carton and Gennadiy Chepurin, Ming Ji and David Behringer for making their model runs, altimetry analyses, and pre-publication manuscripts available to us. This work is a Lamont–Doherty Earth Observatory contribution number xxxx.

Appendix A: Ratio of temporal and spatial variability

Individual waves. Suppose a wave of a unit amplitude

$$s = e^{i(kx-\omega t)}$$

with a wavenumber $k = 2\pi/L$ and frequency $\omega = 2\pi/T$ propagates through a space-time bin with dimensions L_b and T_b respectively. It is easy to calculate the following statistics (notation as in section 2):

$$[s]_{L_b} = \frac{1}{L_b}\int_0^{L_b} e^{i(kx-\omega t)}dx = e^{-i\omega t}\frac{e^{ikL_b}-1}{ikL_b};$$

$$[s]_{L_b\times T_b} = \frac{1}{T_b}\int_0^{T_b}[s]_{L_b}dt = \frac{1}{kL_b\omega T_b}(1-e^{ikL_b})(1-e^{-iwT_b}).$$

$$\sigma^2_{L_b}(s) = [|s|^2]_{L_b} - |[s]_{L_b}|^2 = 1 - |\frac{e^{ikL_b}-1}{kL_b}|^2 = 1 - \frac{2(1-\cos kL_b)}{(kL_b)^2.}$$

Since the last expression does not depend on t,

$$[\sigma^2_{L_b}(s)]_{T_b} = \sigma^2_{L_b}(s) = 1 - \frac{2(1-\cos kL_b)}{(kL_b)^2}$$

as well. Finally,

$$\sigma^2_{T_b}([s]_{L_b}) = [|[s]_{L_b}|^2]_{T_b} - |[s]_{L_b\times T_b}|^2 = \frac{2(1-\cos kL_b)}{(kL_b)^2} - \frac{4(1-\cos kL_b)(1-\cos\omega T_b)}{(kL_b\omega T_b)^2}.$$

Since $kL_b = 2\pi L_b/L$ and $\omega T_b = 2\pi T_b/T$, all variabilities computed above depend only on the ratios

$$\lambda = L/L_b, \qquad \tau = T/T_b,$$

so that the temporal-to-spatial variability ratio is equal to

$$\gamma(\tau,\lambda) = \sqrt{\sigma^2_{T_b}([s]_{L_b})/[\sigma^2_{L_b}(s)]_{T_b}} = \left[\frac{\mathcal{C}(\lambda)(1-\mathcal{C}(\tau))}{1-\mathcal{C}(\lambda)}\right]^{1/2}, \tag{A1}$$

where we introduced

$$\mathcal{C}(\lambda) = \frac{2(1-\cos 2\pi/\lambda)}{(2\pi/\lambda)^2}.$$

Note that $\mathcal{C}(0) = 0$ and $\mathcal{C}(\infty) = 1$. Values of γ are shown in color in both panels of Figure 10.

Ratio for a given dispersion relation. We are interested in in estimating γ for wave packets obeying a dispersion relation $T = T_D(L)$ in the wave length interval from L_1 to L_2. Given the ratio γ, a unit of energy is partitioned into a temporal $\gamma^2(\gamma^2+1)^{-1}$ and a spatial $(\gamma^2+1)^{-1}$

pieces. Assuming power spectrum density $F(k,\omega)$ we can compute the variability ratio for the entire packet by

$$\gamma_D(L_1,L_2)=\sqrt{\frac{\int_{L_1}^{L_2}\gamma(\frac{T_D(L)}{T_b},\frac{L}{L_b})^2\left(\gamma(\frac{T_D(L)}{T_b},\frac{L}{L_b})^2+1\right)^{-1}F(\frac{2\pi}{L},\frac{2\pi}{T_D(L)})L^{-2}dL}{\int_{L_1}^{L_2}\left(\gamma(\frac{T_D(L)}{T_b},\frac{L}{L_b})^2+1\right)^{-1}F(\frac{2\pi}{L},\frac{2\pi}{T_D(L)})L^{-2}dL}}. \tag{A2}$$

Combining two distributions. To compute γ in the presence of two basic wave types with individual ratios γ_1 and γ_2, we have to assume a total variability ratio between these two wave types:

$$\alpha=\sigma_{L_b\times T_b}(s_1)/\sigma_{L_b\times T_b}(s_2).$$

Then a unit of energy is partitioned between these two wavetypes as $\alpha^2/(\alpha^2+1)$ and $1/(\alpha^2+1)$, hence

$$\gamma_{1+2}^2=\frac{\frac{\alpha^2}{\alpha^2+1}\frac{\gamma_1^2}{\gamma_1^2+1}+\frac{1}{\alpha^2+1}\frac{\gamma_2^2}{\gamma_2^2+1}}{\frac{\alpha^2}{\alpha^2+1}\frac{1}{\gamma_1^2+1}+\frac{1}{\alpha^2+1}\frac{1}{\gamma_2^2+1}}=\frac{\alpha^2\gamma_1^2(\gamma_2^2+1)+\gamma_2^2(\gamma_1^2+1)}{\alpha^2(\gamma_2^2+1)+\gamma_1^2+1}. \tag{A3}$$

References

Behringer, D. W., M. Ji, and A. Leetmaa, 1998: An improved coupled model for ENSO prediction and implications for ocean initialization. Part I: The ocean data assimilation system. *Mon. Wea. Rev.*, **126**, 1013-1021.

Cane, M.A. and E.S. Sarachik, 1976: Forced baroclinic ocean motions. I. The linear equatorial unbounded case. *J. Mar. Res.*, **34**, 629-665.

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

Cane, M.A., S.E. Zebiak and S. Dolan, 1986: Experimental forecasts of El Niño. *Nature*, **322**, 827-832.

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

Cañizares, R., A. Kaplan, M.A. Cane, D. Chen, S.E. Zebiak, 2001: Use of data assimilation via linear low order models for the initialization of ENSO predictions. *J. Geophys. Res.*, **106**, 30947–30959.

Carton, J.A., G. Chepurin, X. Cao, and B.S. Giese, 2000a: A Simple Ocean Data Assimilation analysis of the global upper ocean 1950-1995, Part 1: methodology, *J. Phys. Oceanogr.*, **30**, 294-309.

Carton, J.A., G. Chepurin, and X. Cao, 2000b: A Simple Ocean Data Assimilation analysis of the global upper ocean 1950-1995 Part 2: results, *J. Phys. Oceanogr.*, **30**, 311-326.

Chen, D., M.A. Cane, S.E. Zebiak, and A. Kaplan, 1998: The impact of sea level assimilation on the Lamont model prediction of the 1997/1998 El Niño, 1998: *Geophys. Res. Lett.*, **25**, 2837-2840.

Chen, D., M.A. Cane, S.E. Zebiak, R. Cañizares, and A. Kaplan, 2000: Bias correction of an ocean-atmosphere coupled model. *Geophys. Res. Lett.*, **27**, 2585-2588.

Cheney, R.E., J.G.Marsh, and B.D. Beckley, 1983: Global mesoscale variability from collinear tracks of Seasat altimetry data. *J. Geophys. Res.*, **88**,4343-4354.

Cheney, R., L. Miller, R. Agreen, N. Doyle, and J. Lillibridge, 1994: TOPEX/POSEIDON: The 2-cm Solution. *J. Geophys. Res.*, **99**, 24,555–24,564.

Derber, J. and A. Rosati, 1989: A global oceanic data assimilation system. *J. Phys. Oceanogr.*, **19**, 1333-1347.

Ducet, N., P.-Y. Le Traon, and G. Reverdin, 2000: Global high-resolution mapping of ocean circulation from TOPEX/Poseidon and ERS-1 and -2. *J. Geophys. Res.*, **105**, 19,477-19,498.

Fu, L.-L., D.B.Chelton, and V.Zlotnicki, 1988: Satellite altimetry: observing ocean variability from space. *Oceanography*, **1**(2), 4-11 and 58.

Fukumori, I., R.Raghunath, L.-L. Fu, and Y.Chao, 1999: Assimilation of
TOPEX/POSEIDON altimeter data into a global ocean circulation model: How good are the results? *J. Geophys. Res.*, **104**, 25,647-23,665.

Gill, A., 1982: *Atmosphere–Ocean Dynamics*, Academic, San Diego, Calif., 1982.

Giese, B.S., J.A. Carton, and L.J. Holl, 1994: Sea level variability in the eastern tropical pacific

as observed by TOPEX and Tropical Ocean – Global Atmosphere Tropical Atmosphere – Ocean Experiment, *J. Geophys. Res.*, **99**, 24,739-24,748.

JPL, 1998: NASA Scatterometer, Science Data Product (NSCAT-2) User's Manual: Overview and Geophysical Data Products (D-12985). Manual version 1.2, February 1998, Jet Propulsion Laboratory.

Kalnay, E. et al., 1996: The NCEP/NCAR 40-year reanalysis project. *Bull. Amer. Meteor. Soc.*, **77**, 437-471.

Kaplan, A., Y. Kushnir, M. Cane, and M. Blumenthal, 1997: Reduced space optimal analysis for historical datasets: 136 years of Atlantic sea surface temperatures, *J. Geophys. Res.*, **102**, 27835–27860.

Killworth, P.D., D.B.Chelton, R.A. de Szoeke, 1997: The speed of observed and theoretical long extratropical planetary waves. *J. Phys. Oceanogr.*, **27**, 1946-1966.

Mardia, K.V., J.T. Kent, and J.M. Bibby, 1979: *Multivariate Analysis.* Academic Press, Inc., New York. 521 pp.

Miller, R.N., and M.A. Cane, 1989: A Kalman filter analysis of sea level heights in the tropical Pacific, *J. Phys. Oceanogr.*, **19**, 773-990.

Mitchum, G., 1994: Comparison of TOPEX sea surface heights and tide gauge sea levels. *J. Geophys. Res.*, **99**, 24,541-24,553.

Peixoto, J.P., and A.H. Oort, 1995: *Physics of Climate*, American Institute of Physics, New York. 520pp.

Qiu, B., 1999: Seasonal eddy field modulation of the North Pacific Subtropical Countercurrent: TOPEX/Poseidon observations and theory, *J. Phys. Oceanogr.*, , **29**, 2471-2486.

Reverdin, G., A. Kaplan, M. Cane, 1996: Sea level from temperature profiles in the Tropical Pacific Ocean 1975–1992. *J. Geophys. Res.*, **101**, 18,105-18,119.

Schneider, E.K., B.P.Kirtman, Y.Fan, Z. Zhu, 2001: Retrospective ENSO forecasts: The effect of ocean resolution. COLA Tech. Rep. 109, 27pp.

Schneider, E.K., D.G.DeWitt, A.Rosati, B.P.Kirtman, L.Ji, J.J.Tribbia, 2002: Retrospective ENSO forecasts: Sensitivity to atmospheric model and ocean resolution, *Mon. Weather Rev.*, submitted.

Semtner, A.J., and R.M. Chervin, 1992: Ocean general circulation from a global eddy-resolving model. *J. Geophys. Res.*, **97**, 5493-5550.

Smith, T.M., 2000: Tropical Pacific sea level variations (1948-98). *J. Clim.*, **13**, 2757-2769.

Stammer D., 1997: Global characteristics of ocean variability estimated from regional TOPEX/Poseidon altimeter measurements, *J. Phys. Oceanogr.*, , **27**, 1743-1769.

Stammer D. and C.Wunsch, 1999: Temporal changes in eddy energy of the oceans, *Deep Sea Res., II*, **46**, 77-108.

Starr, V.P. and R.M. White, 1954: Balance requirements of the general circulation. Air Force Cambridge Research Center, Geophys. Res. Papers, **35**, 57pp.

Stricherz, J.N., D.M. Legler, and J.J. O'Brien, 1997: TOGA pseudostress atlas, 1985-1994: II, Tropical Pacific Ocean. COAPS Rep. 97-2, 177pp.

Tokmakian, R.T., and P.G. Challenor, 1999: On the joint estimation of model and satellite sea surface height errors. *Ocean Modelling*, **1**, 39-52.

Woodruff, S.D., H.F. Diaz, J.D. Elms, and S.J. Worley, 1998: COADS Release 2 data and metadata enhancements for improvements of marine surface flux fields. *Phys. Chem. Earth*, **23**, 517-526.

Zebiak, S.E., and M.A. Cane, 1987: A model El Niño Southern Oscillation. *Mon. Wea. Rev.*, **115**, 2,262-2,278.

List of Figures

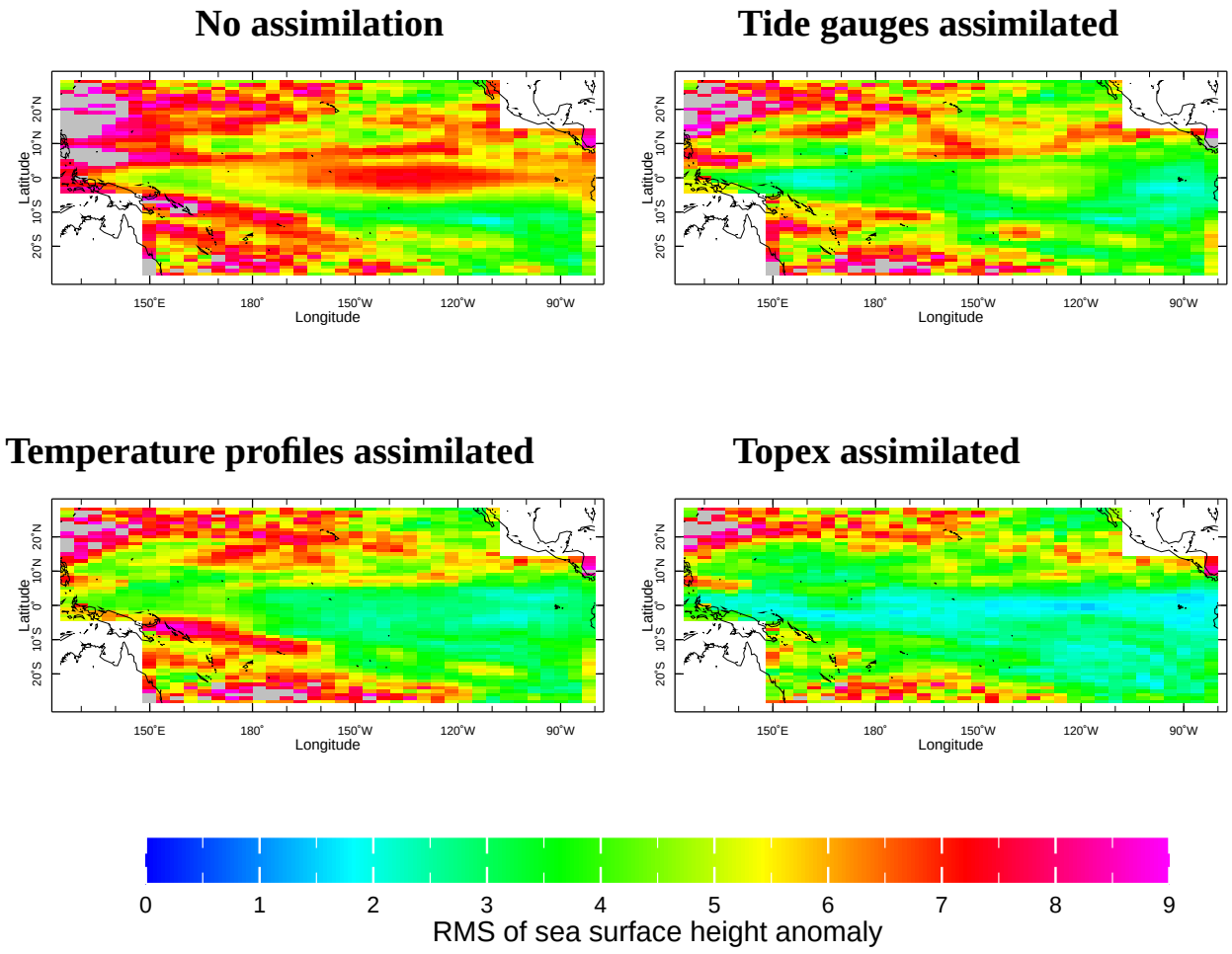


Figure 1: Comparison with the Topex/Poseidon altimetry [*Cheney et al.*, 1994] of sea level height anomalies from a linear anomaly model for the tropical Pacific Ocean [*Cane and Paton*, 1984]. Shown are 1993-2000 RMS differences for four model products: a run with no sea level height assimilation at all, and three optimal smoother products assimilating either 34 tide gauge records, or dynamic heights from the data set of temperature profiles (XBT,BT,TAO), or the Topex altimetry. In all cases the model is forced by the monthly pseudostress anomaly from the FSU [*Stricherz et al.*, 1997].

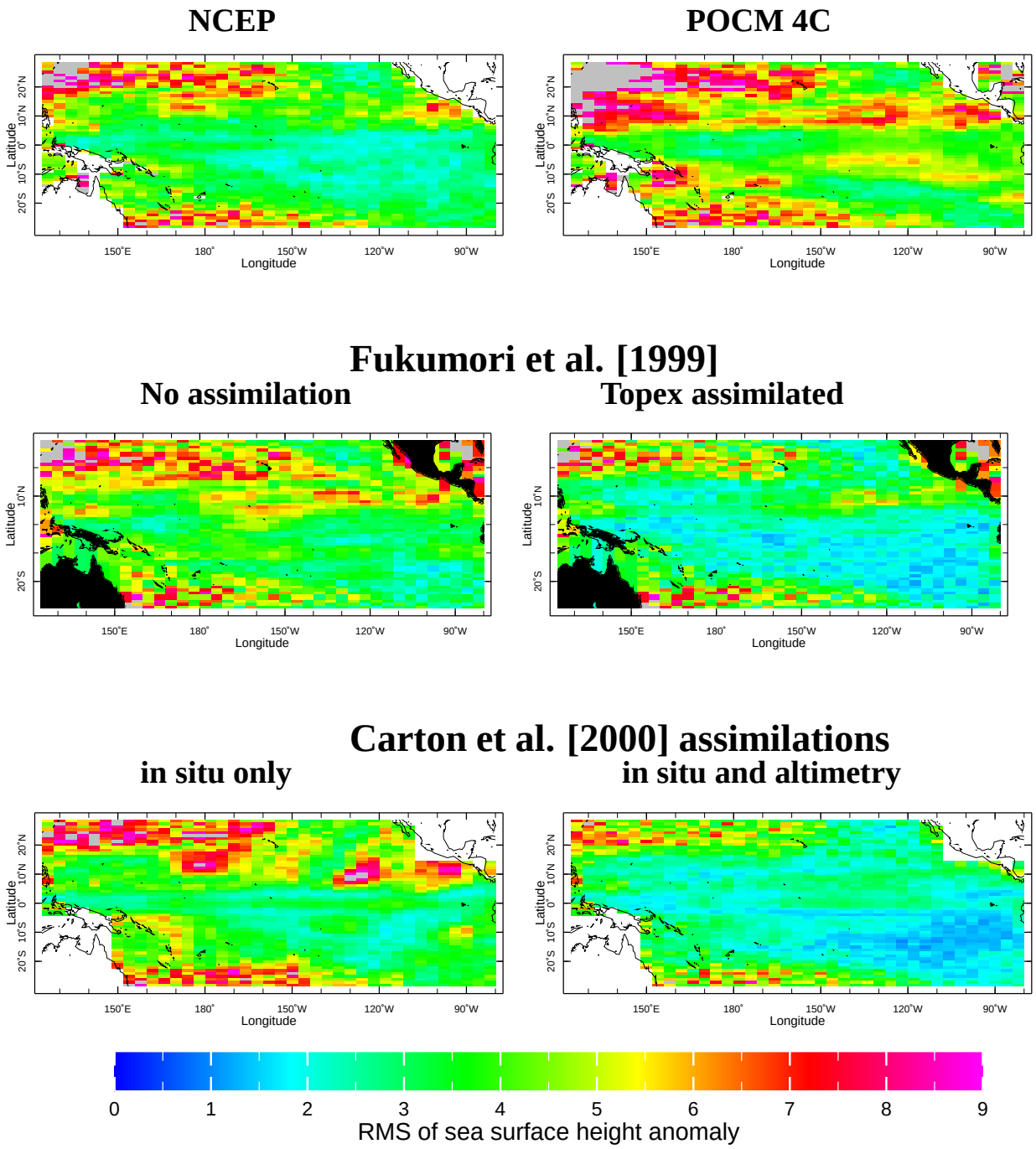


Figure 2: Comparison with the Topex/Poseidon altimetry [*Cheney et al.*, 1994] of sea level height anomalies from six GCM products. Shown are RMS differences for: the NCEP MOM1 version for the Pacific Ocean, assimilating temperature profiles and Topex altimetry [*Behringer et al.*, 1998] (1993-2000); 1/4° simulation from the version POCM 4C of Semtner/Chervin model [*Tokmakian and Challenor,* 1999] (1993-1998); simulation and assimilation of altimetry results from the MOM1 version by *Fukumori et al* [1999] (1993-1995); and assimilations into MOM1 by *Carton et al.* [2000ab].

Small-scale variability and gridded altimetry error

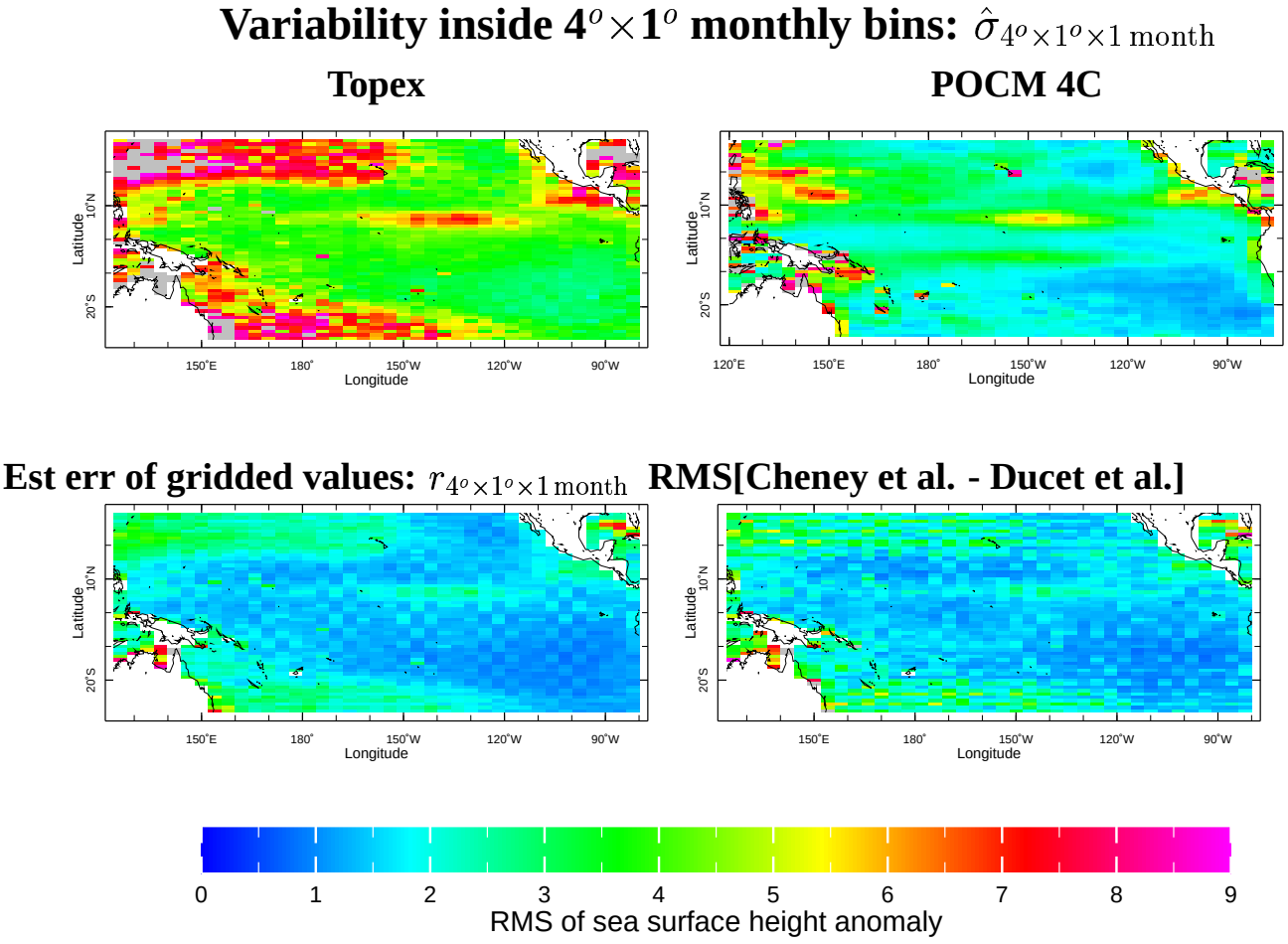


Figure 3: Small-scale, short-term variability and resulting error in sea level height anomaly, cm. Shown are: variability RMS from (top left) the T/P track data and (top right) a high-resolution model (POCM 4C); (lower left) derived error RMS estimates for the gridded *Cheney et al.* [1994] T/P fields; (lower right) RMS difference of two altimetry products.

Validation of Error Estimates

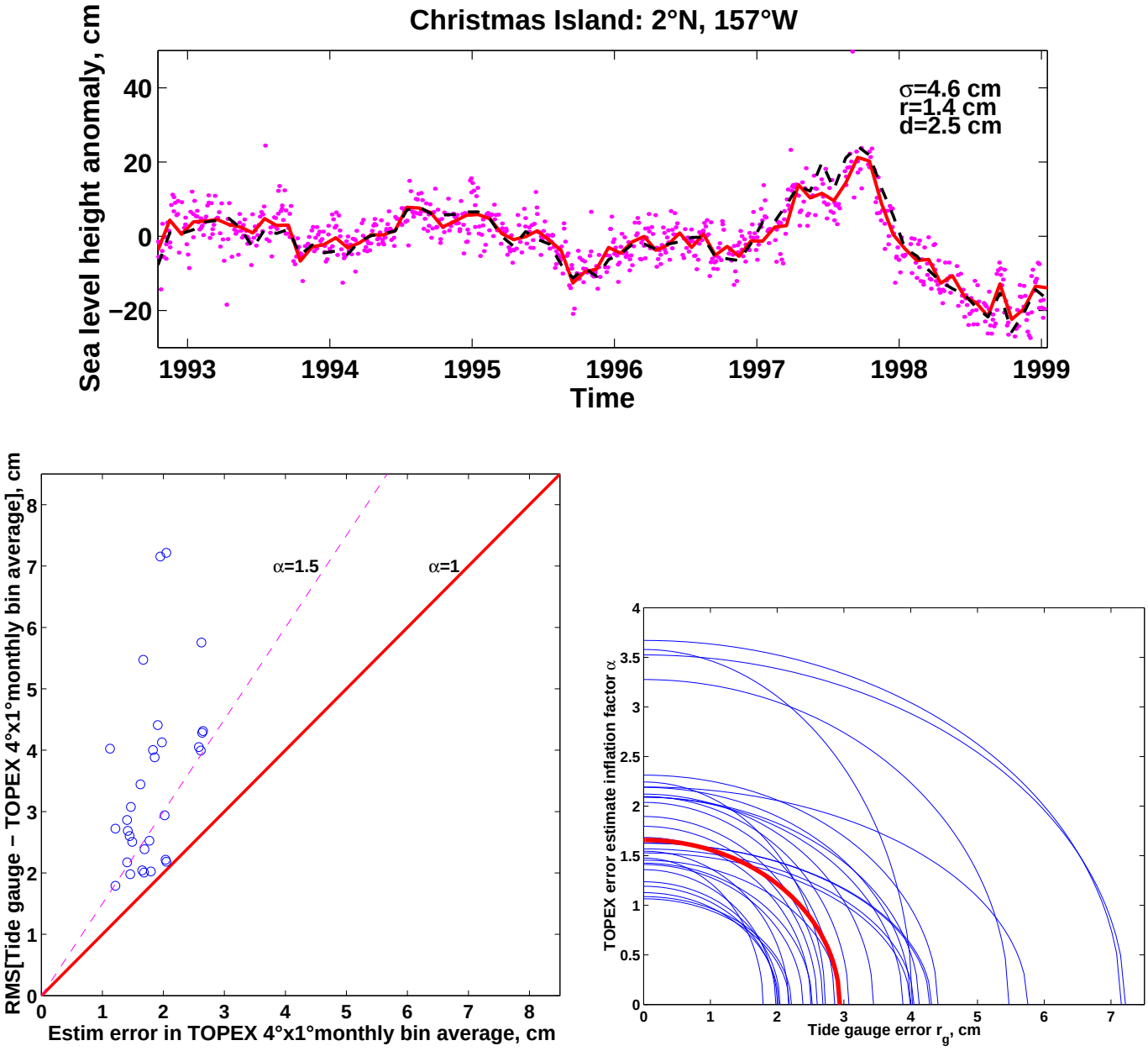


Figure 4: Validation of T/P error estimates by comparison with the tide gauge records, Oct 1992 – Mar 2001. The top panel compares monthly tide gauge sea level height anomalies at Christmas Island (dashes) with altimetric measurements from the corresponding grid box (centered at 2^oN and 158^oW) of the *Cheney et al.* [1994] T/P product. Dots show values from individual altimetry passes, and the solid line shows monthly averages for the grid box. Temporal RMS values of the intrabox variability σ inside this grid box, the sampling error estimate r, and the RMS difference between the grid box and tide gauge monthly means d are indicated as well. In the lower left panel, circles are differences between 31 tide gauges and T/P bins. Differences would fall along the solid line if the only errors were the "optimistic" estimate of T/P errors. The dashed line inflates these optimistic estimates by a factor of 1.5. In the lower right panel, thin lines show constraints on the inflation factor α and tide gauge error r_g imposed by Equation (9) for individual tide gauges. The thick line shows the median constraint.

Small-scale variability in zonal wind $\sigma_{4^o \times 4^o \times 1\ month}$, m/s

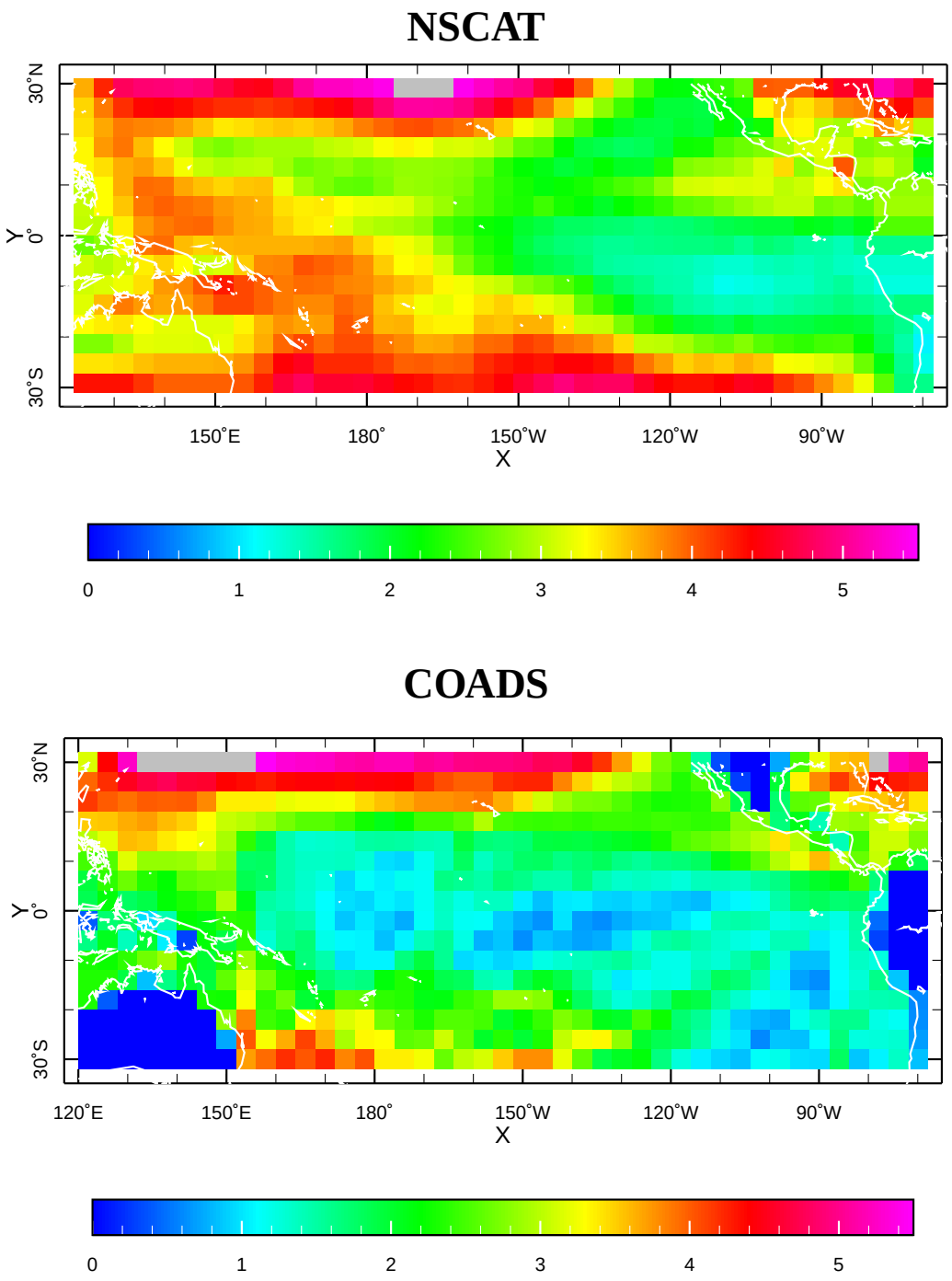
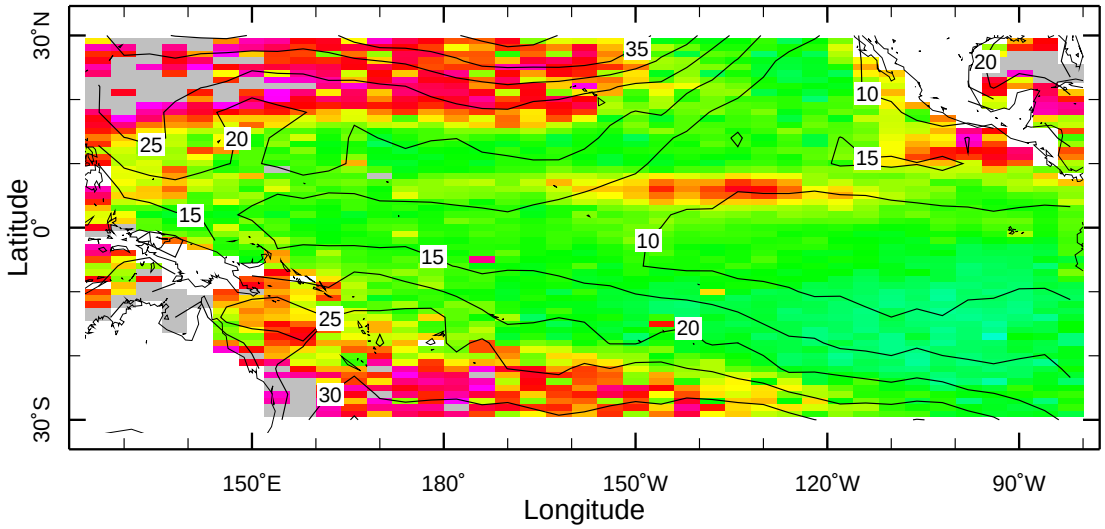


Figure 5: Small-scale and short-term variability for the surface zonal wind estimated from two different sources: (top) satellite scatterometry data and (bottom) ship records from COADS.

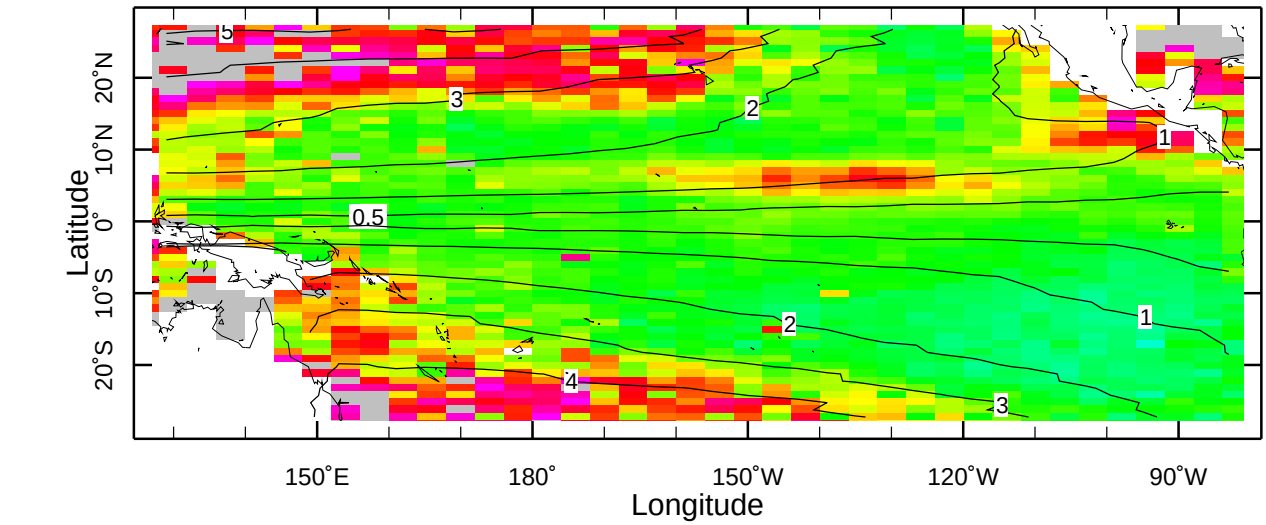
Small-scale variability in wind and sea level height

Zonal pseudostress and sea level height observations



10 m_above_gnd

Sea level height simulation and observations



0.0

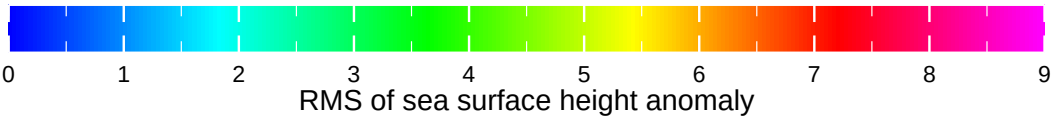
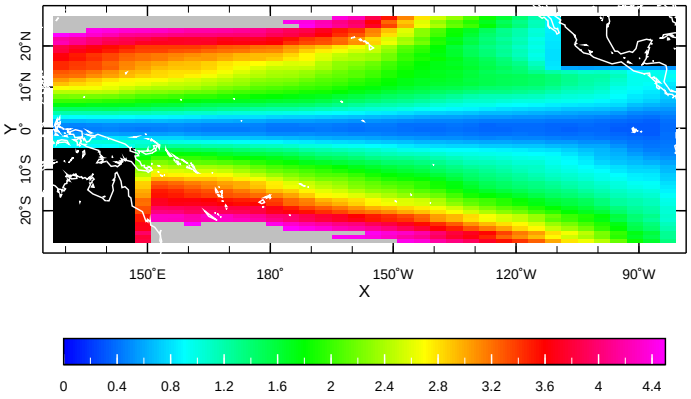


Figure 6: Small-scale variability in pseudostress and sea level height responce: (top) Contours of small-scale variability RMS in zonal wind pseudostress (m^2/s^2 is shown over the color pattern of Figure 3 (upper left); (bottom) same as above but for contours of small-scale variability RMS in the sea level height response of a linear model to the random wind with $1o$ spatial and 0.25 month temporal decorrelation scales. See text for explanations.

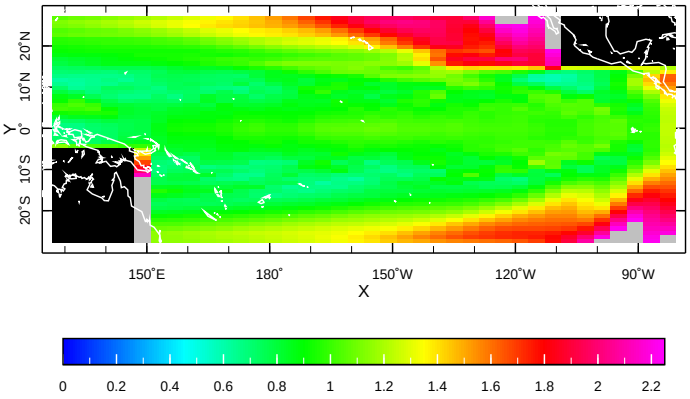
Monte Carlo experiments with a linear model

Variability inside $4^o \times 1^o$ monthly bins: $\sqrt{\left\langle \sigma^2_{4^o \times 1^o \times 1 \text{ month}}(s) \right\rangle_{\text{months}}}$

$1^o \times 1^o$ scaled noise

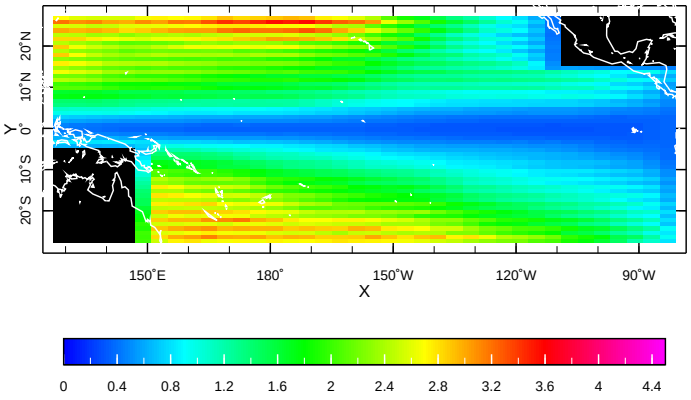


$20^o \times 10^o$ scaled noise



Variability of $4^o \times 1^o$ monthly means: $\sigma_{\text{months}}\left([s]_{4^o \times 1^o \times 1 \text{ month}} \right)$

$1^o \times 1^o$ scaled noise



$20^o \times 10^o$ scaled noise

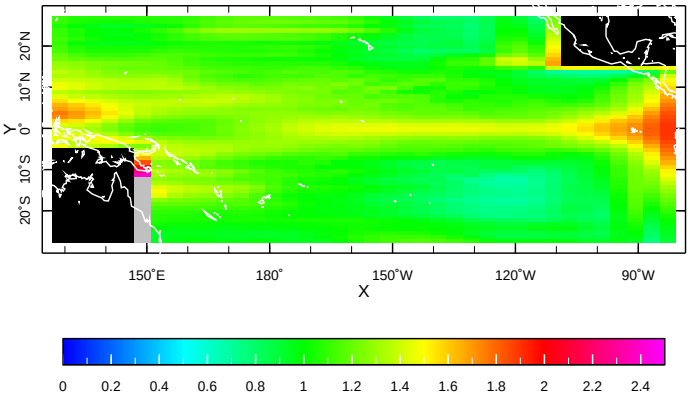


Figure 7: Simulation of sea level height error and small-scale variability in Monte Carlo experiments with a linear model forced by noise designed to imitate errors in the wind forcing. Shown are model responses to the noise forcings with short ($L_x = L_y = 1^o$) and relatively long ($L_x = 20^o,\ L_y = 10^o$) spatial decorrelation scales. Temporal decorrelation scale is 0.25 month. See text for the explanations.

Eastward zonal velocity over small-scale sea level height variability

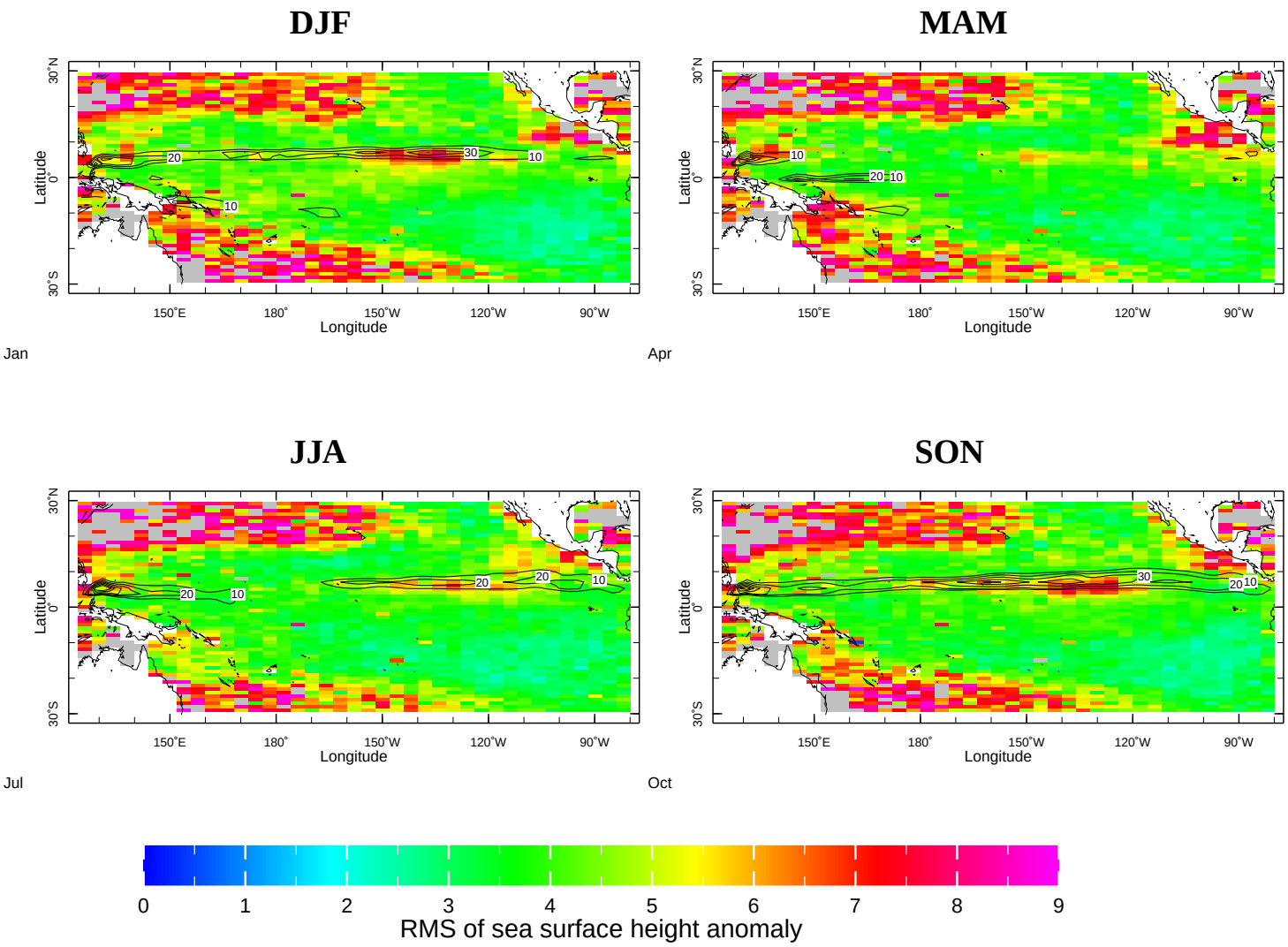


Figure 8: Seasonal means of zonal velocity [*Carton et al.*, 2000ab], cm/s (contours) is shown over seasonal small-scale sea level height variability estimates (colors). Only positive (eastward) velocity values are contoured.

Time-space separation of small-scale sea level height variability

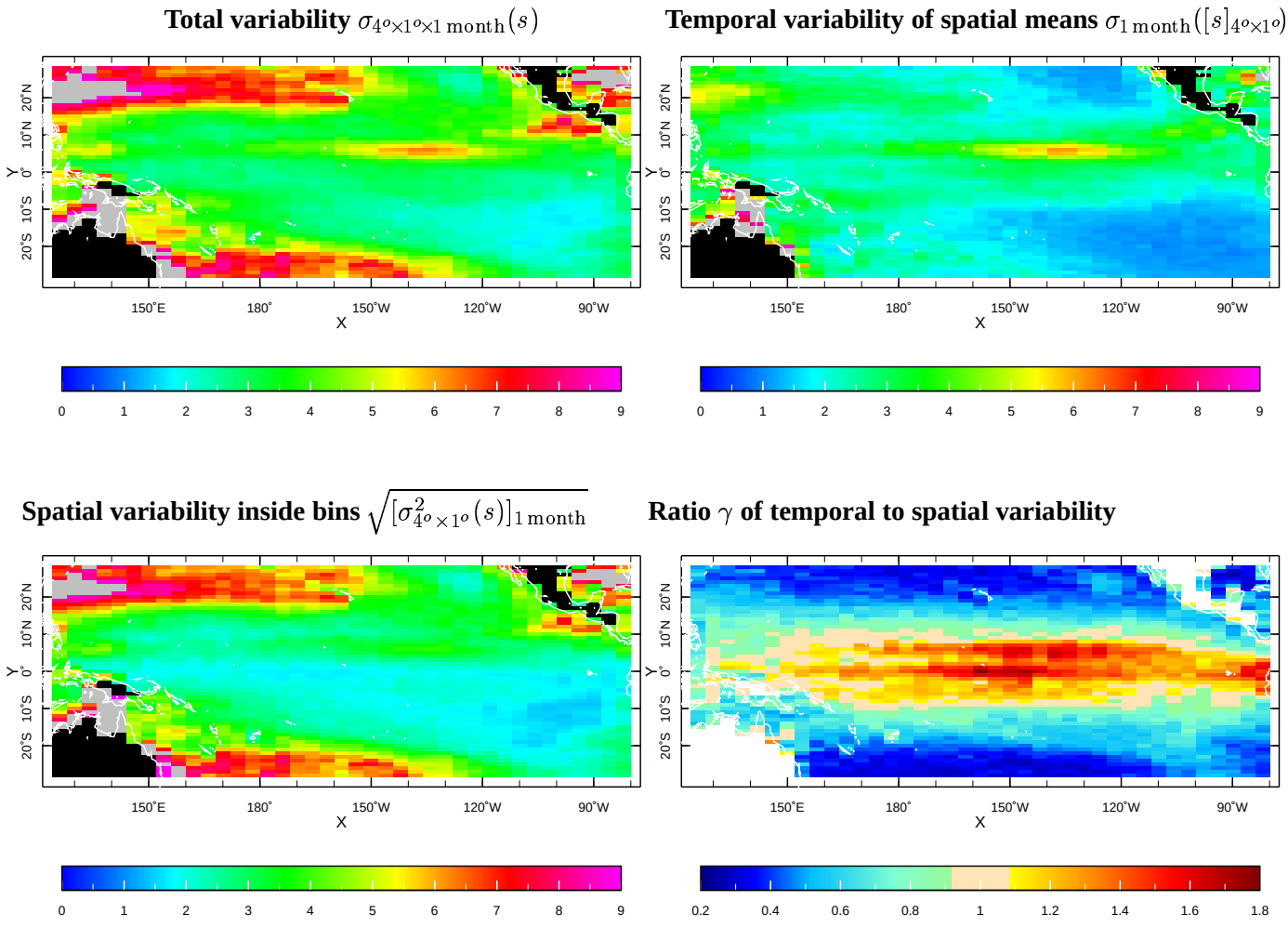


Figure 9: Separation of space-time small-scale sea level height variability into temporal and spatial components for *Ducet et al.* [2000] 0.25^o resolution 10 day gridded altimetry fields. See text for the explanations.

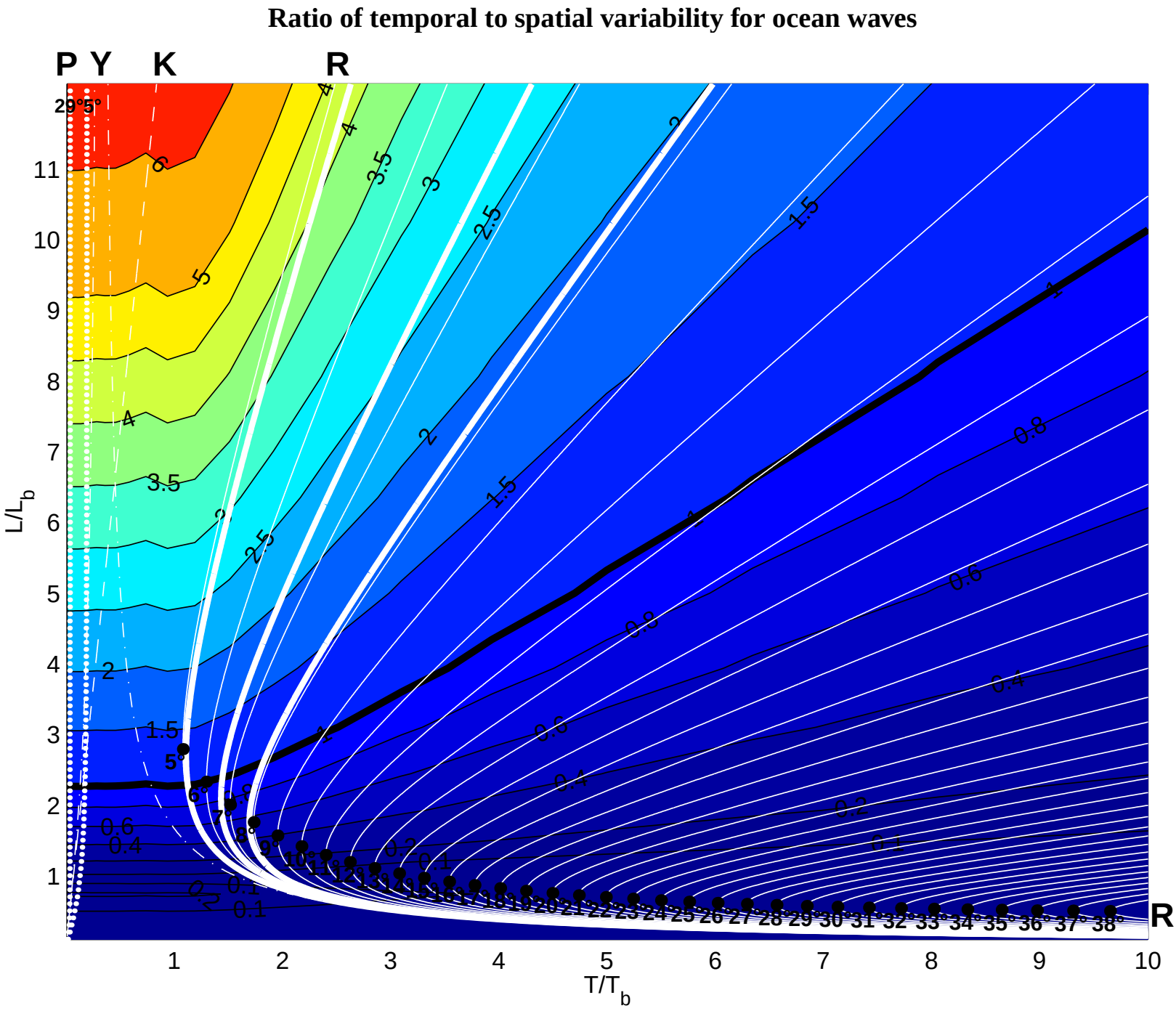


Figure 10: Ratios γ of temporal to spatial variability for ocean waves. Colors show the ratio for a monochromatic harmonic wave with a wavelength L and a period T. White lines show dispersion relations for ocean waves. Solid lines indicate Rossby (R) waves. Thin lines show off-equatorial Rossby waves for different latitudes indicated at black circles that mark points with the minimum allowable wave period for each latitude. Thick lines show the first 3 trapped equatorial Rossby modes. Dashes and dash-points show equatorial Kelvin (K) and Yanai (Y) waves respectively. White dots indicate Poisson (P) waves for the latitudes of 5^o and 29^o. In order to put wave parameters on the dimensionless color diagram, T_b =1 month and L_b =4^o are assumed.

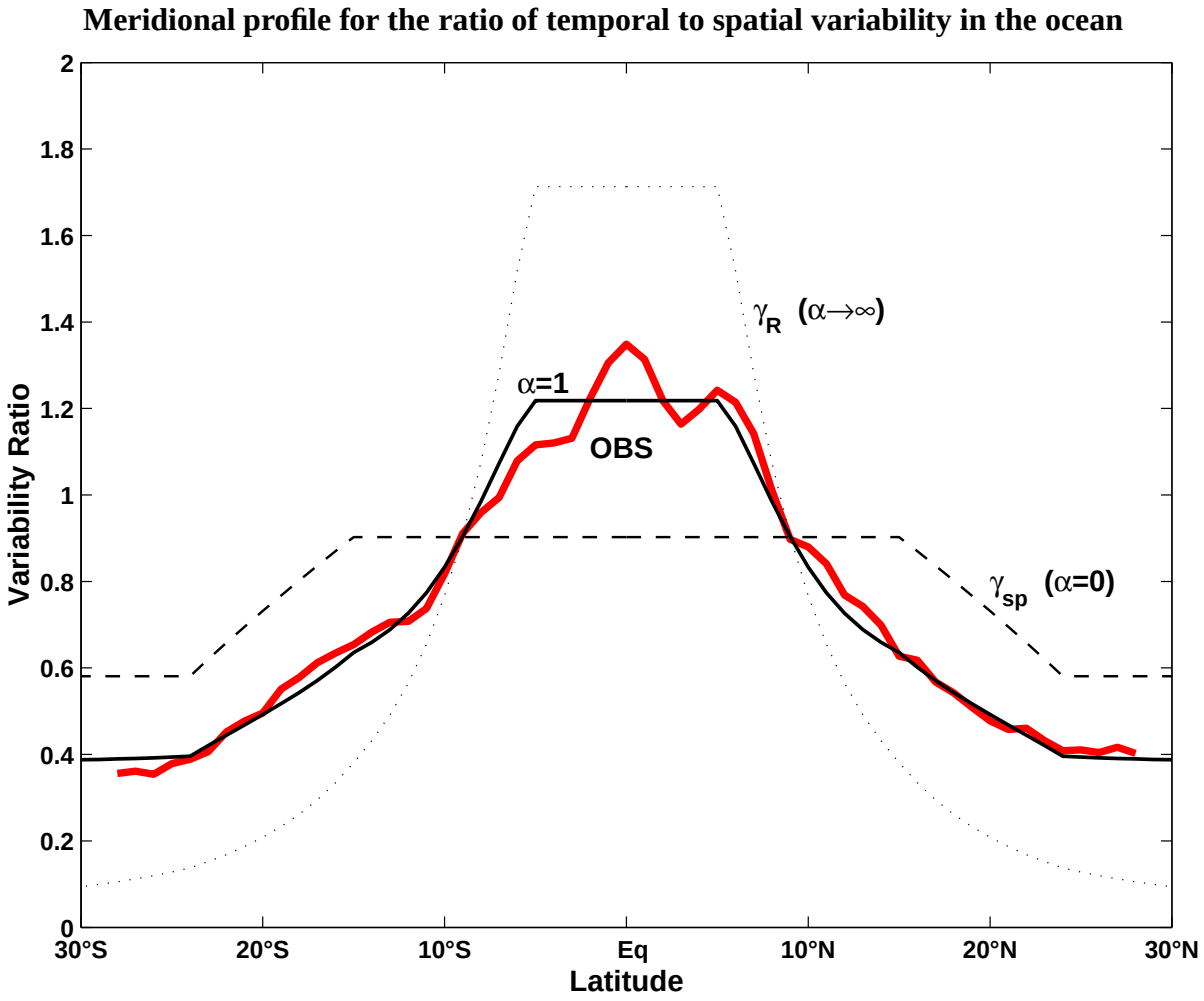


Figure 11: Zonal averages for the ratio of temporal to spatial variability estimated from the *Ducet et al.* [2000] analyzed altimetry fields (thick solid line) and the theoretical estimates by Equation (11) (thin solid line, dashes and dots correspond to $\alpha = 1$,0, and ∞ respectively).

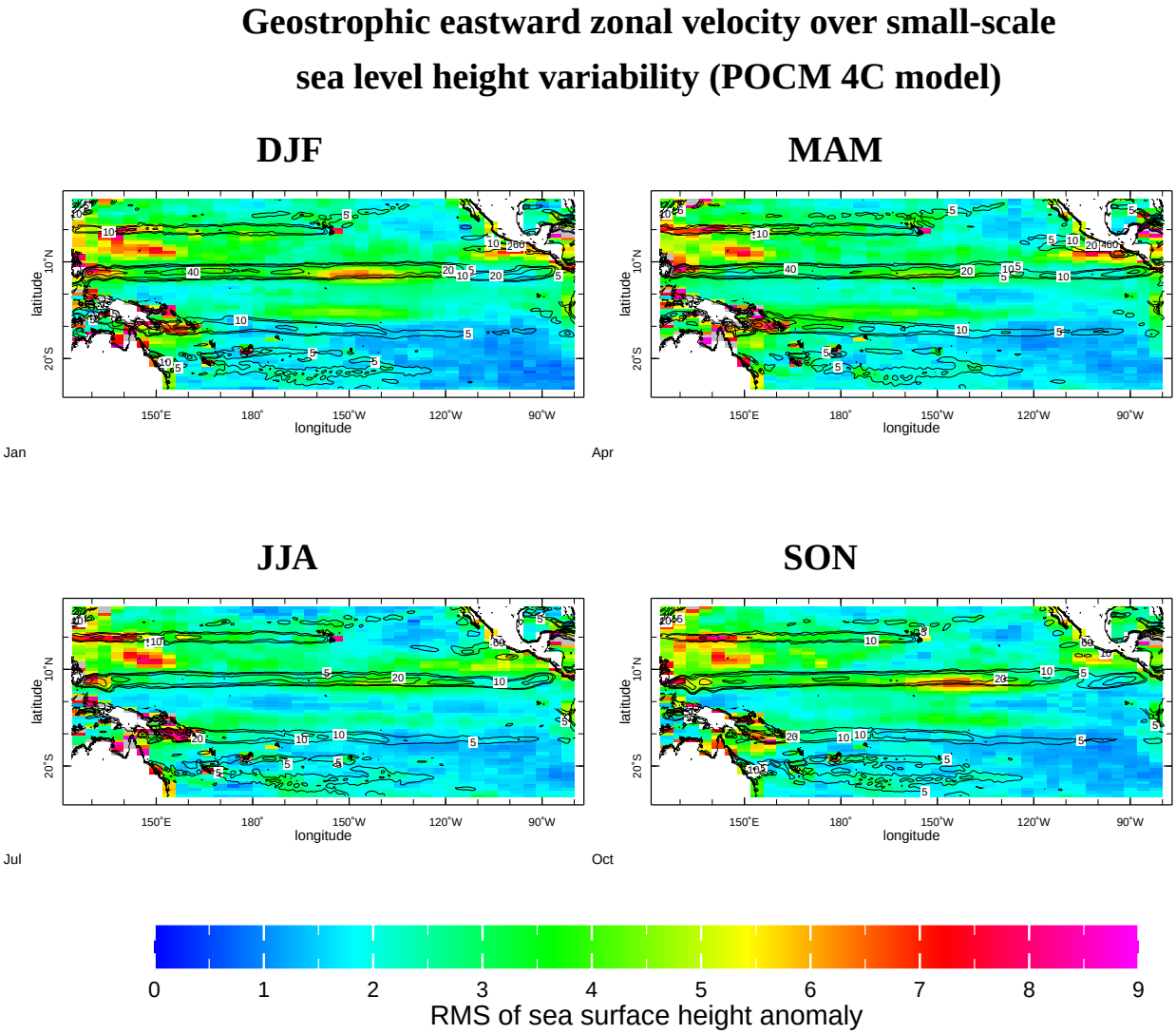


Figure 12: Seasonal means of geostrophic zonal velocity, cm/s (contours) is shown over seasonal small-scale sea level height variability estimates (colors) for POCM 4C model [*Tokmakian and Challenor*, 2001]. Only positive (eastward) velocity values are contoured.

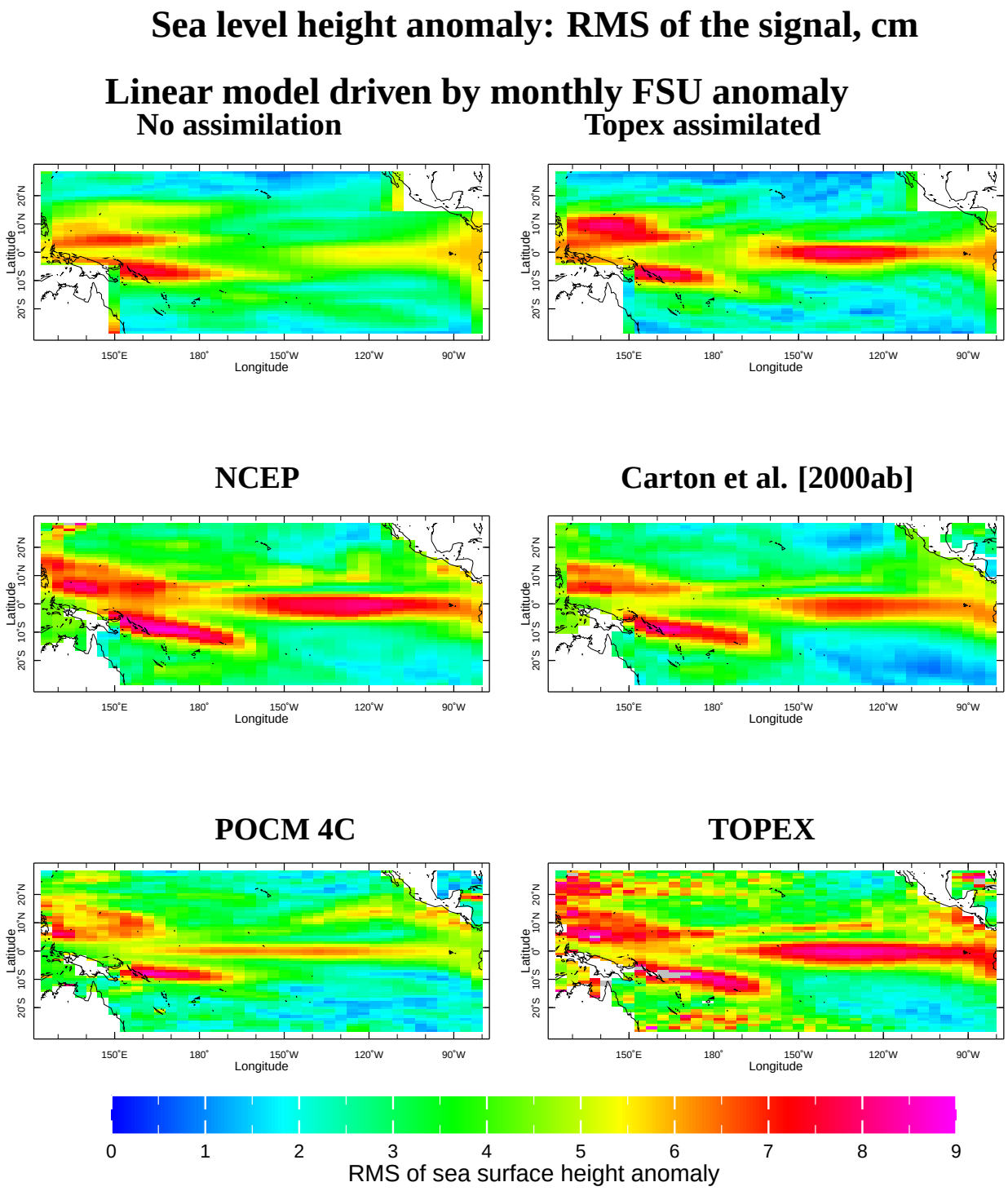


Figure 13: Sea level height anomaly RMS, cm, from different sources: (top) wind-forced and Topex-assimilated runs of the linear model by *Cane and Patton* [1984] (correspond to the upper left and lower right panels of Figure 1); (middle) assimilations of in situ data and TOPEX altimetry by NCEP [*Behringer et al.*, 1998] and *Carton et al.* [2000ab]; (bottom) POCM 4C [*Tokmakian and Challenor,* 1999], and gridded TOPEX sea level height anomaly [*Cheney et al.*, 1994].