Small-scale variability and model error in tropical Pacific sea level

A. Kaplan, M. A. Cane, D. Chen, D. L. Witter[1]

Lamont-Doherty Earth Observatory, Palisades, NY 10964

R.E. Cheney

NOAA/NESDIS Laboratory for Satellite Altimetry, Silver Spring, MD 20910

Short title: SMALL-SCALE VARIABILITY AND ERROR

---

[1]Presently at Kent State University, Department of Geology, Kent, OH 44242

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 toward 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 (SSV) of sea level height. Spatial and temporal components of this SSV are analyzed for predominant variability types. Monte Carlo experiments identify the areas where high SSV 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.

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). State-of-the-art GCM-based schemes of ENSO prediction [*Schneider et al.*, 2003] initialize the ocean component with the states obtained by the assimilation of observed temperature profiles into a surface flux driven ocean GCM.

For the goal of climate prediction it is important to bring the wind-driven ocean model

component into a realistic state by means of data assimilation. However, it is necessary to distinguish between errors in the observed sea surface height and contributions to the signal by processes not simulated by the model, whose gridsize is usually much larger than the observational footprint.

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]. These results are compared to the monthly T/P altimetry fields from *Cheney et al.* [1994], which have a spatial resolution of 4° longitude and 1° 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 variability (SSV) in sea level height 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 by the altimetry product of *Ducet et al.* [2000] and tide gauge data. In Section 5 we calculate the pattern of SSV in surface winds, and use Monte Carlo experiments with a linear ocean model to infer that most of the observed SSV in sea level heights is driven by that in surface winds. Areas where high energy SSV cannot be modeled as wind-driven are associated with eastward ocean currents, particularly

North Equatorial Countercurrent (Section 6). We show that the SSV 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 SSV with a well-known pattern of the mesoscale eddy energy and demonstrate that models heavily underestimate this variability, even if they are of high resolution and are driven by momentum flux fields with adequately represented short-term and small-scale variability. Thus the underestimation of the SSV 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

Definition of signal and small-scale variability. In nature there is variability of physical variables on all spatial and temporal scales. To facilitate its use, a variable is usually represented by a grid of its averages over corresponding 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 small-scale variability (SSV) inside the bins is important for many practical purposes as well. Note, that this division of the total variability of a physical variable into signal variability and SSV depends on the space and time dimensions of a chosen bin.

To make the presentation of our results unambiguous, we introduce the following notation. We will use square brackets not only for zonal averages as in *Starr and White* [1954] , but for averages over any space-time regions; e.g. $[s]_{G\times T}$ is the mean of variable

s over the space-time bin $G \times T$ ("signal"). Similarly, $\sigma_{G \times T}(s)$ is the root-mean-squared (RMS) deviation of s from its mean $[s]_{G \times T}$, i.e. SSV of s. Angle brackets denote statistical ensemble averaging. In practical calculations, under the ergodicity assumption, we use averaging over a temporal sample (thus the resulting estimates are affected by the sampling error). For example, $\langle \sigma^2_{G \times T}(s) \rangle_{\text{all } T}$ is the variance inside $G \times T$ bins averaged over all time periods T available for the spatial gridbox G.

Altimetry data. As a benchmark for comparisons we use the gridded T/P altimetry product of *Cheney et al.* [1994] which is available as 4° longitude by 1° latitude averages. Each $4^o \times 1^o$ bin is sampled between 2 and 4 times every 10 days, providing 6 to 12 altimetry values per month.

We also use sea level height anomaly estimates by *Ducet et al.* [2000] that merge T/P and ERS-1,2 measurements via an advanced global objective analysis. The resulting fields are of high resolution ($0.25^o \times 0.25^o \times 10$ days) but have a temporal gap from December 24, 1993 to March 24, 1995, when ERS-1 was flying along non-repeating tracks. Therefore this product is used for the verification of error estimates we develop and in the analysis of Section 7, where high resolution is necessary.

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.

3. Performance of model products

We start by evaluating the performance of a simple linear wind-driven model for sea level height anomalies. The model has 2° 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 (once a month). 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] (with a reduced space of 32 degrees of freedom). 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].

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°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 might appear 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°N slightly inclined toward the equator from the Central American coast, and low values on the equator and in the southeastern tropical Pacific. Either the problem lies with the altimetry fields, or there must be some similarity in the problems these models have with simulating sea level height variability. The latter makes the errors in wind forcing a prime suspect. These two possible problems, error in the altimetry and wind fields, are investigated in the next two sections.

4. SSV and error of gridded altimetry products

Figure 3 (top left panel) shows that the error pattern identified in the previous section is remarkably similar to the pattern of the SSV in the ocean surface height, which we define here as the variability inside $4^o \times 1^o$ monthly bins. The gridded $4^o \times 1^o$ altimetry product by [*Cheney et al.*, 1994] is constructed as averages of the values from those 6 to 12 TOPEX tracks that pass through each $4^o \times 1^o$ gridbox every month. Here, instead of taking averages, we compute the SSV from a sample of these 6–12 values for each gridbox and month. The results were averaged over all months in the period from October 1992 to March 2001, to

estimate the SSV (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 values. Since an estimate involves averaging of N observations, it is affected by a sample error whose standard deviation can be estimated as $\hat{\sigma}_{4^o\times1^o\times1 \text{ month}}/\sqrt{N_{4^o\times1^o\times1 \text{ month}}}$, assuming optimistically that all N altimeter passes in the month T sample uncorrelated deviations of the sea surface height from its gridbox mean $[s]_{4^o\times1^o\times1 \text{ month}}$. By averaging monthly error estimates over the entire time period, we obtain an estimate $r_{4^o\times1^o\times1 \text{ month}} = \sqrt{\langle \hat{\sigma}^2_{4^o\times1^o\times1 \text{ month}}/N_{4^o\times1^o\times1 \text{ month}}\rangle_{\text{all months}}}$ of the sampling error contribution to the total error of the *Cheney et al.* [1994] gridded altimetry product (Figure 3, lower left). The values $r_{4^o\times1^o\times1 \text{ month}}$ provide a lower bound to the error in the gridded fields, 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 $4^o\times1^o$ *Cheney et al.* [1994] grid boxes with more than 6 track crossings per month on average.

For a given station g the squared difference $d_g^2 = \langle (s_g - \widehat{[s]}_{4^o\times1^o\times1 \text{ month}})^2\rangle_{\text{all months}}$ is expected to be

$$d_g^2 = \alpha^2 r^2_{4^o\times1^o\times1 \text{ month}} + r_g^2, \tag{1}$$

where r_g is the RMS error of tide gauge values as estimators of the grid box mean. 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 individual altimetry passes, whose scatter reflects both the SSV and the altimetry net error.

Figure 4 (lower left panel) shows d_g versus $r_{4^o \times 1^o \times 1 \text{ month}}$ for all 31 stations. As expected from (1) with $\alpha \geq 1$, the former exceeds the latter: all circles are higher than the solid line. With d_g and $r_{4^o \times 1^o \times 1 \text{ month}}$ computed, (1) gives a constraint on α and r_g. The median values are $\alpha \approx 1.5$, $r_g \approx 1$ cm. For all but 4 of the stations α must be less than 2.3. Of the remaining 4, which allow $\alpha > 3$, three (Johnston, Rarotonga, Hilo) are located at around 20° latitude where the comparison is known to be problematic [*Mitchum*, 1994] because 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. (The remaining outlier is Nauru).

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 (Figure 3, lower right panel). We expect the error of the *Ducet et al.* [2000] product 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 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 significantly higher than $\alpha = 1.5$, their relatively small size suggests that there is much more to the model–altimetry differences than just the altimetry error.

5. SSV 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 SSV (not shown).

Figure 5 suggests that any gridded wind stress analysis produced from observations of incomplete temporal and spatial coverage will be affected by the error with a spatial variance pattern similar to that of the SSV. This error corresponds to the SSV in the wind removed by the wind stress gridding procedure. It has short spatial and temporal decorrelation scales.

We estimate $\sigma_{p\tau_x}$, the standard deviation of the corresponding SSV 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 sea level height SSV (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 SSV 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. Following *Miller et al.* [1996] we simulate the error in wind forcing by a random multivariate Gaussian error $e(x, y, t)$ in the

pseudostress components with covariance

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

where W is the spatially variable standard deviation of the estimated pseudostress SSV

(viz Figure 6 top panel). Wind errors at different time steps (1/4 month here) are taken

to be uncorrelated, while the spatial covariance structure is controlled by decorrelation

scales $L_x = L_y = 1^o$. We choose 1° since the decorrelation scales of SSV should be less

than the grid dimensions. An independent pseudostress field (both zonal and meridional

components) was generated at each model time step.

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 $\sqrt{\langle \sigma^2_{4^o \times 1^o \times 1 \text{ month}}(s)\rangle_{\text{months}}}$ (SSV averaged

over the entire Monte Carlo run of 2000 months), and (lower left panel) the RMS of the

means of these bins $\sigma_{\text{months}}([s]_{4^o \times 1^o \times 1 \text{ month}})$ (signal variability). The simulated spatial pattern

of the SSV (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.

The variability at signal resolution has 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 much longer spatial decorrelation

scales: $L_x = 20^o$, $L_y = 10^o$. The SSV and signal variability patterns are quite dissimilar in

this case, with the latter resembling the variability of the response to the "signal" in wind.

Both, however, are very different from the variance pattern for the response to the spatially uncorrelated wind shown in the lefthand panels.

6. Influence of eastward currents

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. Note that these areas are associated with the North Equatorial Countercurrent. Indeed the sea level SSV is colocated with the area of strong eastward currents indicated by positive zonal velocities at 7.5 m in the ocean analysis by *Carton et al.* [2000ab]. Moreover, the current strength and the sea level SSV manifest a similar timing within the seasonal cycle (Figure 8): the NECC and the SSV 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 sea level height SSV

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 separately the spatial and temporal SSV. Since there were only 3 temporal samples per month, the temporal variance estimates were multiplied by 3/2 to produce an unbiased estimate [*Mardia et al.*, 1979].

The pattern of $\sigma_{4^o \times 1^o \times 1\,\text{month}}$ computed from *Ducet et al.* [2000] values for each $4^o \times 1^o \times 1$ month bin (Figure 9, upper left panel) is quite similar to the one computed from T/P track data (Figure 3, upper left panel). 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 $5–8^o$N between 170^oW and 110^oW predominantly involves temporal variability of the bin spatial means. This division is consistent with two different variability mechanisms (wind bursts and instability waves) in these two adjacent areas as suggested by *Giese et al.* [1994].

The ratio γ of temporal to spatial variability (Figure 9, lower right panel) manifests a remarkably zonally uniform meridional variability structure. 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. Figure 10 demonstrates that the observed zonal means of this ratio can be explained by the combined contribution of Rossby and short-period (mostly inertia-gravity) waves. Details of this estimation are given in Appendix. Note that 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. The contribution from the short-period waves makes the latitudinal change more realistic.

The observed peak at around 5^oN is due to the area of instability waves associated with the NECC. These waves have the longer wavelengths than the waves propagating in stable conditions at these latitudes. Similarly, the equator and 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 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 needed in optimal data assimilation procedures.

Sea level height SSV and eddy kinetic energy. The pattern of the T/P SSV (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. 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 sea level height SSV resembles the structure of the variance in their seasonal cycle of the eddy kinetic energy.

The similarity between the patterns of SSV and eddy kinetic energy is interesting and was noted before [e.g. *Stammer*, 1997]. 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. For an explicit mathematical relation between these two see *Kaplan* [2003].

Causes for the SSV and its modeling. Connection between the SSV 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 SSV in the wind stress. 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 toward 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 [*Qiu*, 1999; *Ducet et al.*, 2000].

The interpretation of the eastern Pacific variability band is more complicated. In fact, it 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 variability maximum at $5-8^oN$ corresponds to the areas where horizontal shear between NECC and SEC is likely to be high (Figure 8).

This maximum is completely missed in our Monte Carlo experiments, emphasizing the importance of the instability mechanisms compared to the local wind forcing. This area is also identified by a particularly high ratio of temporal to spatial variability (Figure 9, lower right) pointing toward the predominance of very long waves in this area.

Simulation of the SSV due to currents' instabilities needs a high resolution non-linear model. For example, for the 0.25^o resolution POCM 4C model [*Tokmakian and Challenor*, 2001] all maxima in SSV are either in the areas of eastward currents or near islands. In particular, the POCM 4C model successfully produced variability maxima in the areas of NECC and SEC. However, the major variability maxima due to the wind in the northwestern and southwestern corners of the tropics are almost completely missing. At the same time, some variability is produced seasonally 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 wind bursts 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.

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; e.g. note the absence of an error maximum there in the NCEP assimilation run (Figure 2).

SSV 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 sea level height SSV, even when driven by winds with a good representation 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 SSV in the northwestern and southwestern tropical Pacific – only variability due to currents' instability gets simulated (cf upper panels of Figure 3). This underestimation is most likely due to the friction parameterization in ocean GCMs, which overly damps smaller scales.

Underestimation of the SSV by the model products seems to be connected to underestimation of signal variability in the same areas (Figure 11). Areas of high variability poleward of 20° in the tropical western Pacific, present in the T/P panel of Figure 11, 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 just attenuate in time, or are not allowed at all.

In other words, the 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 SSV in the ocean is high. Simulation error patterns additionally include areas where the signal variability is high, since the model response to wind products' errors having long (signal-type) decorrelation scale has a pattern similar to the response to large-scale wind signal. 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 1d) has some features which resemble the signal

variance, e.g. equatorial maximum in the tropical Pacific (cf. Figure 11, 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 SSV 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 has many features of the SSV pattern, the same pattern that influences the model error.

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 areas of large error are near the northwestern and southwestern corners of the domain, as well as in a band around 10^oN in the eastern part of the domain. A similar pattern of differences was found in the comparison of a wide variety of GCM simulation and assimilation products with the T/P altimetry fields.

This difference pattern was identified as the pattern of high SSV in the tropical sea

level height. Accordingly, the gridding error in the T/P altimetry product has a similar structure. However, since it is smaller in size, the error of model products must have this structure as well. The explanation is that (1) most of the features of this pattern of variability can be caused by the SSV (and error) in wind, and (2) most GCMs underestimate SSV even if driven by 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 wind SSV are those of high shear and current instabilities in the ocean, and have their own host of simulation problems. In other words, we have used SSV to present a "model invalidation" study (and an answer to *Oreskes et al.* [1994] who "have never seen a paper in which the authors wrote 'the empirical data invalidate this model' ").

Systematic underestimation of variance by the ocean models should be taken into account by data assimilation systems. Comparison of the top panels of Figure 11 with the left panels of Figure 7 suggests 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 contrasting types of error patterns: in response to small-scale noise and to 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.

Appendix: 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):

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

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

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

Since the last expression does not depend on t,

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

as well. Finally,

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

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

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

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 by color in Figure A1.

Ratio for a given dispersion relation. We need to estimate 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 γ_t and a spatial γ_s pieces:

$$\gamma_t = \frac{\gamma^2}{\gamma^2 + 1} = \frac{\mathcal{C}(\lambda)(1 - \mathcal{C}(\tau))}{1 - \mathcal{C}(\lambda)\mathcal{C}(\tau)}, \qquad \gamma_s = \frac{1}{\gamma^2 + 1} = \frac{1 - \mathcal{C}(\lambda)}{1 - \mathcal{C}(\lambda)\mathcal{C}(\tau)}.$$

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

$$\gamma_D(L_1, L_2) = \sqrt{\frac{\int_{L_1}^{L_2} \gamma_t(T_D(L)/T_b, L/L_b) F(2\pi/L, 2\pi/T_D(L)) L^{-2} dL}{\int_{L_1}^{L_2} \gamma_s(T_D(L)/T_b, L/L_b) F(2\pi/L, 2\pi/T_D(L)) L^{-2} dL}}. \tag{A2}$$

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 know the 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

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

Ocean waves. To attribute parameters of the actual ocean waves to the points of the diagrams in Figure A1, we let $L_b = 4^o$ and $T_b = 1$ month. Then a given dispersion relationship $T = T_D(L)$ defines a curve on the $(T/T_b, L/L_b)$ plane.

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

$$T = T_R(L) = \frac{f(y)^2 L^2/c^2 + 4\pi^2}{\beta(y)L} \tag{A4}$$

[*Gill, 1982*]. (Unlike the wavenumber, we take the wavelength L to be positive for waves propagating either to the west or east). The minimum allowable period and wavelength are

$$T_{\min}(y) = 4\pi f/\beta c, \quad L_{\min}(y) = 2\pi c/f, \tag{A5}$$

and so are functions of latitude.

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 (A4) is applicable to trapped modes as well, if f is evaluated at the turning latitude y_T, and β has its equatorial value (which is very close to $\beta(y_T)$ for low number modes) [*Cane and Sarachik*, 1976].

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 A1 (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° and 29°.

Therefore we have two distinct types of dispersion relationships: $T = T_R(L)$ for Rossby waves, defined by (A4), and $T = T_{\mathrm{sp}} \approx 0$ 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. Based on the study by *Stammer* [1997] we take the power spectral density of tropical sea level heights in the form

$$F(k, \omega) \sim k^{-2.5}\omega^{-0.5} \tag{A6}$$

for waves longer than 100 km with periods between 30 and 200 days. Since the spectral density (A6) favors the long waves, Rossby waves are integrated from $L_{\min}(y)$ (defined by (A5)) up along their dispersion curves:

$$\gamma_R = \sqrt{\frac{\int_{L_{\min}(y)}^{L_{\max}} \gamma_t(T_R(L)/T_b, L/L_b)\sqrt{LT_R(L)}dL}{\int_{L_{\min}(y)}^{L_{\max}} \gamma_s(T_R(L)/T_b, L/L_b)\sqrt{LT_R(L)}dL}}. \tag{A7}$$

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.

For the short period waves take $T_D \to 0$ in (A2) to obtain

$$\gamma_{\rm sp} = \sqrt{\frac{\int_{L_{\min}}^{L_{\max}} \mathcal{C}(L/L_b)\sqrt{L}dL}{\int_{L_{\min}(y)}^{L_{\max}} (1 - \mathcal{C}(L/L_b))\sqrt{L}dL}}. \tag{A8}$$

Modeling observed latitudinal profile of γ. We attempt to estimate observed values of γ via formula (A3), by mixing contributions from Rossby and short-period waves. Comparison of the variability in pentad (five day periods) averages of $4^o \times 4^o$ 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 we assume that a roughly similar amount of energy goes into Rossby and short-period waves, hence $\alpha = 1$.

Individual ratios γ_R and $\gamma_{\rm sp}$ are estimated numerically using (A7) and (A8). The lower integration limit in (A8) is taken $L_{\min} =$100 km, results not being very sensitive to this choice. Results are, however, quite sensitive to the upper integration limit $L_{\max}$ in both (A7) and (A8). We found that not an unreasonable latitudinal profile of $L_{\max}$ values ($1.9 \cdot 10^3$km within 15° of the equator, $1.3 \cdot 10^3$km poleward of 25°, and the linear change with latitude in between) results in a good agreement of formula (A3) with observations (Figure 10).

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., 2003: Eddy kinetic energy and small-scale sea level height variability. *J. Phys. Oceanogr.*, in revision. [Available from Institute for Mathematics and its Applications, preprint

#1913: `http://www.ima.umn.edu/preprints/mar2003/1913.pdf`].

Kaplan, A., Y. Kushnir, M. Cane, and M. Blumenthal, 1997: Reduced space optimal analysis for historical data sets: 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.

Oreskes, N., K. Shrader-Frechette, and K. Belitz, 1994: Verification, validation, and confirmation of numerical models in the earth sciences, *Science*, **263**, 641-646.

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., D.G.DeWitt, A.Rosati, B.P.Kirtman, L.Ji, J.J.Tribbia, 2003: Retrospective ENSO forecasts: Sensitivity to atmospheric model and ocean resolution, *Mon. Weather Rev.,* in press.

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

Sea level height anomaly: RMS[Topex − Linear model], cm

No assimilation

Tide gauges assimilated

Temperature profiles assimilated

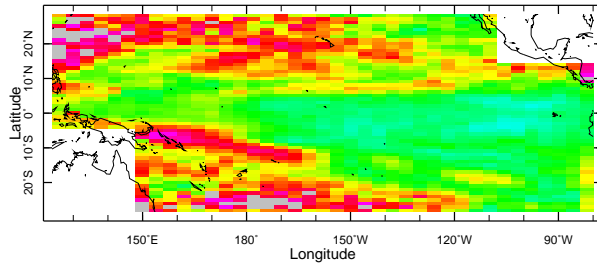## Topex assimilated

RMS of sea surface height anomaly

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].

Sea level height anomaly: RMS[Topex − models], cm

NCEP

POCM 4C

Fukumori et al. [1999]

No assimilation

Topex assimilated

Carton et al. [2000] assimilations

in situ only

in situ and altimetry

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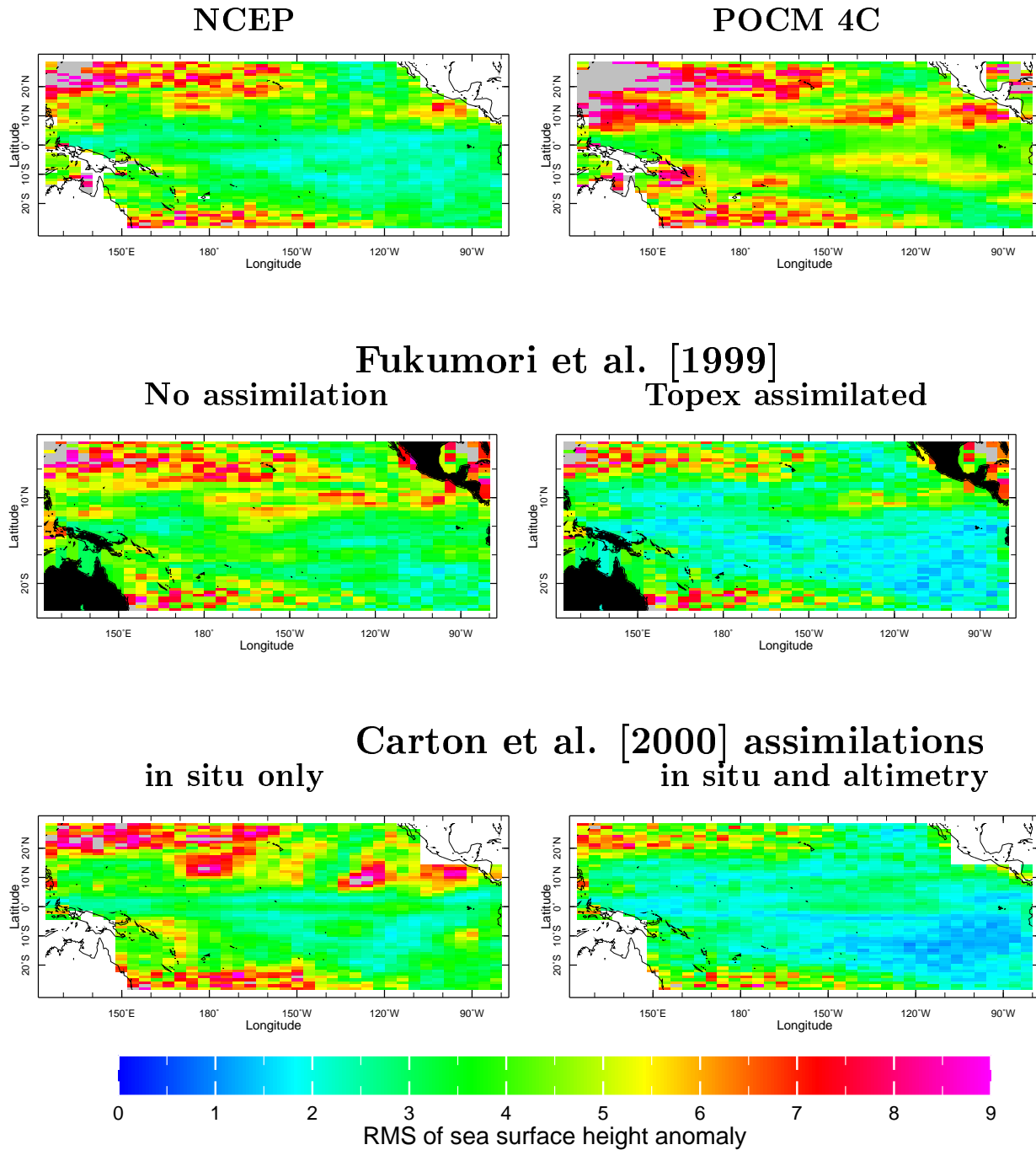
Figure 2. Comparison with the Topex/Poseidon altimetry [*Cheney et al.*, 1994] of sea level height anomalies from six GCM products. Shown are RMS differences for: the NCEP MOM1 version for the Pacific Ocean, assimilating temperature profiles and Topex altimetry [*Behringer et al.*, 1998] (1993-2000);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

Small-scale variability and gridded altimetry error

Variability inside $4^o \times 1^o$ monthly bins: $\hat{\sigma}_{4^o \times 1^o \times 1\,\text{month}}$

Topex

POCM 4C
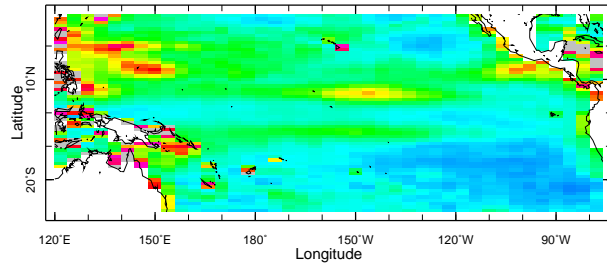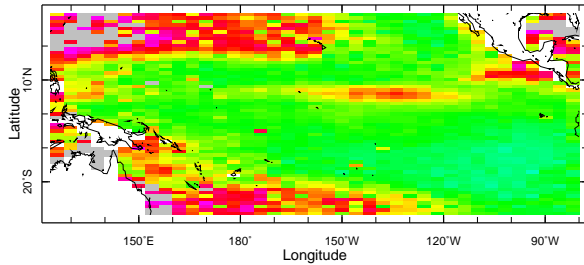
Est err of gridded values: $r_{4^o \times 1^o \times 1\,\text{month}}$ RMS[Cheney et al. - Ducet et al.]
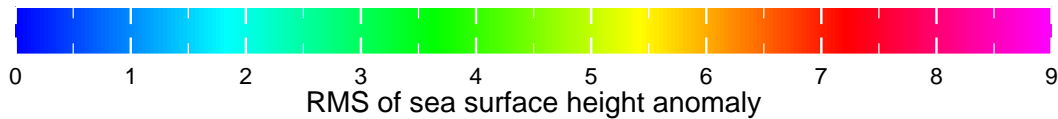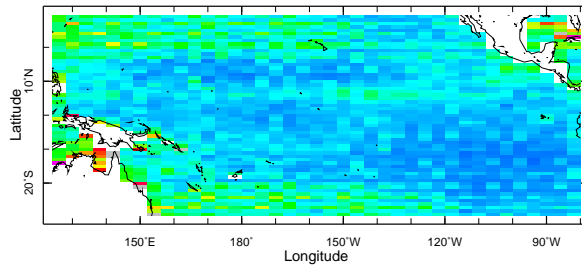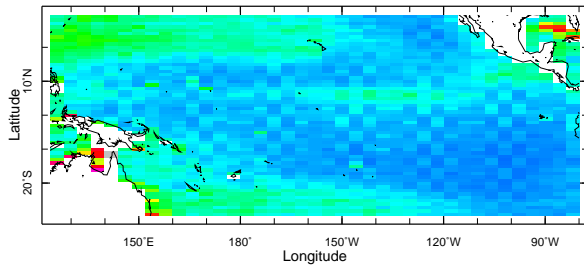
RMS of sea surface height anomaly

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.
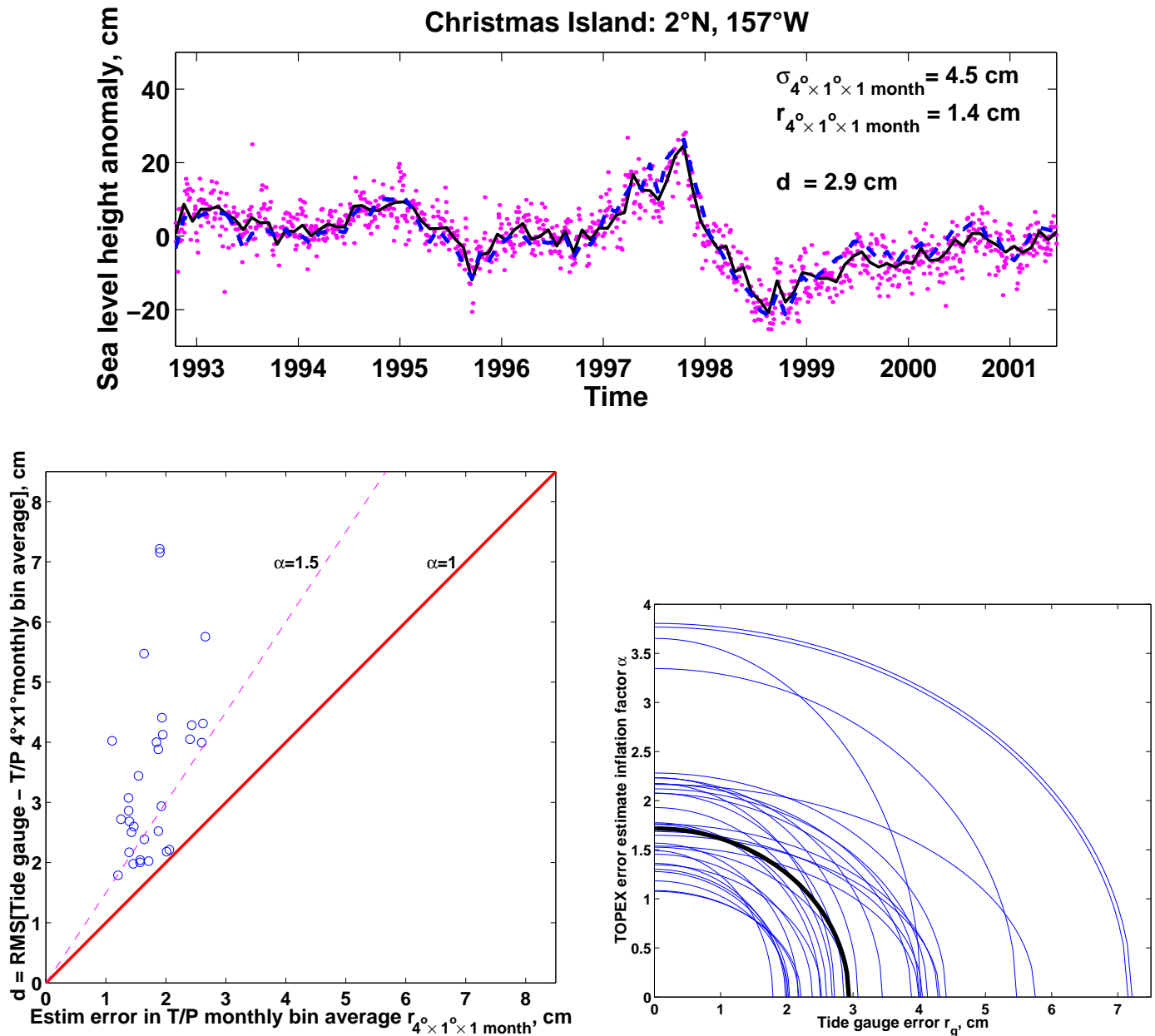
Validation of Error Estimates

Figure 4. Validation of T/P error estimates by comparison with the tide gauge records, October 1992 – March 2001. The top panel compares monthly tide gauge sea level height anomalies at Christmas Island (dashes) with altimetric measurements from the corresponding gridbox (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 their monthly averages for this gridbox. Temporal RMS values of the intrabox variability $\sigma_{4^o \times 1^o \times 1\ \text{month}}$ inside the gridbox, the sampling error estimate $r_{4^o \times 1^o \times 1\ \text{month}}$ for the gridbox mean, and the RMS difference between the gridbox 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 (1) for individual tide gauges. The thick line shows the median constraint.

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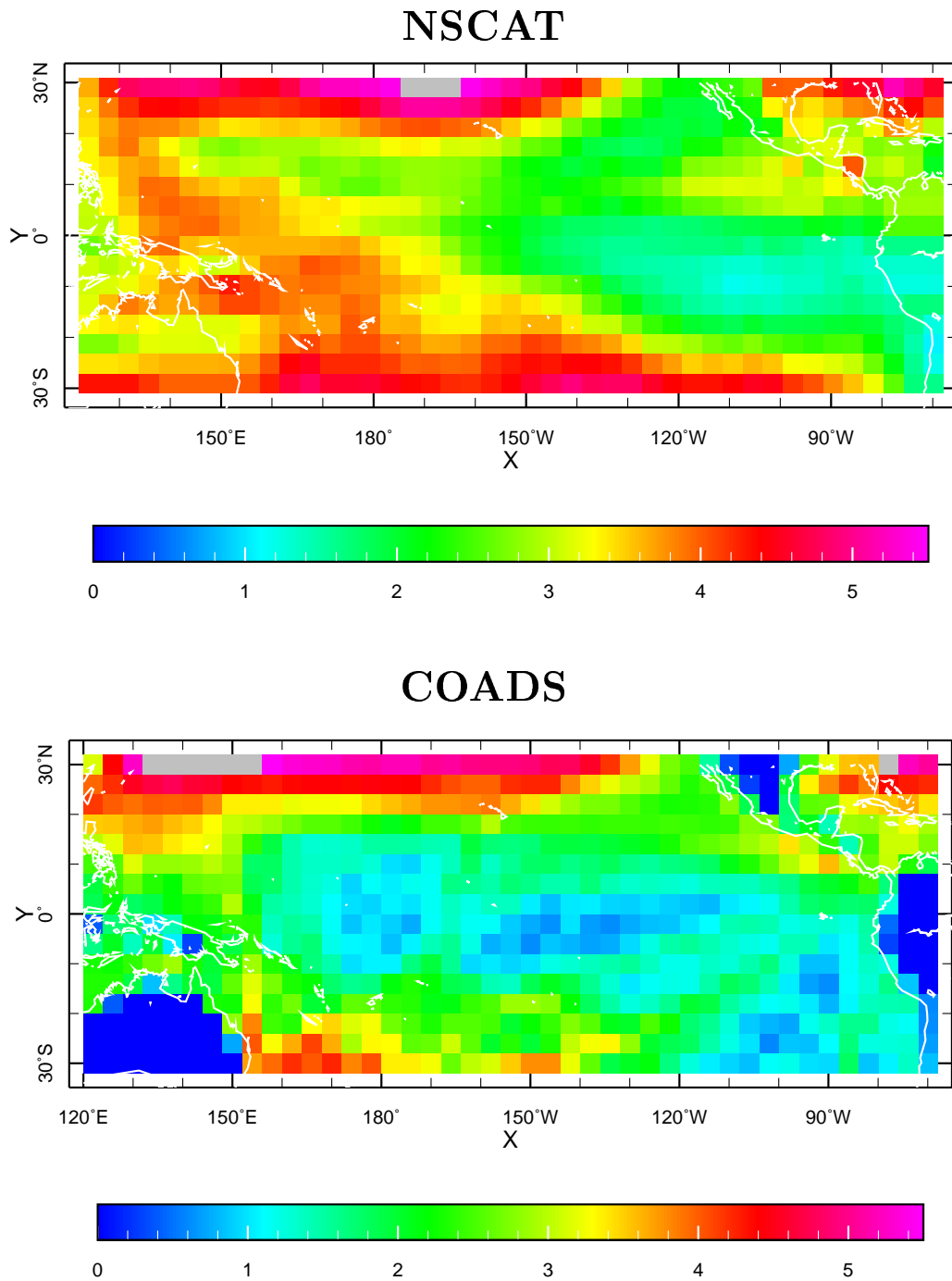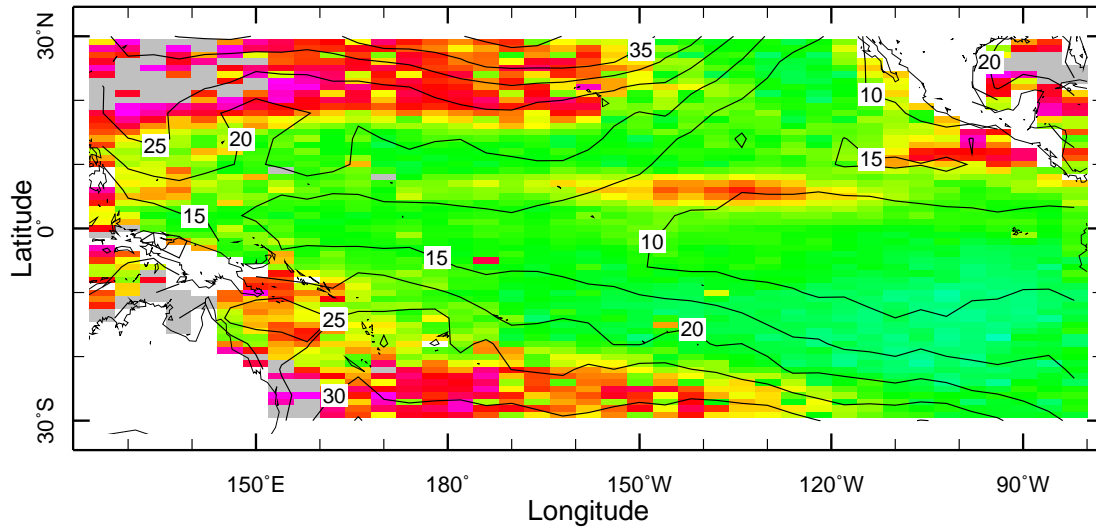
NSCAT

COADS

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.
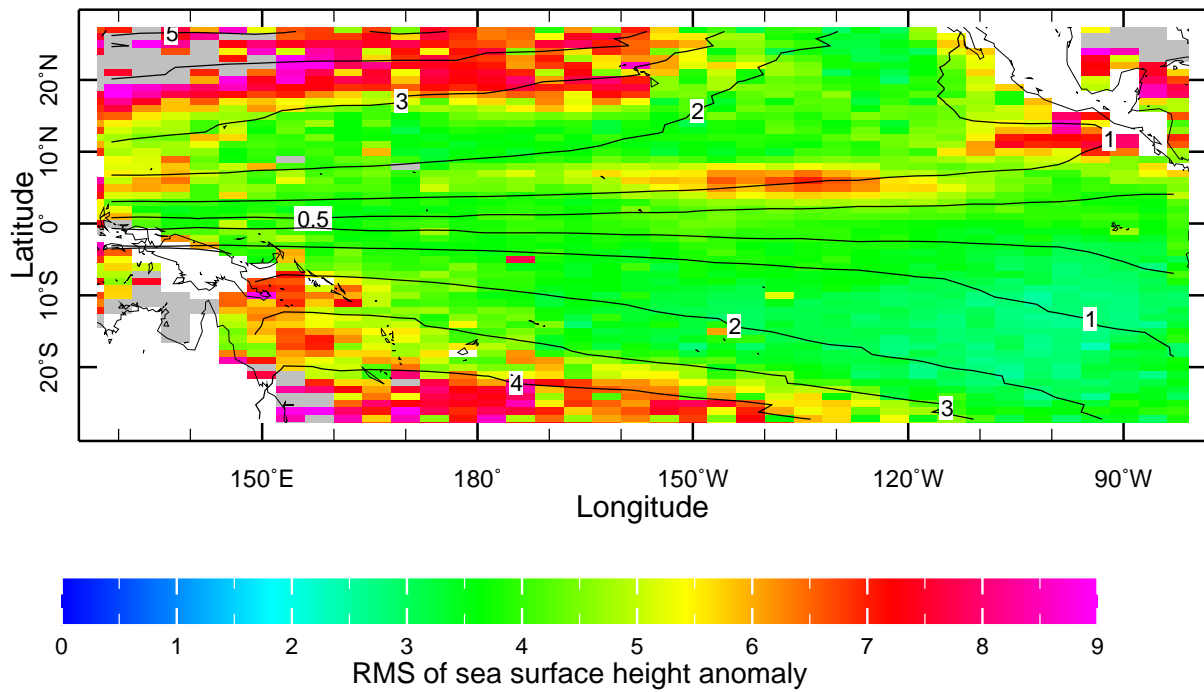
Small-scale variability in wind and sea level height

Zonal pseudostress and sea level height observations

10 m_above_gnd
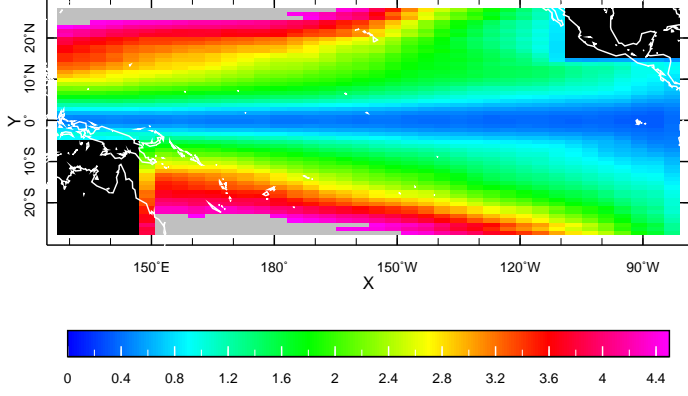
Sea level height simulation and observations

0.0

RMS of sea surface height anomaly

Figure 6. SSV in pseudostress and sea level height responce: (top) Contours of SSV 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 the SSV RMS in the sea level height response of a linear model to the random wind with 1° spatial and 0.25 month temporal decorrelation scales. See text for explanations.
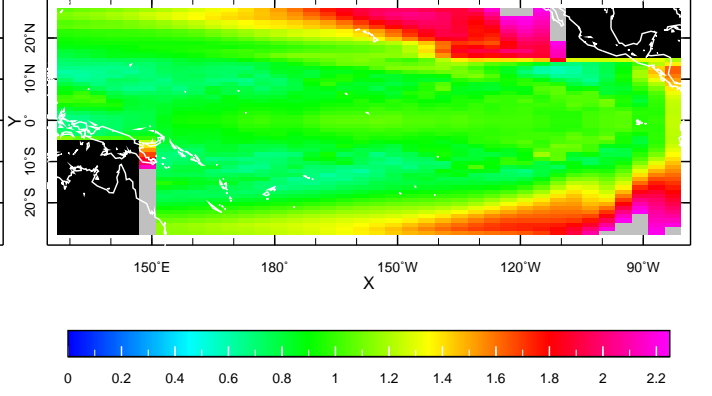
Monte Carlo experiments with a linear model

Variability inside $4^o{\times}1^o$ monthly bins: $\sqrt{\langle\sigma^2_{4^o\times1^o\times1\text{ month}}(s)\rangle_{\text{months}}}$

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

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

Variability of $4^o{\times}1^o$ monthly means: $\sigma_{\text{months}}([s]_{4^o\times1^o\times1\text{ month}})$

$1^o{\times}1^o$ scaled noise
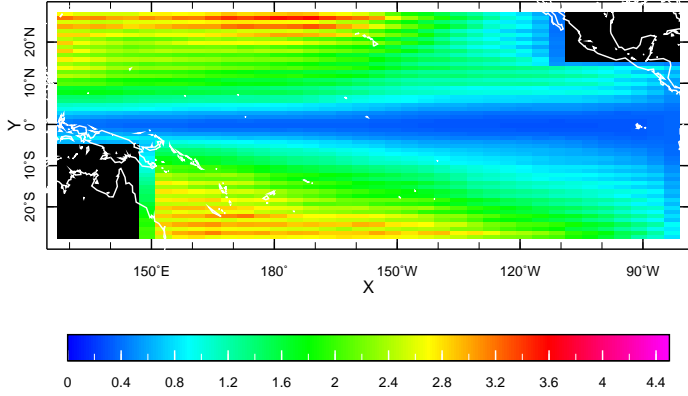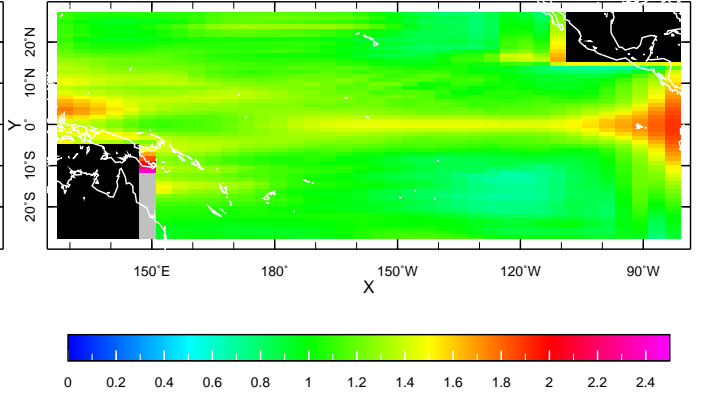
$20^o{\times}10^o$ scaled noise

Figure 7. Simulation of sea level height error and SSV 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 explanations.

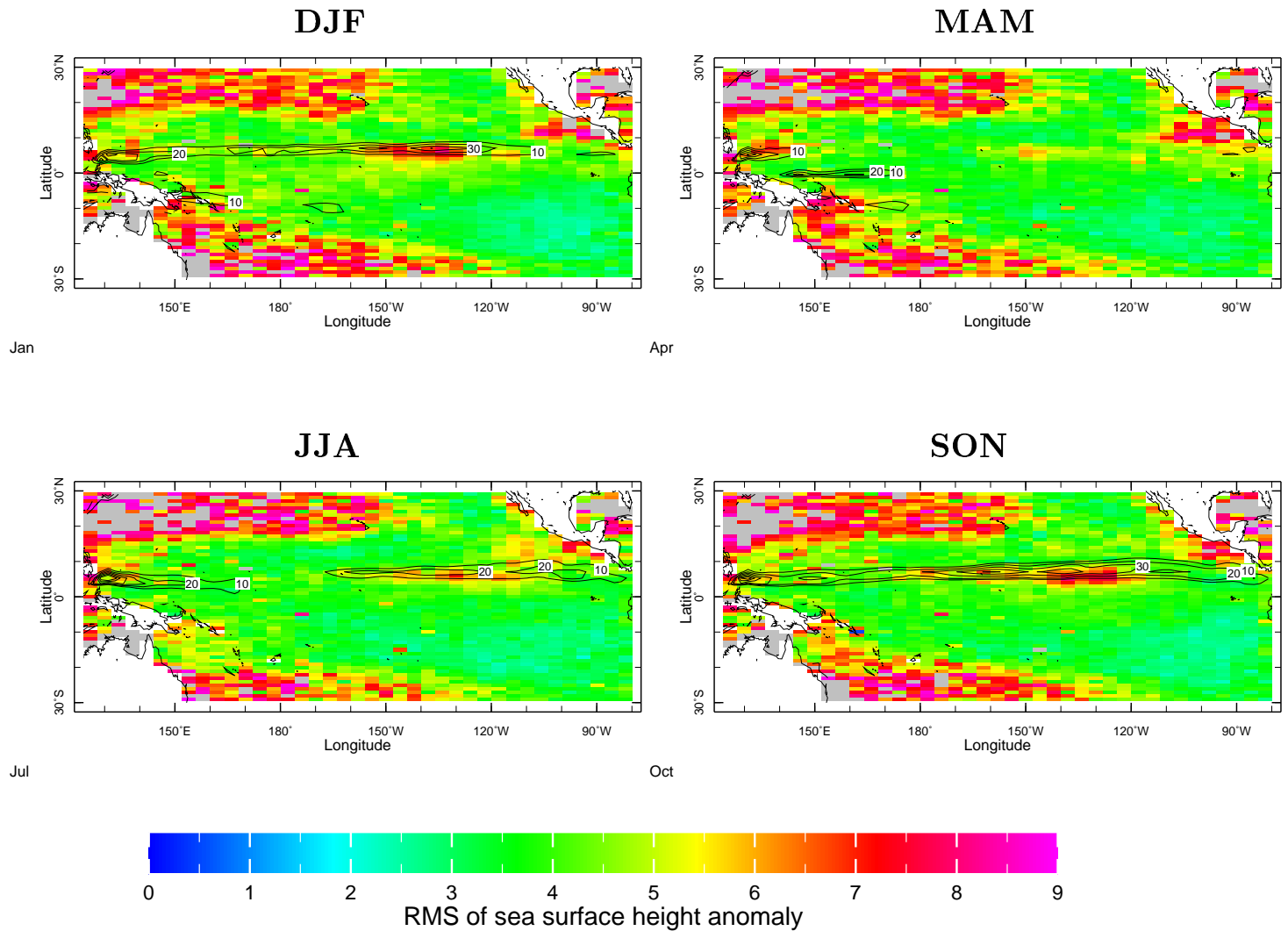# Eastward zonal velocity over small-scale sea level height variability

Figure 8. Seasonal means of zonal velocity [*Carton et al.*, 2000ab], cm/s (contours) are shown over seasonal SSV estimates for sea level height (colors). Only positive (eastward) velocity values are contoured. The sea level SSV is estimated from 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) $\sqrt{\langle \hat{\sigma}^2_{4^o \times 1^o \times 1 \text{ month}} \rangle_{\text{DJF}}}$, March-April-May (MAM) $\sqrt{\langle \hat{\sigma}^2_{4^o \times 1^o \times 1 \text{ month}} \rangle_{\text{MAM}}}$, etc.

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

Total variability $\sigma_{4^o \times 1^o \times 1\,\mathrm{month}}(s)$

Temporal variability of spatial means $\sigma_{\mathrm{months}}([s]_{4^o \times 1^o})$

Spatial variability inside bins $\sqrt{[\sigma^2_{4^o \times 1^o}(s)]_{1\,\mathrm{month}}}$

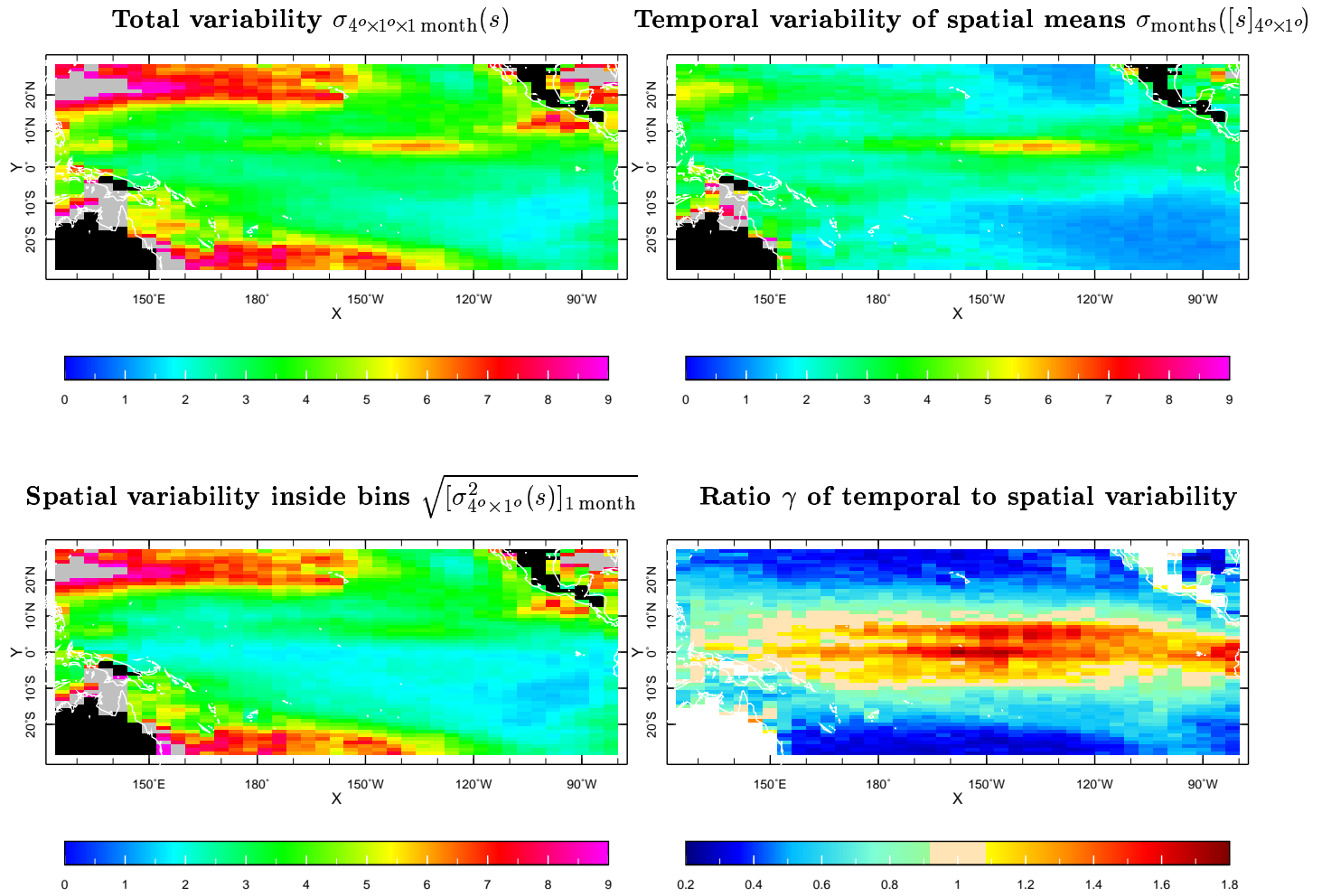**Ratio γ of temporal to spatial variability**

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

Meridional profile for the ratio of temporal to spatial variability in the ocean
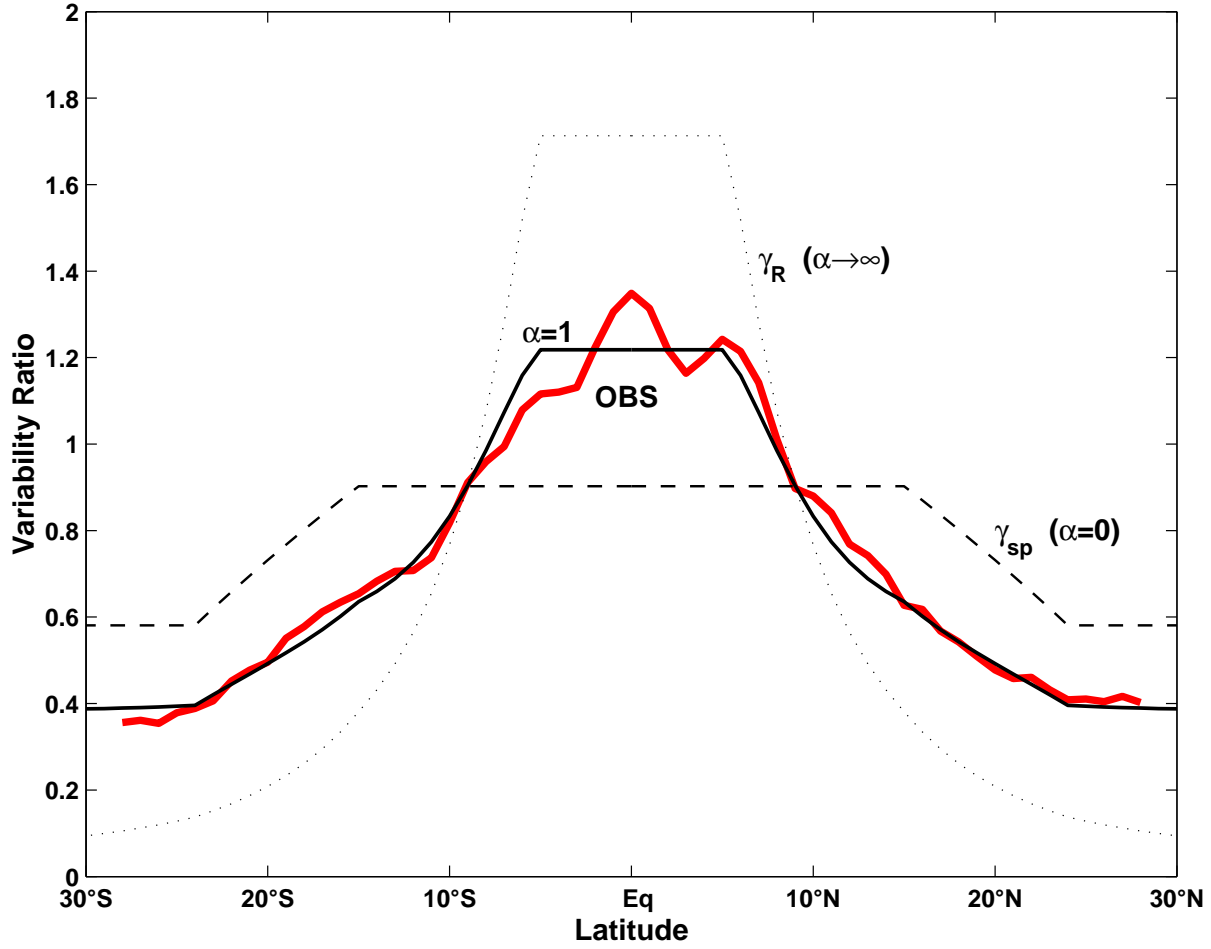

Figure 10. Zonal averages for the ratio of temporal to spatial variability estimated from the *Ducet et al.* [2000] analyzed altimetry fields (thick solid line) and the theoretical estimates by equation (3) (thin solid line, dashes and dots correspond to $\alpha = 1,0$, and ∞ respectively).

Sea level height anomaly: RMS of the signal, cm

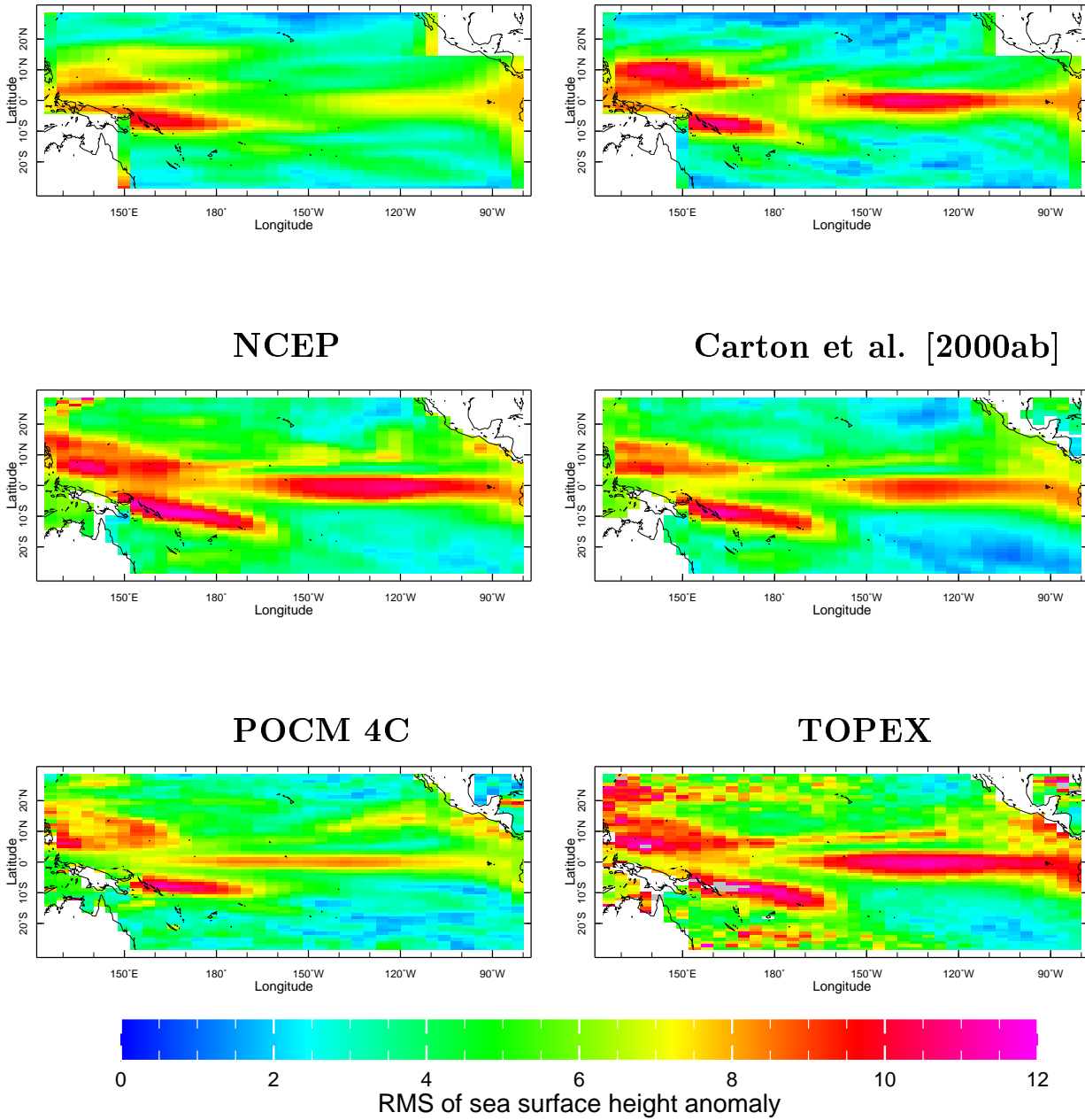## Linear model driven by monthly FSU anomaly

Figure 11. 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].

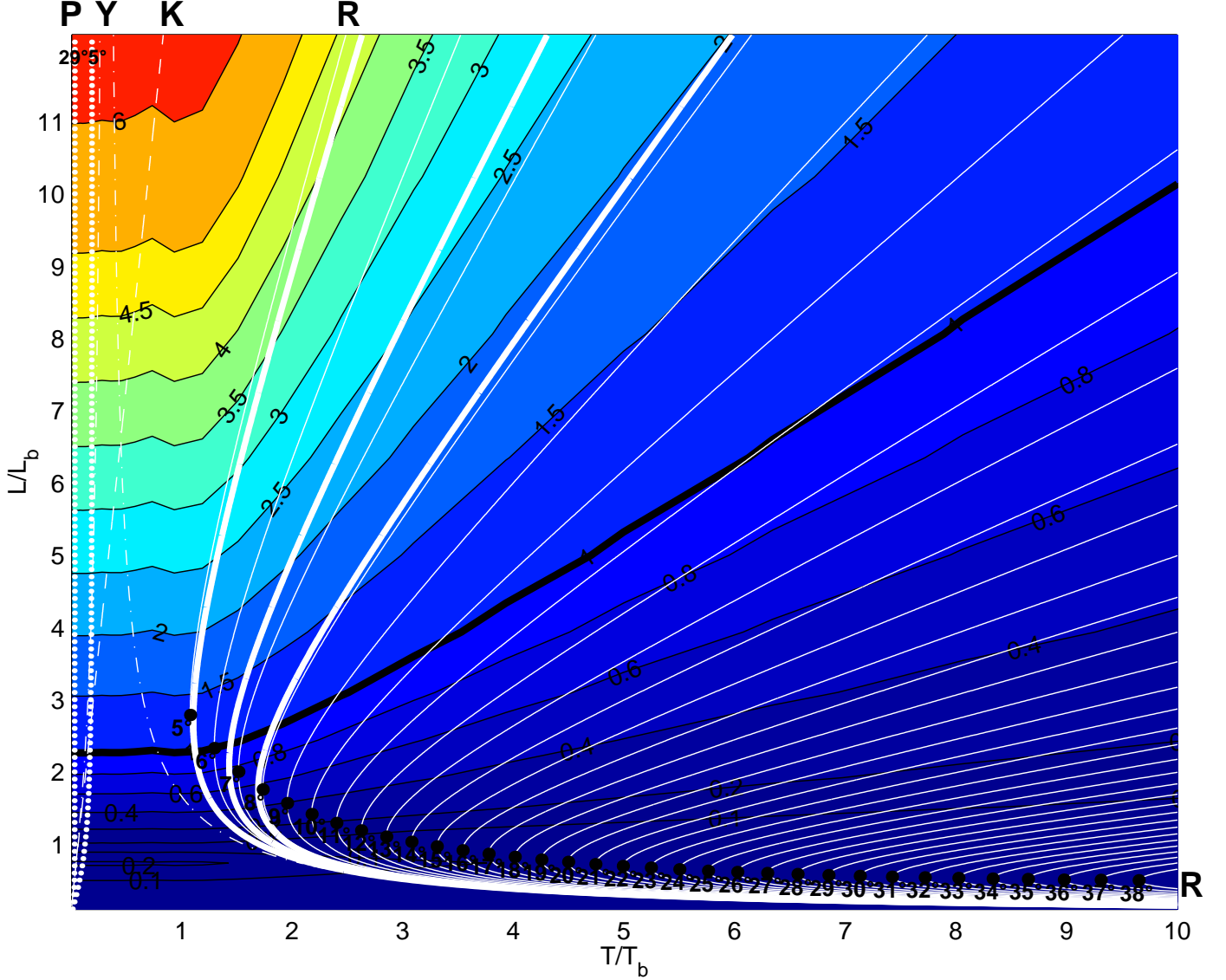# Ratio of temporal to spatial variability for ocean waves

Figure A1. 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° and 29°. In order to put wave parameters on the dimensionless color diagram, $T_b = 1$ month and $L_b = 4^o$ are assumed.