An Adaptive Procedure for Tuning a Sea Surface Temperature Model*

NATHALIE SENNÉCHAEL AND CLAUDE FRANKIGNOUL

Laboratoire d'Océanographie Dynamique et de Climatologie, Université Pierre et Marie Curie, Paris, France

MARK A. CANE

Lamont–Doherty Earth Observatory, Palisades, New York

(Manuscript received 3 November 1993, in final form 23 February 1994)

ABSTRACT

To determine the value of the adjustable parameters of an ocean model required to optimally fit the observations, an adaptive inverse method is developed and applied to a sea surface temperature (SST) model of the tropical Atlantic. The best-fit calculation is performed by minimizing a misfit between observed and simulated data, which depends on the observational and the modeling errors. An adaptive procedure is designed in which the model being tuned is also used to construct a model of the observational errors. This is done by performing the optimization on the mean seasonal cycle and using the SST anomalies obtained for different years and plausible forcing fields as additional information to construct a sample estimate of the observational error covariance matrix. Assuming idealized modeling errors, the procedure is applied to the SST model of Blumenthal and Cane, yielding refined estimates for several models and heat flux parameters. The simulation of the mean annual SST is improved, but not the simulation of seasonal and interannual variability. The model–observation discrepancies remain too large to be solely attributed to atmospheric and oceanic data uncertainties and are linked to the model's rudimentary geometry and its incorrect representation of SST cooling by upwelling. The existence of larger model deficiencies than was originally assumed in the model errors is confirmed by a statistical test of the correctness of the assumptions in the inverse calculation.

1. Introduction

All oceanic models contain parameterizations of such physical processes as convection, near-surface mixing, and horizontal mixing due to subgrid-scale motions. Parameterizations are based on physical ideas but typically yield forms that contain parameters whose values are not known precisely from theory. When a parameter is model dependent (e.g., mixing is a function of grid spacing), parameter tuning will be in part model dependent. Surface forcing also depends on poorly known parameters. Even in data-rich regions where random measurement and sampling errors are small, the fluxes are poorly known because of systematic data biases and bulk formulation uncertainties. Modeling these quantities may be considered part of the problem of creating an oceanic model (Seager et al. 1988). In view of their inherent imprecision, the uncertain parameters should be tuned against observed data. At the same time, models should be consistent with known physics to within the tolerances allowed by the approximations made.

Particularly in the tropics where observations are sparse, both forcing and verification data are imprecisely known. Hence, the accuracy to be expected in model simulations is limited, even if the physics is perfectly represented, and data uncertainties should be taken into account in parameter tuning. Frankignoul et al. (1989) developed a multivariate model testing procedure that provides an objective measure of the fit between ocean model simulations and observations, taking into account the data uncertainties. Using a trial and error approach, the method can be used for model tuning. This is illustrated by Duchêne and Frankignoul (1991) and Braconnot and Frankignoul (1993), who determined the vertical resolution of the linear model of Cane (1984), which provided an optimal fit to observations of surface currents and thermocline depth in the tropical Atlantic. However, this approach requires that the number of adjustable parameters be small.

A more efficient tuning approach is that of Blumenthal and Cane (1989), who used inverse modeling procedures to determine the parameter values required to optimally fit sea surface temperature (SST) in a simplified SST model. A priori knowledge constraining the parameter range was included in the calculation,

* Lamont–Doherty Earth Observatory Contribution.

Corresponding author address: Dr. Claude Frankignoul, LODYC, Université Pierre and Marie Curie, Tour 14, 2e étage, 4, Place Jussieu, Paris, Cedex 05 75252 France.

but only a highly idealized model was used for the data errors. The error model enters the measure of the misfit between observed and predicted data which is minimized in the best-fit calculation. Thus, the atmospheric forcing uncertainties need to be properly represented, as they introduce large uncertainties in SST model response (e.g., Harrison et al. 1990).

As the forcing uncertainties have large and poorly known correlation scales, the error estimates are best derived from direct simulations. We have thus developed an adaptive tuning procedure, where the model that is being tuned is also used to construct the observational error model for the best-fit calculation. The tuned model is then tested against observations and if it agrees with the data to within expected errors, it will be judged adequate. Such an adaptive technique combines the model tuning of Blumenthal and Cane (1989) and the model testing of Frankignoul et al. (1989). Although the procedure is developed in the context of the simplified SST model of Blumenthal and Cane (1989), it is general as long as the parameter dependence is linear, and is readily applicable to more complex models. The adaptive procedure requires little computation and programming, and is much simpler to implement than the adjoint method. However, since the effective degree of freedom of the error estimates is limited by the length of the sample, the number of parameters that can be tuned is limited.

The paper is organized as follows. In section 2, the physical model is presented and applied to a simulation of the tropical Atlantic SST from 1965 to 1986; the mean seasonal cycle is then validated using a multivariate approach. Section 3 presents the adaptive tuning procedure in a general way applicable to models with a linear parameter dependence. Several technical issues are addressed: construction of the error models, design of a data adaptive inverse technique, statistical uncertainties in parameter estimation, and model testing. In section 4, the adaptive procedure is applied to improve the SST model. The performance of the tuned model is discussed in section 5.

2. Modeling sea surface temperature variations

a. Ocean model and surface heat flux

The ocean model is that of Blumenthal and Cane (1989, hereafter BC) and has three major parts: a wind-driven model for velocity and pressure fields (Cane 1984), an advective/diffusive SST equation (Zebiak and Cane 1987), and a parameterization for surface heat fluxes (Seager et al. 1988, henceforth SZC). The dynamical variables are predicted with a linear, multimode equatorial beta-plane model with a surface mixed layer of constant depth $h = 35$ m, which adds a direct Ekman flow to the modal currents. The model has five vertical modes that are characteristic of mean tropical Atlantic conditions and have gravity wave speed of 2.36, 1.38, 0.89, 0.69, and 0.53 m s^{-1}, respectively. The model basin extends from 30°N to 20°S and has a simplified geometry; its resolution is 1° in longitude and 0.5° in latitude and the time step is one week. The equations are solved in the longwave approximation, making the model inappropriate for simulation of currents near the western boundary. In the following, we only consider the domain in Fig. 1, which should not be affected by the model's artificial boundaries.

The SST is determined from a nonlinear advective equation. The temperature is assumed to be uniform in the mixed layer, and determined from the net balance of horizontal advection, upwelling, horizontal diffusion, and surface heat exchanges:

$$\partial_t T + u\partial_x T + v\partial_y T + \gamma w(T - T_d)/h = \kappa(\partial_{xx} + \partial_{yy})T + Q/\rho C_p h, \quad (1)$$

where w is the vertical velocity at the mixed layer base in the case of entrainment, and zero otherwise; T_d is the temperature below the mixed layer, κ a horizontal diffusion coefficient; and Q the surface heat flux into the mixed layer, positive downward. Note that the upwelling term is usually written as $w(T - T_e)$, where T_e is the temperature of the water entrained into the mixed layer, but the two forms are equivalent if

$$T_e = (1 - \gamma)T + \gamma T_d. \quad (2)$$

The "entrainment efficiency" γ is an adjustable parameter that should be less than unity, as T_e is somewhere between T and T_d. As in SZC and BC, the parameterization of T_d is done in two parts: first the observed temperature at the mixed layer base is fit to the depth of the 20°C isotherm in the equatorial zone using the Levitus (1982) data, then the 20°C isotherm depth is fit to the model prediction of the thermocline depth (the latter fit is sensitive to the wind stress and has been adapted to each of the wind stress products below). During the course of this study, it was noticed that this simple parameterization could lead to temperature inversions, hence to a mixed layer warming by upwelling. Thus, the upwelling flux is set to zero when T_e is larger than T, which slightly improves the model performances (cf. the results reported in Frankignoul et al. 1993). Note that Zebiak and Cane (1987) and Cane et al. (1986) use a simpler parameterization of T_d.

The surface heat flux parameterization is that of SZC, which was designed to avoid using either the (poorly measured) air–sea temperature differences found in the bulk formulas or the artificial feedback to a prescribed climatological air temperature often imposed in ocean simulations. This parameterization solely includes effects that are externally imposed on the SST: the only measured variables used are wind speed v^a and fractional cloud cover C. The air temperature, to a large extent, is fixed by the SST and so has been eliminated. The heat flux is written as

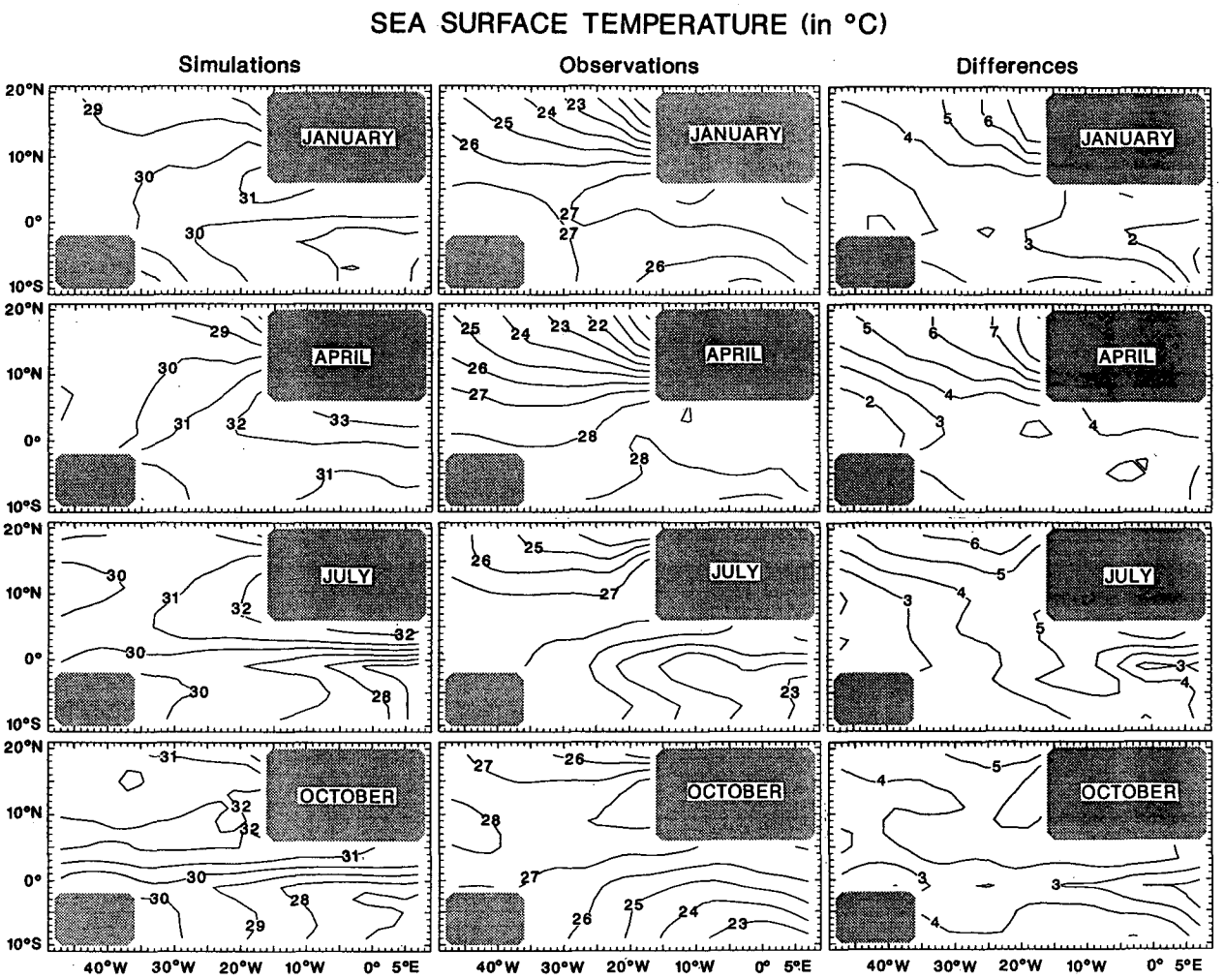

FIG. 1. (a) Mean SST (in °C) during January, April, July, and October as predicted by the multimode model using the a priori values of the parameters for the period 1965–1986. (b) Corresponding SST as derived from the observations by Servain et al. (1985). (c) Differences between simulations and observations.

$$Q = 0.94\, Q_0(1 - a_c C + a_\alpha \alpha) - \rho C_E L v^a a_{rh} q_s(T) - a_T(T - T_r). \quad (3)$$

The first term is the usual shortwave radiation formula, where Q_0 is the clear sky solar flux reduced at the surface by the effects of a constant surface albedo (0.06) and by the absorption and reflection of the atmosphere, which depends on cloud amount C and solar angle α. The second term represents the latent heat flux, computed from the standard bulk formula using a fixed percentage a_{rh} of the saturation humidity $q_s(T)$ as the evaporation potential $q_s(T) - q_s(T_{\text{air}})$; this assumes that the moisture content of the air has equilibrated with the ocean temperature, which is a reasonable assumption sufficiently far from the coasts. To compensate for the loss of variability in using monthly winds, the wind velocity v^a is not allowed to fall below 4 m s^{-1}. The smaller sensible heat flux and back radiation are simply modeled together in the last term as being proportional to T minus a constant reference temperature T_r. This formulation gives reasonable results in the tropics (SZC) but should not be used in midlatitudes (Liu and Niiler 1990).

In the SST equation and the heat flux formulation, there are a number of parameters not precisely known, but which were assigned a "reasonable" value by SZC. Here we assume that seven parameters are adjustable within reasonable ranges: the entrainment efficiency γ, the horizontal diffusion κ, and the heat flux parameters a_c, a_α, a_{rh}, a_T, and $a_T T_r$ in (3), which we represent below by the seven-dimensional vector $\mathbf{a}$. The a priori values of the tunable parameters, denoted by $\mathbf{a}_p$, are those of SZC for the tropical Pacific, namely $\gamma = 0.5$, $\kappa = 2\ 10^8$ m^2 s^{-1}, $a_c = 0.62$, $a_\alpha = 0.0019$, $a_{rh} = 0.3$, $a_T = 1.5$ W m^{-2} K^{-1}, and $T_r = 273.15$ K. The drag coefficient for the wind stress is not allowed to vary as in BC, since its uncertainty is simulated explicitly.

b. *Simulation of the tropical Atlantic SST climatology*

After spinup, the model is forced by a monthly wind stress derived from ship reports for the period 1964–1986. As described in Frankignoul et al. (1989, henceforth FDC), the wind stress was constructed by interpolating and smoothing monthly averaged fields of pseudo wind stress provided on a 2° × 5° grid by J. Servain. To simulate the drag coefficient uncertainty, we follow the Monte Carlo approach of Braconnot and Frankignoul (1993) and use five different, equally plausible drag coefficients in the bulk formula. They are calculated by prescribing a relative humidity of 80% and using either a constant air–sea temperature difference of −1°C (as in FDC), or a climatological monthly air–sea temperature difference derived from the COADS data [for the parameterizations of Large and Pond (1981), Liu et al. (1979), Smith (1988), and Isemer and Hasse (1987)]. To avoid smoothing, the monthly mean wind stresses were corrected to insure that linear interpolation on the model time step would not alter the original means (Duchêne 1989). Cloudiness data are of poorer quality, so that cloud cover is prescribed from the monthly climatology of Esbensen and Kushnir (1981), with an added normal noise of 0.1 standard deviation to crudely simulate its short space–time scale variability.

Ignoring the first year to eliminate the effects of the unknown initial conditions, we have 5 × 22 simulations of the SST annual cycle whose dispersion is representative of both the interannual variability and the drag coefficient uncertainty. The mean cycle of simulated SST is warmer than the observations, as illustrated in Fig. 1 for January, April, July, and October by a comparison with the mean SST over the same period calculated from the data of Servain et al. (1985).

The differences between the SST predictions and the observations are due to (i) errors in the atmospheric data (wind stress, cloud) and the SST observations; (ii) model shortcomings due to oversimplification of the physics; and (iii) poor choice of the model parameters. To assess the validity of the SST model, we must take (i) into account and minimize (iii) by an optimal tuning; remaining discrepancies should then point to the model deficiencies (ii).

Root-mean-square SST differences between the mean model response and observations on the 2° × 2° grid of the latter are given in Table 1 (left column), where we distinguish between annual mean, mean seasonal variations around the annual mean (hereafter the mean seasonal variability), and SST anomalies. The model-observation differences are large for the long-term mean, which is strongly affected by a 3.9°C mean bias. The mean seasonal variability seems better reproduced, with differences (rms 0.7°C) small compared to the amplitude of the observed signal (rms 1.3°C). On the other hand, the SST anomaly differences (after removing a linear trend to reduce the influence of artificial trends in the wind measurements—e.g., Cardone et al. 1990) are comparable (rms 0.7°C) to the observed anomalies (rms 0.5°C), with no significant correlation between observations and simulations, except north of 10°N (Fig. 2, top). These results underestimate the model skill, however, as they do not take data errors into account. For instance, the forcing uncertainties alone create an rms SST anomaly uncertainty of 0.5°C. Furthermore, the SST data of Servain et al. (1985), solely based on ship observations transmitted in real time, are noisy; the rms differences with the SST anomalies in the more elaborate but smoother SST product of Reynolds (1988), which also uses satellite data, are more than 0.3°C for the 1979–1988 period. This, however, is not sufficient to explain the poor model performance.

TABLE 1. Rms difference (in °C) between detrended observed and modeled (averaged response) SST for the yearly mean, the mean seasonal variability, and SST anomalies for various model versions in the 10°S–20°N region. The annual mean difference and the correlation between detrended observed and simulated monthly anomalies during 1965–1986 are given in italic.

($SST_{mod} - SST_{obs}$)	Before tuning	After tuning	After tuning (BC)
Annual mean	4.0	1.4	1.0
(mean bias)	*(3.9)*	*(1.1)*	*(0.5)*
Seasonal variability	0.69	0.75	0.74
Anomalies	0.67	0.65	0.64
(anomaly correlation)	*(0.26)*	*(0.23)*	*(0.26)*

A more quantitative estimation of the model performances taking into account some of the uncertainties in the oceanic observations and the atmospheric forcing, as well as their space–time correlations, has been made for the mean seasonal cycle. Following the multivariate approach of FDC, we calculate the misfit (signal-to-noise ratio)

$$T^2 = (\langle \bar{\mathbf{T}} \rangle - \overline{\mathbf{T}_o})' \mathbf{D}^{-1} (\langle \bar{\mathbf{T}} \rangle - \overline{\mathbf{T}_o}), \qquad (4)$$

where $\langle \bar{\mathbf{T}} \rangle$ and $\overline{\mathbf{T}_o}$ describe the mean seasonal cycle of modeled and observed SST, respectively, the vector space including all grid points (on the observational grid) and the 12 months. The overbar denotes the 22-year mean and angle brackets the average over the five 22-year runs; the prime indicates vector transpose, and $\mathbf{D}$ is the error covariance matrix of $(\langle \bar{\mathbf{T}} \rangle - \overline{\mathbf{T}_o})$.

To estimate $\mathbf{D}$, we use the five 22-year samples. A first contribution to $\mathbf{D}$ arises from the uncertainties in the mean seasonal variations that are due to interannual variability and *nonsystematic* observational errors of SST, wind, and cloud cover, which affect randomly each realization of the observed and modeled seasonal cycles. Assuming for simplicity that each year, denoted by the superscript t, is statistically independent, which is acceptable for the tropical Atlantic, these uncertainties are estimated by

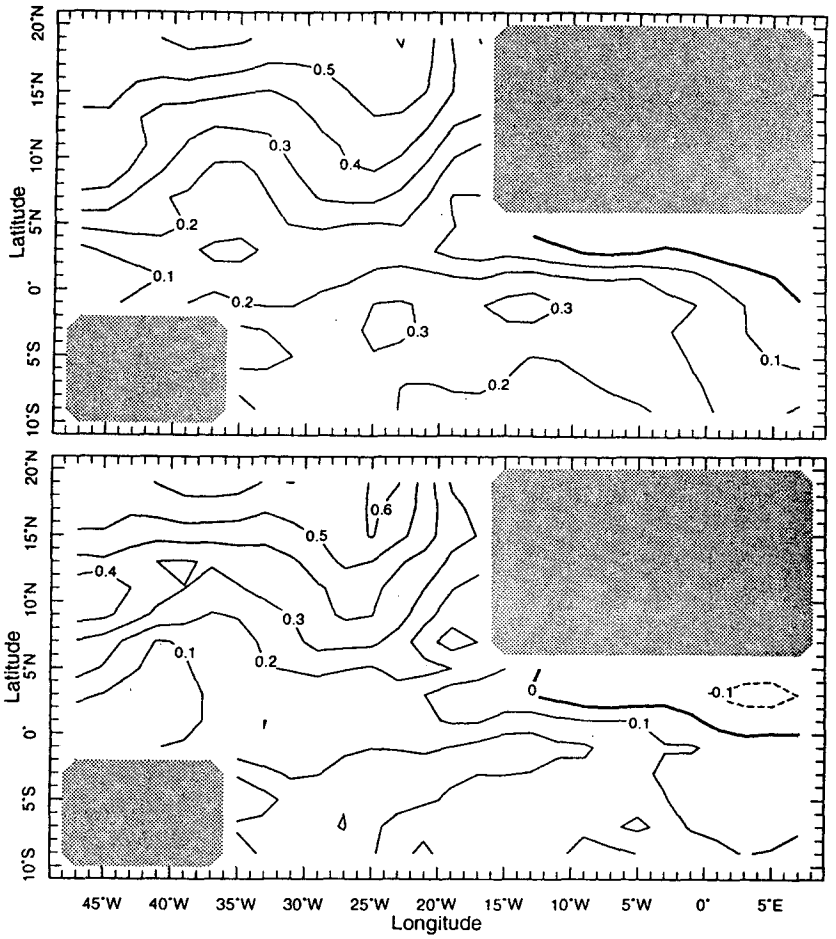

FIG. 2. Correlation coefficient between detrended observed and modeled SST anomalies over the 1965–1986 period for the initial (top) and the tuned (bottom) model. The 5% significance level is approximately 0.3.

$$\mathbf{D}_1 = \frac{1}{21*22}\sum_{t=1}^{22}[(\langle \mathbf{T}\rangle^t - \mathbf{T}_o^t) - (\overline{\langle \mathbf{T}\rangle} - \overline{\mathbf{T}_o})] \times [(\langle \mathbf{T}\rangle^t - \mathbf{T}_o^t) - (\overline{\langle \mathbf{T}\rangle} - \overline{\mathbf{T}_o})]' \quad (5)$$

with 21 degrees of freedom. Another contribution to **D** represents the drag coefficient uncertainty and is estimated by

$$\mathbf{D}_2 = \frac{1}{4*5}\sum_{i=1}^{5}(\bar{\mathbf{T}}^i - \langle \bar{\mathbf{T}}\rangle)(\bar{\mathbf{T}}^i - \langle \bar{\mathbf{T}}\rangle)' \quad (6)$$

with 4 degrees of freedom if we assume that the responses to the different forcing, denoted by the superscript i, are statistically independent. Since the two sources of errors are independent, $\mathbf{D} = \mathbf{D}_1 + \mathbf{D}_2$, with approximate degrees of freedom ν (for instance, $\nu \approx 24$ in the case of the seasonal variations below). Not represented in **D** are *systematic* observational errors (e.g., incorrect Beaufort scale, SST biases), lack of high-frequency variability, and limited resolution of the wind stress curl.

Since the dimension of the SST field is much larger than the degrees of freedom of **D**, the misfit (4) is calculated in a truncated space that is sufficiently small to calculate **D** reliably while representing the main space–time patterns of $(\langle \bar{\mathbf{T}}\rangle - \overline{\mathbf{T}_o})$. As described in appendix A, the annual mean is appraised in a one-dimensional space and the mean seasonal variability in a 12-dimensional space.

If the SST fields are multinormal, the null hypothesis that the model response to the true forcing is equal to the true SST (no model errors) can be tested as the test statistic (4) is then Hotelling's single-sample T^2 statistic with dimension p (the dimension of the space) and approximate degrees of freedom ν. Its distribution is related to the F-distribution with p and $\nu + 1 - p$ degrees of freedom by $T^2 = \nu p/(\nu + 1 - p)\,F$ (e.g., see Morrison 1976). The results, given in Table 2 (left), show that T^2 is much larger than the critical value at the 5% level (right) and the null hypothesis rejected. They confirm that the model performances are rather poor. Although systematic observational errors in SST,

TABLE 2. Misfit between model and observations for the yearly mean (t-value) and the mean seasonal variability in the 20°N–10°S region, before and after tuning. The tuning is done in the 10°S–20°N region. The critical values for rejecting the null hypothesis of no model error are also given.

Misfit	Before tuning	After tuning	After tuning (BC)	Critical value (5% level)
Long-term annual mean	15.2	7.4	7.0	2
Seasonal variability	1751	1636	1666	35

cloud cover, and wind stress have not been considered in the test, the data uncertainties are likely to be insufficient to explain the model–observation discrepancies, which must be mainly attributed to model shortcomings and poor parameter tuning.

3. An adaptive procedure for model tuning

a. Linear model corrections

To see how the tunable parameters enter the calculation of the SST in the numerical model, it is convenient to write Eq. (1) in matrix form

$$\mathbf{L}(\mathbf{T}) + \mathbf{M}(\mathbf{T})\mathbf{a}_p = 0, \tag{7}$$

where the vector $\mathbf{T}$ represents temperature at all the points in space and time where a model solution has been obtained, $\mathbf{a}_p = (\gamma, \kappa, a_c, a_\alpha, a_{rh}, a_T, a_T T_r)$ is the vector of a priori parameter values, $\mathbf{M}(\mathbf{T})$ and $\mathbf{L}(\mathbf{T})$ are linear operators determined at all space/time points by retaining the terms of the model equations (1) and (3) that are and are not affected by parameter changes, respectively. Specifically, the ith row of $\mathbf{L}(\mathbf{T})$ includes the contribution at space/time point i from

$$\partial_t T + u\partial_x T + v\partial_y T - 0.94Q_0,$$

while the ith row of $\mathbf{M}(\mathbf{T})$ correspondingly represents the transpose of the terms

$$\begin{bmatrix} w(T - T_d)/h \\ -(\partial_{xx} + \partial_{yy})T \\ 0.94Q_0C \\ -0.94Q_0\alpha \\ -\rho C_E L v^a q_s(T) \\ T \\ -1 \end{bmatrix}.$$

Both $\mathbf{L}$ and $\mathbf{M}$ depend on the atmospheric forcing, which is imperfectly known, so even if the model was perfect and the uncertain parameters optimally chosen, the model predictions would differ from the observations.

Since SST is a relatively well-measured variable, we follow BC and estimate the "corrective heat flux" $\delta\mathbf{q}$ that, for the a priori values of the uncertain model parameters, would be needed to make the model SST match the observed SST exactly. To do so, we run the model again, using the observed SST, denoted by $\mathbf{T}_o$, instead of the calculated one, after interpolation on the model grid. Equation (7) is then only satisfied by adding a "heat flux correction" $\delta\mathbf{q}$:

$$\mathbf{L}(\mathbf{T}_o) + \mathbf{M}(\mathbf{T}_o)\mathbf{a}_p + \delta\mathbf{q} = 0. \tag{8}$$

As expected from the limited SST agreement, the heat flux correction is rather large, showing that additional cooling would be needed for realistic simulations (Fig. 3a).

Since $\delta\mathbf{q}$ depends linearly on the tunable model parameters, the estimation of their optimal value can be formulated as an inverse problem

$$\delta\mathbf{q} = \mathbf{M}(\mathbf{T}_o)\delta\mathbf{a}, \tag{9}$$

where $\delta\mathbf{a} = (\delta\gamma, \delta\kappa, \ldots, \delta a_T T_r)$ represents the parameter changes that minimize the heat flux correction $\delta\mathbf{q}$, yielding

$$(\delta\mathbf{q})_{\min} = \delta\mathbf{q} - \mathbf{M}(\mathbf{T}_o)\delta\mathbf{a}. \tag{10}$$

A good estimator of the parameter correction $\delta\mathbf{a}$ must take errors into account, as well as our knowledge of the expected parameter range.

There are many sources of errors in the estimates appearing in (9). The wind stress and cloud data used to force the model have significant errors, resulting in model response uncertainties with large correlation scales, particularly in the equatorial waveguide. The observed SST is noisy as well, although to a lesser extent. When the best-fit calculation is based on a mean seasonal cycle as in this paper, there are also sampling errors that reflect the interannual variability and have large correlation scales. Finally, there are "irreducible" modeling errors inherent in the ocean model formulation, for example, errors due to subgrid-scale phenomena, or to the oversimplification of the ocean dynamics and the air–sea fluxes. These cannot be expected to be reduced by model tuning. The modeling errors (called system errors in the Kalman filter literature) thus represent the errors that would exist if there were no observational errors and the uncertain parameters were at their true value.

Tarantola (1987) discusses the general inverse problem in the case of an inaccurate theory, using a Bayesian viewpoint. When the forward problem is linear as in (8) and there are Gaussian modeling errors in $\mathbf{M}$, described by the covariance $\mathbf{C}_\mathbf{T}$, the solution of the inverse problem takes a simple form if the observational errors in $\delta\mathbf{q}$ are Gaussian and statistically independent of the modeling errors. If the a priori value of the parameter correction $\delta\mathbf{a}$ is zero, as in the present case, the optimal solution is given by the minimum of the misfit function

$$S(\delta\mathbf{a}) = [(\mathbf{M}\delta\mathbf{a} - \delta\mathbf{q})'\mathbf{C}^{-1}(\mathbf{M}\delta\mathbf{a} - \delta\mathbf{q}) + \delta\mathbf{a}'\mathbf{C}_a^{-1}\delta\mathbf{a}]/2 \tag{11}$$

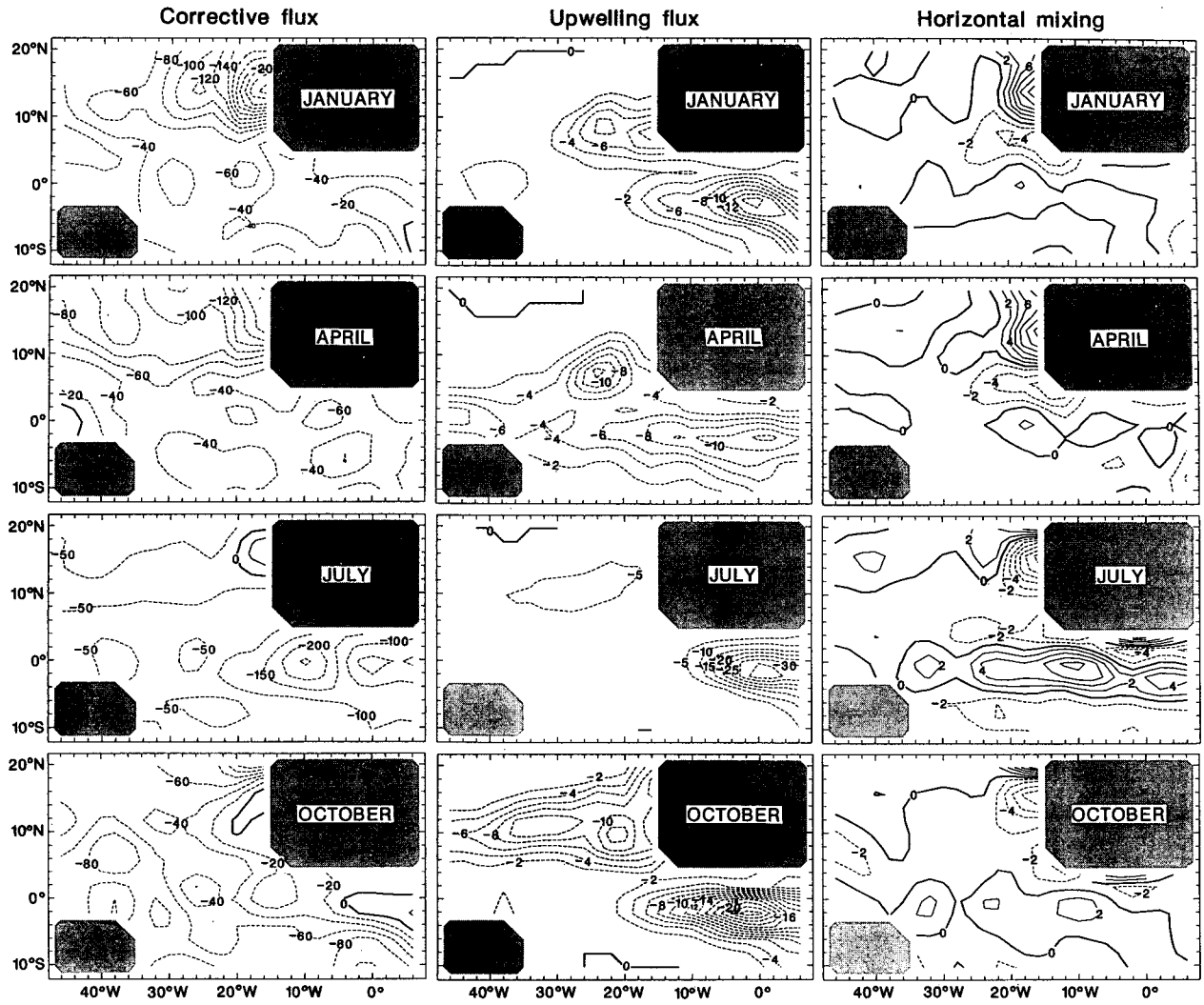

FIG. 3. (a) Mean heat flux correction (in W m^{-2}) during January, April, July, and October for the period 1965–1986, when using the a priori values of the model parameters. Corresponding values of (b) the upwelling flux, (c) horizontal diffusion, (d) cloud factor, (e) solar angle, and (f) latent heat flux, as scaled *after* tuning.

with $\mathbf{C} = \mathbf{C}_T + \mathbf{C}_o$, where $\mathbf{C}_o$ is the error covariance matrix of the observations $\delta\mathbf{q}$, and the covariance matrix $\mathbf{C}_a$ describes the a priori uncertainty of $\delta\mathbf{a}$. The solution is

$$\delta\mathbf{a} = (\mathbf{M}'\mathbf{C}^{-1}\mathbf{M} + \mathbf{C}_a^{-1})^{-1}\mathbf{M}'\mathbf{C}^{-1}\delta\mathbf{q}. \quad (12)$$

Blumenthal and Cane followed this formalism, assuming for simplicity that the observational noise only affected the model matrix $\mathbf{M}$, and the modeling error only the heat flux correction $\delta\mathbf{q}$. On the basis of order of magnitude estimates, they used constant rms errors of 10 and 35 W m^{-2} with a simple exponential decay for the modeling and total error, respectively.

There are a number of simplifications in BC's approach: 1) As shown by (8), both $\delta\mathbf{q}$ and $\mathbf{M}$ depend on the input data (e.g., the surface wind stress affects both the heat exchanges and the ocean dynamics), hence they are both affected by the data uncertainties and the modeling errors. The errors in $\delta\mathbf{q}$ and $\mathbf{M}$ are thus *not statistically independent,* and the model matrix really is a stochastic regression matrix. Unfortunately, ordinary and generalized least-squares estimators are in general not consistent in this case of nonlinear coupling between model and data errors (e.g., Mardia et al. 1979; Judge et al. 1988). Alternative estimating procedures that will give consistent estimators have been used in econometrics but they are not general and are hard to apply to the present problem. 2) The error models used by BC are highly idealized and could be improved. Since the results of the tuning are sensitive to the assumed error models, we adopt a more elaborate strategy to achieve a refined estimate.

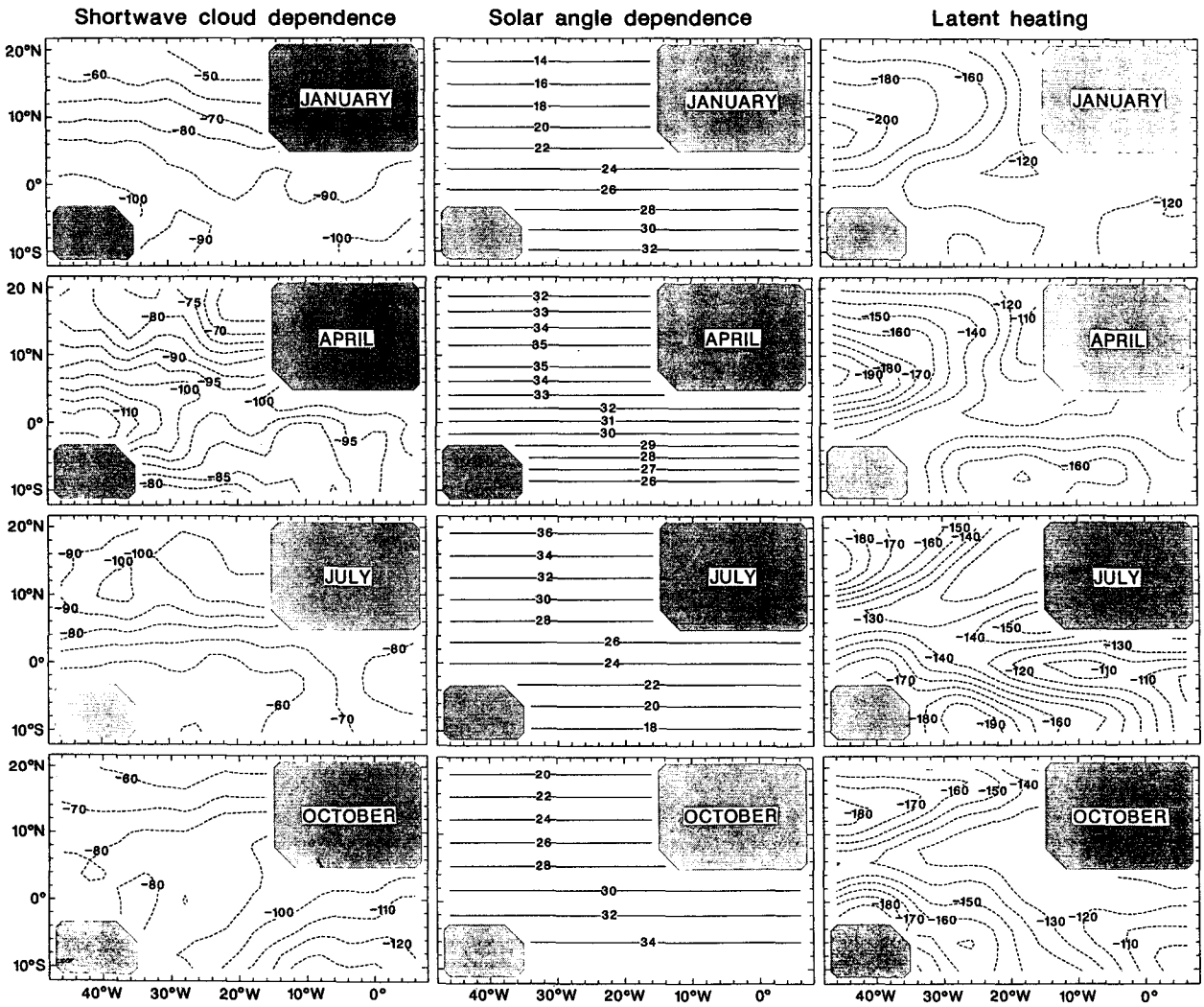

FIG. 3. (*Continued*)

b. The adaptive procedure

The correlation scales of the model response errors due to forcing and SST uncertainties are large and complex, hence difficult to represent a priori. However, they can be estimated by performing the optimization on the mean seasonal cycle, which is least noisy, and using the dispersion of the model seasonal responses as independent information to construct a more realistic model for the observational errors.

Assuming that the parameters do not vary in time, we can write for each year t (here $t = 1, 22$) and for each forcing i (here $i = 1, 5$), denoted by the upper index, that the linear model (9) holds

$$\mathbf{L}^{t,i}(\mathbf{T}_o^t) + \mathbf{M}^{t,i}(\mathbf{T}_o^t)\mathbf{a}_p + \delta\mathbf{q}^{t,i} = 0. \quad (13)$$

Denoting long-term sample means by an overbar and the mean over the different forcing by an angle brace, we write relation (8) under the form

$$\langle\overline{\mathbf{L}(\mathbf{T}_o)}\rangle + \langle\overline{\mathbf{M}(\mathbf{T}_o)}\rangle\mathbf{a}_p + \langle\overline{\delta\mathbf{q}}\rangle = 0. \quad (14)$$

The errors in (13) and (14) are due to forcing and SST uncertainties, and to model inadequacies. Since the nonsystematic errors associated with data uncertainties should decrease with averaging, they will be smaller in (14) than in (13). On the other hand, the systematic errors, primarily due to model deficiencies, should remain practically unchanged.

Let us write the parameter estimation as the linear statistical model

$$\langle\overline{\delta\mathbf{q}}\rangle = \langle\bar{\mathbf{M}}\rangle\delta\mathbf{a} + \langle\bar{\mathbf{e}}\rangle, \quad (15)$$

where $\langle\bar{\mathbf{e}}\rangle$ represents the random errors, assumed to be Gaussian, with zero mean and unknown true covariance matrix $\mathbf{C}$. Because of the statistical dependence between $\langle\overline{\delta\mathbf{q}}\rangle$ and $\langle\bar{\mathbf{M}}\rangle$, an estimate of $\delta\mathbf{a}$ is required before one may estimate the random errors from the sample. Thus, an adaptive approach is used, where the

estimates of the observational error covariance and the model parameters are updated as part of an iterative procedure. If we have a first estimate of $\delta\mathbf{a}$, say $\delta\mathbf{a}_0$, which we will take equal to zero, then we can estimate for each year t the mean error over the different forcing, $\langle \mathbf{e}_1^t \rangle$, by

$$\langle \mathbf{e}_1^t \rangle = \langle \delta\mathbf{q}^t \rangle - \langle \mathbf{M}^t \rangle \delta\mathbf{a}_0. \tag{16}$$

A first sample estimate of the error covariance matrix associated with the random wind, cloud, and SST errors is

$$\mathbf{S}_{r1} = \frac{1}{21*22} \sum_{t=1}^{22} (\langle \mathbf{e}_1^t \rangle - \overline{\langle \mathbf{e}_1 \rangle})(\langle \mathbf{e}_1^t \rangle - \overline{\langle \mathbf{e}_1 \rangle})', \tag{17}$$

where we have assumed for simplicity that observations are independent at yearly intervals. We can also estimate for each forcing i the long-term mean error, $\overline{\mathbf{e}_1^i}$, by

$$\overline{\mathbf{e}_1^i} = \overline{\delta\mathbf{q}^i} - \overline{\mathbf{M}^i}\delta\mathbf{a} \tag{18}$$

and a first sample estimate of the error covariance matrix associated with the drag coefficient uncertainties is

$$\mathbf{S}_{f1} = \frac{1}{4*5} \sum_{i=1}^{5} (\overline{\mathbf{e}_1^i} - \langle \overline{\mathbf{e}_1} \rangle)(\overline{\mathbf{e}_1^i} - \langle \overline{\mathbf{e}_1} \rangle)'. \tag{19}$$

A first estimate of the error covariance associated with the observational uncertainties, say $\mathbf{S}_{o1}$, can then be obtained by

$$\mathbf{S}_{o1} = \mathbf{S}_{r1} + \mathbf{S}_{f1}$$

and it can be used to compute an estimated generalized least-squares estimate of $\delta\mathbf{a}$, say $\delta\mathbf{a}_1$. As in (12), we incorporate the modeling errors and our a priori knowledge on the model parameters:

$$\delta\mathbf{a}_1 = (\bar{\mathbf{M}}'\mathbf{S}_1^{-1}\bar{\mathbf{M}} + \mathbf{C}_a^{-1})\bar{\mathbf{M}}'\mathbf{S}_1^{-1}\overline{\delta\mathbf{q}}, \tag{20}$$

with

$$\mathbf{S}_1 = \mathbf{S}_{o1} + \mathbf{C}_T. \tag{21}$$

The procedure is repeated by using $\delta\mathbf{a}_1$ in (16) to get an improved estimate $\mathbf{S}_2$, leading to the parameter correction $\delta\mathbf{a}_2$, and so on. If the initial estimate $\delta\mathbf{a}_0$ represents a reasonable first guess and the inverses in (20) are well conditioned, the procedure should converge rapidly. The end result is a data error structure consistent with the multiyear model run, and thus presumably a better parameter estimation.

The error model $\mathbf{S}_n$ represents most of the nonsystematic data and model errors; it also includes such data errors as artificial trends in wind and SST data. The true interannual variability is not treated as an error since it appears in both $\delta\mathbf{q}^t$ and $\mathbf{M}^t$ in (16). The weighting in the least-squares fit is therefore based on data noise and uncertainties and it takes into account, at least approximately, the lack of independence between $\bar{\mathbf{M}}$ and $\overline{\delta\mathbf{q}}$. On the other hand, the weighting is not affected by the systematic errors that recur every year. Model deficiencies or systematic data biases must be dealt with explicitly.

Because of the limited sample, the error covariance matrix $\mathbf{S}_{on}$ is of strongly reduced rank and the inverse of $\mathbf{S}_n$ dominated by unreliable information. Hence, the problem is ill conditioned. To circumvent the difficulty, we reduce the dimension of the fields and tune the model in a highly truncated space, using furthermore a singular value decomposition to perform some of the matrix inversions. The construction of the reduced base is more difficult than in FDC as both the heat flux correction $\langle \overline{\delta\mathbf{q}}(x, t) \rangle$ and the model $\langle \bar{\mathbf{M}}(x, t) \rangle$ must be properly represented. There are several ways to define the basis vectors of the truncated space, and we found them to yield rather similar results as long as the problem remains well conditioned. Also, we verified (on a small domain) that tuning in reduced space is very similar, but not identical, to tuning in gridpoint space.

The iterative method is implemented in reduced space: for each forcing, each individual year is projected onto the reduced base, thereby defining a reduced heat flux correction and a reduced model matrix. By projection, a reduced modeling error matrix is also constructed. The sample error covariance matrix associated with the observational uncertainties and the optimal parameter corrections are then directly calculated in reduced space, so that the computational costs are very limited.

As discussed in Tarantola (1987), the probable errors in the parameter estimates are described by the a posteriori covariance operator

$$\hat{\mathbf{C}}_a = (\bar{\mathbf{M}}'\mathbf{S}_n^{-1}\bar{\mathbf{M}} + \mathbf{C}_a^{-1})^{-1}. \tag{22}$$

The diagonal terms give the variance of the a posteriori parameters, and the off-diagonal terms their covariances. The correlation matrix shows whether the posterior uncertainties are uncorrelated. A small correlation indicates that two parameters have been resolved independently by the dataset; a large one that only some linear combination of the parameters is resolved. A resolution operator given by

$$\mathbf{R} = \mathbf{I} - \hat{\mathbf{C}}_a\mathbf{C}\mathbf{a}^{-1} \tag{23}$$

can also be calculated; it shows the relative contributions of the dataset and the a priori information in resolving the parameters.

c. *Model testing*

For simplicity, BC based their assessment of model validity on the sole basis of visual comparisons: if the residual errors were of the order of the a priori ones (35 W m^{-2}, as mentioned before), the SST model was judged to be valid; if that figure was exceeded, it was

concluded that the SST model had larger errors than assumed. Such "univariate" approach ignores useful information on correlation scales; if the residuals had large space/time patterns, they would not be consistent with the error models assumed by BC and the SST model might not be acceptable, even though it would appear so in univariate tests.

To take the multidimensional aspects of the fields into account, we generalize a multivariate test derived by Tarantola (1987). For true error covariance matrix $\mathbf{C}$, the minimum of the misfit function (11) is given by

$$S(\delta\mathbf{a}) = \overline{\delta\mathbf{q}}'(\bar{\mathbf{M}}\mathbf{C}_a\bar{\mathbf{M}}' + \mathbf{C})^{-1}\overline{\delta\mathbf{q}} \qquad (24)$$

and it should have a χ^2 distribution with degrees of freedom η given by the dimension of the reduced data space (because the rank of the operator $\mathbf{M}$ equals the dimension of the parameter space). Hence, (24) can be compared to the critical χ^2_η values at a given level of significance. If it is too large, then some of the assumptions are unlikely to be acceptable.

The test is generalized to our model-tuning procedure by replacing the true error covariance $\mathbf{C}$ in (24) by the sample estimate $\mathbf{S}_n = \mathbf{S}_{dn} + \mathbf{C}_T$. The null hypothesis that the only errors besides the observational ones are the modeling errors can be tested since the test statistic (24) is then distributed as Hotelling's T^2 with degrees of freedom η (the reduced dimension) and τ (the equivalent degrees of freedom of $\mathbf{S}_n$). If (24) exceeds the critical value, the errors are underestimated. The most likely interpretation is that the modeling errors have been underestimated, since, except for possible biases, the observational uncertainties are represented by an error model which is, by construction, consistent with the available observations. This is a statistical model-testing procedure analogous to that of FDC, but more general, as model tuning is included in the test. By separating model testing and tuning, FDC ignored the possible influence of model tuning on the statistics, which can be an oversimplification.

Since the multivariate character of the fields is represented in our method, we expect the test to be more stringent than the comparison done by BC. Model flaws should thus be more apparent, opening the way to further model improvement.

4. Tuning the tropical Atlantic SST model

a. Results of the adaptive procedure

The monthly values of $\delta\mathbf{q}^{t,i}$ and $\mathbf{M}^{t,i}$ are first spatially smoothed with a $5° \times 5°$ running average. The fit is then done in the region between 10°S and 20°N by considering January, April, July, and October, which are representative of the various SST regimes. The data dimension p is 322 (number of grid points) $\times$ 4 (number of months) = 1288.

The mean heat flux correction $\langle\overline{\delta\mathbf{q}}\rangle$ is represented in Fig. 3a. The rms value is large (70 W m^{-2}), and negative values in excess of -100 W m^{-2} are found off Africa and in the Gulf of Guinea, mostly where the largest SST differences are observed, although there is no one-to-one correspondence between the two fields. The tuning can be viewed as determining the best fit of the mean heat flux correction vector in Fig. 3a by the seven column vectors of $\langle\overline{\mathbf{M}(\mathbf{T}_o)}\rangle$, which are represented in Figs. 3b–f, taking their errors into account. To facilitate the interpretation, the amplitudes of the latter are given in W m^{-2} by using the results of the inverse calculation [i.e., each structure has been multiplied by $(\mathbf{a}_p + \delta\mathbf{a}_n)$]. The "upwelling pattern" (Fig. 3b) has a large signal in the Gulf of Guinea peaking near 2°S, 0°, with maximum amplitude during the upwelling season in July; a smaller signal is also seen in the ITCZ with maximum amplitude off Africa, except in April where the peak is near 22°W. As the observed SST fields of Servain et al. (1985) are relatively smooth (see Fig. 1b), the meridional scale of the "diffusion pattern" (Fig. 3c) is only slightly smaller than that of upwelling, with maximum amplitude off the African coast and, in July, along the equator. The "cloud pattern" (Fig. 3d) has broader scales and larger amplitudes, and its seasonal changes reflect those of Q_0 and C. The "evaporation pattern" (Fig. 3f) has equally large amplitudes, a large meridional scale, and strong zonal gradients. Additional patterns are the "insolation pattern" (Fig. 3e), a constant, and the observed SST pattern, with small amplitudes.

The data compression is done by working in the space defined by orthonormalizing the eight vectors consisting of $\langle\overline{\delta\mathbf{q}}\rangle$ and the seven column vectors of $\langle\bar{\mathbf{M}}\rangle$. As the dimension η of the subspace is the number of adjustable parameters plus one (8 in the present case), the inverse problem remains formally overdetermined. As described in section 3, $\delta\mathbf{q}^{t,i}$ and $\mathbf{M}^{t,i}$ are projected onto the reduced base for each year t, and the sample error covariance matrix $\mathbf{S}_{dn}$ directly estimated in reduced space at each iteration n. Since $\mathbf{S}_{dn}$ has limited degrees of freedom, its elements are inaccurately known (large sampling errors) and the condition number of the matrix is very large. Lacking precise information on the modeling errors, we use BC's model (Gaussian noise with short correlation scales), but double the rms error to 20 W m^{-2}, which seems more appropriate to the simplicity of the SST and heat flux models. This modeling error matrix is not sufficient to insure good conditioning, so a singular value decomposition is used to invert $\mathbf{S}_n$ in (20). In practice, we apply a taper to the error covariance matrices (17) and (19), which is an estimate of the sampling uncertainty of their elements (taken to be the trace divided by the matrix dimension) derived from the χ^2 distribution. As the other inverses in the calculation are well conditioned, the results are not too sensitive to the tapering.

For simplicity, we use zero for the initial parameter correction, $\delta \mathbf{a}_0 = 0$, but the results are similar when using a different initial value. Convergence is reached in two or three iterations, with the largest changes occurring after the first iteration. Figure 4 shows the a priori and a posteriori values of the adjustable parameters with twice their standard deviation (an approximation to the 95% confidence interval). Two parameters undergo large changes and reach values largely outside their expected range (Fig. 4): the upwelling efficiency γ, which strongly decreases, and the horizontal diffusion that remains positive, but not significantly different from zero at the 5% level. Both parameters are well resolved by the dataset (Table 3) and independently resolved (uncorrelated), so their changes provide useful information on the model physics. The reason for the large changes is apparent in the flux patterns. In the upwelling zone off Africa, the model is too warm and cooling is needed in the first part of the year (Fig. 3a), but horizontal mixing heats the offshore waters (Fig. 3c) and it thus tends to be decreased by the fit. At the same time, cooling by upwelling takes place too far south and offshore to provide the required cooling, particularly in April. In the Gulf of Guinea, horizontal mixing is strong in July but again out of phase with the heat flux correction, while the upwelling flux is alternatively in (January, April, July) and out (October) of phase, and too far east at its maximum in July, so it also tends to be reduced. The fit in the two upwelling regions thus inhibits the upwelling-induced SST cooling, as shown by the small magnitude of the upwelling flux in Fig. 3b. Note that horizontal diffusion is a coarse parameterization that represents a variety of local phenomena (including the Reynolds' heat flux during the instability wave season), so that its small tuned value remains acceptable, while the strong decrease in upwelling efficiency is more puzzling. Although the changes in the cloud factor a_c and the latent heat flux a_{rh} are also well resolved by the dataset, they are not statistically significant at the 5% level, which suggests that the a priori choices were good, needing only little adjustment. Moreover, these changes are difficult to interpret as the two parameters are not independently resolved and are anticorrelated, and correlated with the three remaining parameters, a_α, a_T and $a_T T_r$. The latter are poorly resolved by the dataset, and their changes are small (significant for $a_T T_r$). The total number of parameters resolved by the dataset, given by the trace of the resolution matrix, is 4, while 3 remain determined by the a priori choice. Of interest is that the trace of $\mathbf{S}_{dn}$ decreases by 22% during the iterations, so that the interannual variance of the heat flux correction is reduced by the optimization, suggesting indeed some model improvement.

Figure 5a shows the heat flux correction (10) after tuning. The amplitudes are smaller than in Fig. 3a: the rms value has dropped from 70 to 32 W m^{-2} and the space–time average to -8 W m^{-2} (Table 4), suggesting

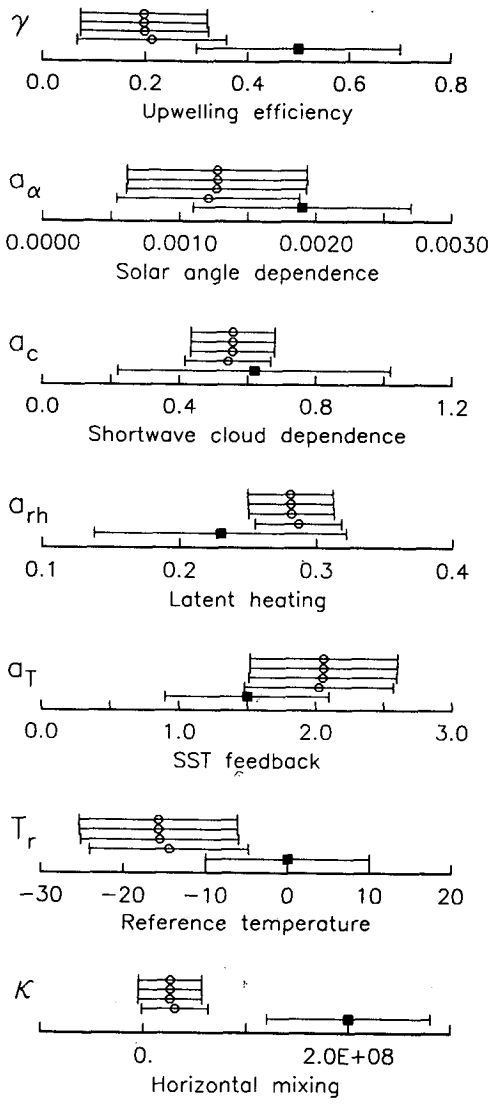

FIG. 4. Changes in the parameters as a function of the number of iterations. The solid square is the initial choice and the open circles proceeding upward are the iterations. The error bars represent the 95% confidence intervals.

that the warm SST bias seen in Fig. 1a should be somewhat corrected, but that the tuning may be unable to do more. Heat flux corrections larger than 100 W m^{-2} can still be seen off the North African coast during winter and in the equatorial upwelling region during summer, which seems much too large to be explainable by the assumed modeling errors and the data uncertainties. Blumenthal and Cane noted that the model does not work well near the African coasts, in part because the simplification of the coastal geometry limits the accuracy of surface currents and coastal upwelling and because the parameterization for the interior temperature T_d is based on equatorial data; also, the parameterization of latent heat flux assumes that the air

TABLE 3. Value of the adjustable parameters before and after tuning. Bold characters indicate parameters that change significantly at the 5% level. The resolution is derived from the diagonal terms of the resolution operator (23).

	Parameter		Resolution by	
	A priori value	After tuning	Observations (%)	A priori info (%)
Upwelling efficiency	**0.5**	**0.2**	62	38
Cloudiness	0.62	0.56	91	9
Latitude dependence	1.9×10^{-3}	1.3×10^{-3}	31	69
Latent heat flux	0.23	0.27	88	12
SST negative feedback	1.5	2.1	19	81
Constant flux	**0**	**−16**	8	92
Horizontal diffusion	$\mathbf{2 \times 10^8}$	$\mathbf{0.3 \times 10^8}$	85	15

temperature has equilibrated with the SST, which is unlikely near the African coast. These model flaws seem correctable, but the large heat flux correction in the equatorial upwelling region is more perturbing since the model was designed to simulate the equatorial variability.

To verify the consistency of the inverse calculation, we apply the statistical test of section 3c. Although the critical value of the test statistic (24) is difficult to establish as the total error covariance is the sum of a sample one and an (assumed to be) true one, upper and lower bounds can easily be found. For true covariances, the critical value, given by the χ^2 distribution with 8 degrees of freedom (the dimension of the reduced space), would be 16 at the 5% level (lower bound). For sample covariance matrices, it would be given by Hotelling's T^2 and equal to 32 (upper bound). The test value is 336, which largely exceeds both critical values. It must be concluded that the modeling errors have been strongly underestimated. In particular, there are large modeling biases, not only random modeling errors as assumed.

Since the tuning minimizes the heat flux correction (more precisely a weighted form of it), it is of interest to verify whether the SST predictions have been improved by the parameter changes. The tuned model was thus run with the same forcing fields as before. As expected, a more realistic SST field is obtained (Fig. 5b), although model–observation differences of a few degrees can still be seen in the upwelling region off Africa during the first part of the year and in the Gulf of Guinea during the second part (Fig. 5c). Table 1 (center) suggests that the model improvements are limited to a strong decrease of the warm SST bias, even though it still averages to 1.1°C. Although the mean bias can be reduced further to 0.4°C by performing the fit on the annual means only (not shown), it may reflect the lack of statistical independence between $\delta\mathbf{q}$ and $\mathbf{M}$, which leads to biased estimators, as mentioned above. The mean seasonal variability seems slightly less realistic, and modeled and observed SST anomalies remain largely uncorrelated (the reduction in upwelling efficiency has even increased the areas with negative correlations, as seen in Fig. 2, bottom). The more accurate multivariate statistical test (Table 2) confirms the improved model ability at simulating the yearly mean and shows that the model performances for the mean seasonal variability have in fact been improved, although not significantly. The multivariate test also indicates that the SST model remains largely inconsistent with the observations: the tuning is unable to compensate the model shortcomings.

The sensitivity of the tuning results to various factors has been investigated. The dependence on the details of the procedure (tapering, data reduction) is relatively small, as long as the problem remains well conditioned. When the fit is done on the annual mean or on a particular month only, the tuning results are similar (i.e., within the error bars), except for the upwelling efficiency and, to a lesser extent, horizontal diffusion, although a decrease is found in all cases. The decrease in upwelling efficiency is particularly noteworthy, as the values for the four individual months fail to bracket that for the four months together (Fig. 6). Rather, the decrease is a nearly linear function of the *resolution* by the dataset. The upwelling flux is even nearly cancelled when no a priori constraint is prescribed (but several parameters then take unacceptable values). The observations are thus inconsistent with the predicted upwelling flux and clearly stress the inadequacy of the model representation of SST cooling by upwelling.

The consistency of the adaptive tuning procedure has also been investigated, using a set of artificial data: the "observed" SST was simply taken as the averaged SST predicted by the tuned model when it was forced by the five wind stress fields. The tuning was then performed for various choices of the a priori parameters $\mathbf{a}_p$, to evaluate how well the adaptive procedure would recover the original parameters, and whether the results depended on the a priori choices. As illustrated in Fig. 7, the true value of the parameters was in all well-resolved cases (including the upwelling efficiency) within the error bars of the tuned values, and the latter were insensitive to the a priori choice, except for the horizontal diffusion κ. The latter is well-resolved by the dataset and (nearly) independently resolved, yet its es-

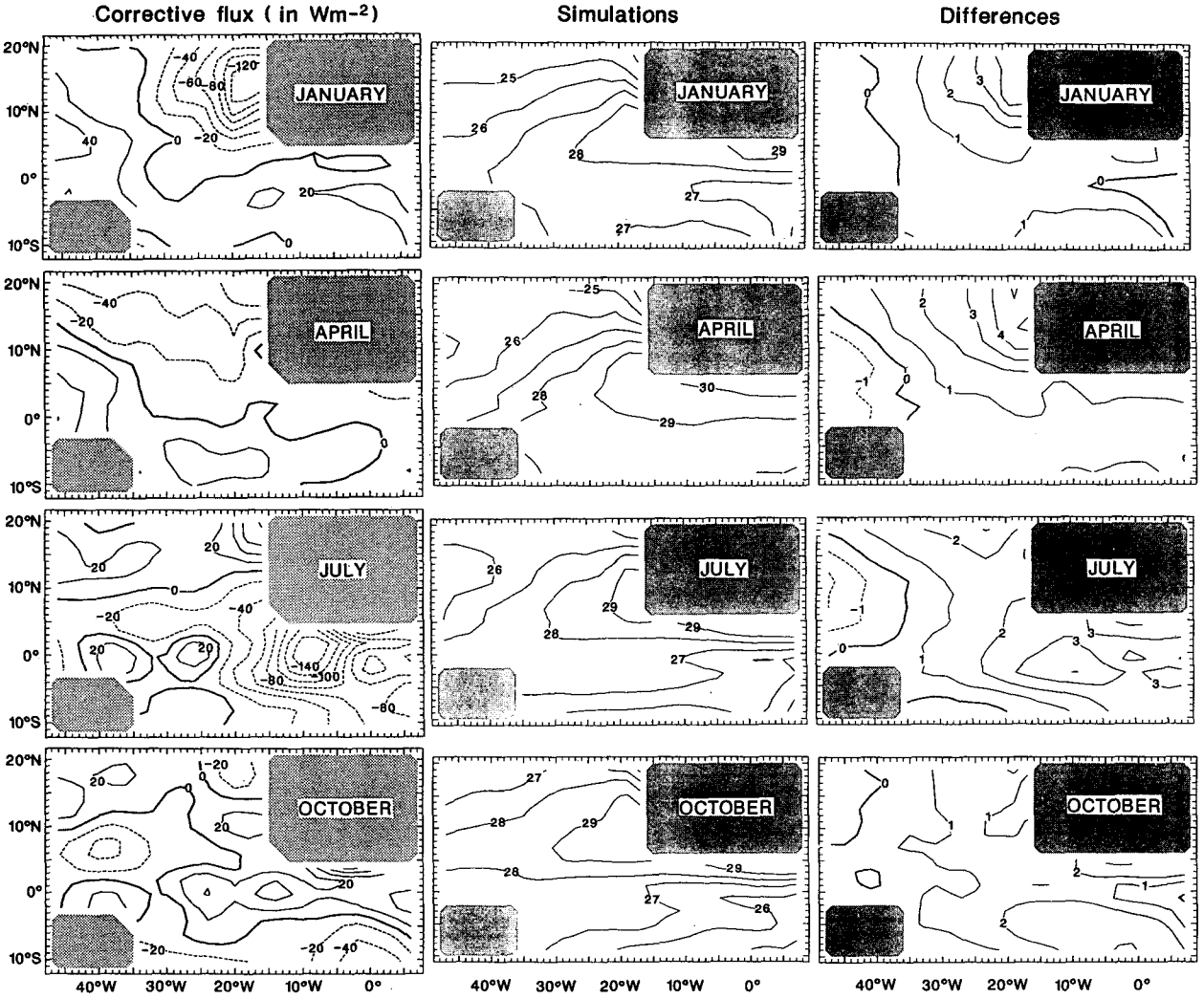

FIG. 5. (a) Mean heat flux correction in W m^{-2} during January, April, July, and October, after optimization. (b) Mean SST predicted by the multimode model using the optimal parameter values, for the period 1965–1986. (c) Differences between SST simulations and observations.

timates are strongly constrained by the a priori choice, which is puzzling. Nevertheless, the consistency of the results is altogether satisfactory, in view of the nonlinearity of the inverse problem.

b. *Adaptive tuning in the equatorial zone*

Since the poor model performances off Africa are in part linked to the inadequacy of the latent heating formulation near the coasts and the use of equatorial data to estimate T_d, even in off-equatorial areas (section 2a), a tuning only based on the equatorial zone (4°S–4°N) was also performed. The results are similar, but the horizontal diffusion only decreases to $(0.8 \pm 0.4)10^8$ m^2 s^{-1} and the upwelling efficiency to 0.36. However, the latter is then less resolved by the dataset (48%), so it remains in line in Fig. 6: the upwelling parameterization is not unambiguously better adapted to the equatorial zone. The test (24) still largely exceeds the critical values, although being smaller (158) than for the larger domain. When the model was rerun with the optimal parameters, the performances were similar to the case above, even when considered over the 20°N–10°S region [see Sennéchael-Scoffier (1994) for details].

c. *Comparison with BC's solution*

For comparison, BC's best-fit procedure was applied to the average of the five simulations with different bulk formulas, using our improved parameterization of SST cooling by upwelling (section 2a). The tuning is done in full space in the 20°N–10°S region, using BC's error models. The results differ from BC's, how-

TABLE 4. Rms value (in W m^{-2}) of the corrective heat flux for the yearly mean and the mean seasonal variability. The annual mean difference is given in italic.

Rms of the corrective flux	Before tuning	After tuning
Long-term annual mean	63	17
(mean bias)	*(−61)*	*(−8)*
Seasonal variability	31	26

ever, because of differences in the data sources: BC used the climatological wind stress of Hellerman and Rosenstein (1983) and SST of Reynolds (1982). Also, BC considered that the drag coefficient for the wind stress was an additional adjustable parameter. For the grid size of the atmospheric forcing and the SST observations used in the present paper, the "effective" number of degrees of freedom in the inversion is $\nu = 53$ (see BC, appendix B).

The tuning results are different from those of the adaptive tuning procedure, with less resolution by the dataset (only 1.9 parameters in total) and thus smaller parameter changes. Only one parameter changes significantly at the 5% level, a_{rh}, which increases to 0.39 ± 0.07; this parameter is mostly resolved by the dataset (60%) and is correlated with all of the other parameters but $\delta\gamma$, $\delta\kappa$, and δa_α. The upwelling efficiency is uncorrelated and only decreases to 0.48 ± 0.18, but is little resolved by the dataset (13%), so it remains in line in Fig. 6. The horizontal diffusion is little resolved (26%) and decreases to $(1.4 \pm 0.7)\ 10^8\ m^2\ s^{-1}$. To verify that these differences are due to differences in the observational error models, the fit was also performed in reduced space, with nearly identical results.

Since lesser weight is given to the dataset, it seems that BC overestimated the observational errors, thereby increasing the influence of the a priori information. BC assumed an rms observational error of 34 W m^{-2}, which does not seem unrealistic for the monthly timescale. Indeed, we find that the rms sample uncertainty in monthly variations due to interannual variability is 37 W m^{-2}. However, as the interannual variability affects nearly randomly each realization of the seasonal cycle, the corresponding errors over the 22-year mean are strongly reduced by averaging. A similar argument can be made for the drag coefficient uncertainties, and in both cases our sample value for the *mean* seasonal cycle reduces to only 8 W m^{-2}. Blumenthal and Cane's choice thus leads to an overestimation, in particular since its form was shown to be equivalent to prescribing uncorrelated observational errors with rms as large as 264 W m^{-2}.

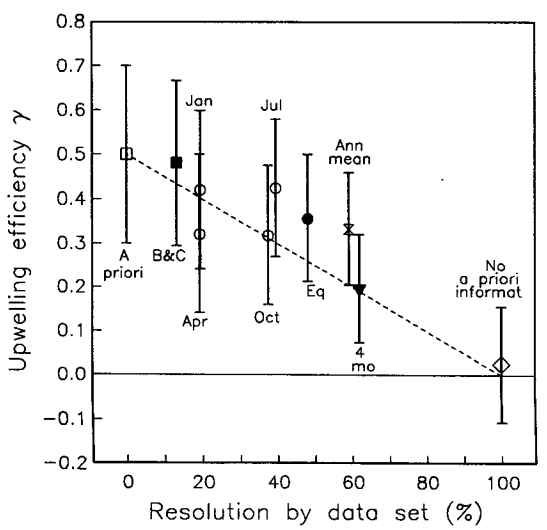

FIG. 6. Upwelling efficiency vs resolution by the dataset for various fits in the 20°N–10°S region.

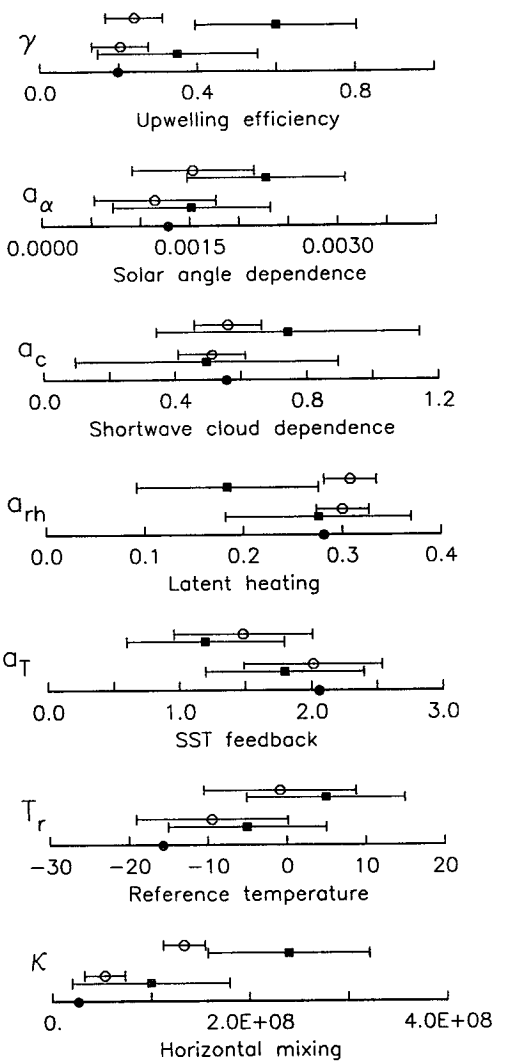

FIG. 7. Tuned parameters (open circle) for the initial choice given by the solid square below, in the case of the artificial SST data. The true value of the parameters is given by the solid circle, and the error bars represent the 95% confidence intervals.

The sensitivity of the tuning results to the observational error model and to the particular dataset being used supports our contention that it is crucial to accurately take into account the observational uncertainties. However, running the SST model with the parameters estimated with BC's method leads to performances that are not significantly different from the case of adaptive tuning (Table 1 and 2). Thus, the better representation of the errors provided by the adaptive procedure does not produce better SST simulations. What the more complex procedure does do is better reveal the magnitude and the nature of the model errors, thus pointing the way to improvement. Thus, it is to be expected that the superiority for tuning of the more refined adaptive method will become more apparent with a more realistic model.

5. A discussion of the SST model

An assessment of the dynamical model ability at simulating surface currents in the tropical Atlantic has been given in FDC and in Duchêne and Frankignoul (1990), who used as observational basis the mean surface currents derived from ship drifts. Perhaps in part because the model predicts only averaged currents for the mixed layer, it was not very successful at simulating the yearly mean conditions: the surface current was strongly underestimated along the equator, and the westward jet observed a few degrees north of the equator was not reproduced. The mean seasonal variability was better simulated, although many discrepancies remained. Duchêne and Frankignoul (1991) tested the model ability at reproducing the seasonal cycle of the surface dynamic topography, estimated by the 0/400 db dynamic height. The model performed better than for the surface currents, and also simulated the seasonal variations better than the annual mean. However, the discrepancies with the observations remained too large to be explainable by the data uncertainties, and more sophisticated models performed significantly better (Février 1993).

The evolution of the 20°C isotherm depth during the 1982–84 FOCAL/SEQUAL experiment was considered in Braconnot and Frankignoul (1993). There were differences for the 3-year mean; in particular, the equatorial slope was less pronounced in the model, and the meridional gradients near 6°S strongly underestimated. The linear model simulated successfully many features of the thermocline depth variability. Discrepancies with the observations were mainly associated with an unrealistic representation of the vertical displacement associated with equatorial upwelling, and the model failed to properly represent the eastward progression of the thermocline shallowing that is observed in the Gulf of Guinea during summer. Also, the model underestimated the interannual variability of the thermocline depth.

In these papers, the optimal number of vertical modes was searched. For surface currents, the linear model worked best with one vertical mode, although the differences with versions using three or more modes were not statistically significant. Surface dynamic topography was better simulated with a multimode version, and the FOCAL/SEQUAL data showed that the three-mode version was significantly better than the two-mode one. The five-mode version used in BC and the present paper is equally good, although more time consuming.

The thermodynamic version of the model was discussed by BC, who showed that it did not perform as well in the tropical Atlantic as in the tropical Pacific, in part because the Atlantic basin is smaller and more influenced by the boundaries, which are poorly represented. Blumenthal and Cane also stressed other reasons why the model could not realistically represent the SST off the North African coast: the parameterization of the temperature of the entrained water is adapted to the equator but not to off-equatorial upwelling, and the latent heat flux parameterization should only be applicable far from the coasts.

The present analysis confirms that the model cannot satisfactorily simulate SST off Africa, partly because the maximum cooling by coastal upwelling occurs too far south and offshore (see Fig. 3b). It also shows that the model does not properly simulate summer SST cooling by equatorial upwelling, even though it takes place away from the coast and the parameterization of entrained water should be well-adapted: neither the timing nor the geographical location of the equatorial upwelling flux is correct, so that the upwelling efficiency needs to be reduced. As a consequence, the model ability at representing SST anomalies in the Gulf of Guinea is decreased (Fig. 2, bottom). These flaws correspond to the inability of the model to correctly represent the evolution of the thermocline in the Gulf of Guinea (Braconnot and Frankignoul 1993) and seem linked both to an underestimation of westward advection by the equatorial currents (FDC) and a poor representation of vertical mixing. Experiments with general circulation models have stressed the need to correctly distribute the vertical momentum over depth to satisfactorily simulate the thermocline variability in this region (Blanke and Delecluse 1993). Our analysis also points to weaknesses in BC's parameterization of the temperature of the entrained water, which should probably be defined locally rather than for the whole basin.

The need to reduce the upwelling efficiency may seem at odds with the results of Miller et al. (1993), who found that, in the tropical Pacific, the SST anomalies in the model of Zebiak and Cane (1987) were best (and well) represented near the eastern equatorial boundary, where upwelling and vertical mixing primarily control the SST variability. However, Zebiak and Cane (1987) used a simpler parameterization of the temperature of the entrained water and only one vertical mode, so the increased complexity of BC's ver-

sion of the model may well have led to a decrease in model skill.

This discussion suggests that the model flaws are in part associated with the oversimplifications of the model dynamics and geometry, which introduce systematic biases that are modulated seasonally rather than random errors as assumed here for the modeling errors, and in part with the simplified representation of the surface heat exchanges. As the model has not demonstrated any skill in predicting SST anomalies in the tropical Atlantic, its usefulness is debatable for this ocean basin, except as a handy tool for methodological development. However, as pointed out by BC, the tropical SST model works much better in the larger Pacific Ocean, presumably because of its larger size.

6. Conclusions

We have developed an adaptive inverse method to tune the adjustable parameters of a tropical SST model in a way that optimally takes into account the large uncertainties of the atmospheric forcing and the oceanic data, the expected modeling errors, and our a priori knowledge of the parameter values. This is achieved by performing the model optimization for the mean seasonal SST cycle and using the dispersion of the model responses for each year and (equally plausible) forcing field as independent information to construct a sample estimate of the observational error covariance matrix. The procedure is more refined than that of BC in that the nonlinear nature of the inverse problem is partly taken into account and the large correlation scales of the forcing uncertainties are represented realistically. The method is general as long as the parameters enter the SST equation linearly, and it could be extended to the nonlinear case by using an iterative approach. Since the optimization is performed in a strongly reduced space, the computational cost is limited. However, a proper estimation of the observational errors requires that several multiyear model runs be available.

The method has been applied to tuning the SST model of BC in the tropical Atlantic. The optimization reduces the warm SST bias of the model, but brings no significant improvement in its ability at representing the seasonal or interannual SST fluctuations. A statistical test of the correctness of the assumptions in the inverse calculation shows that the modeling errors are much larger than assumed. In particular, the tuning procedure clearly reveals that there are model biases linked to its rudimentary geometry, to the simplifications in the latent heat formulation, and to its inability at properly representing SST cooling by upwelling. As the tuning is unable to improve the representation of SST changes, it is concluded that a more sophisticated model is needed to simulate SST variability in the tropical Atlantic. A different conclusion is likely to hold for the larger tropical Pacific, where the SST model is more realistic (BC) and has demonstrated significant skill in SST predictions (Cane et al. 1986).

The present study stresses that modeling errors are substantial in this SST model, and this is likely to remain true for more sophisticated ones. A proper representation of the modeling errors is thus needed to achieve an optimal model tuning, as a best fit should properly weight observations and simulations. The need to properly weight observational and modeling errors also holds for data assimilation methods, and not too much faith should be given to model reconstituted fields based on methods that do not properly represent the modeling errors.

Finally, the adaptive tuning procedure provides an alternative to imposing the "flux correction" that is often needed to avoid climate drift when coupling an SST model to an atmospheric model. Indeed, the decrease in mean SST bias should decrease climate drift in the coupled mode without introducing the drawbacks of the flux correction method (see Latif et al. 1993), since the correction more properly takes place via model parameters, without altering the SST dynamics.

Acknowledgments. We would like to thank B. Blumenthal and P. Braconnot for their help. Useful discussions with C. Wunsch and F. Martel are also acknowledged. N.S. was supported by a grant from the DRET. This research was made possible in France by grants from the PNEDC and the Commission of the European Communities (EV4C006), and done in part at the Lamont–Doherty Earth Observatory and supported by ONR Contract N00014-J-15951 with Columbia University, and at the Massachusetts Institute of Technology, whose hospitality to one of us (C.F.) is gratefully acknowledged.

APPENDIX

Data Compression for the SST Model Testing

To test the simulations of the long-term mean SST, denoted by the superscript y, we work in the (normalized) space of the yearly mean temperature *difference* $(\langle \mathbf{T}^y \rangle - \mathbf{T}_o^y)$. Mean differences between yearly mean observations and model response to the five forcing fields are projected in this space for each of the 22 years. Their sample average (over the 22 years and the five forcings) is tested for zero mean (null hypothesis). As the test statistic T^2 then reduces the square of a t variable, the latter is used in Table 2, using a one-sided distribution to calculate the critical value.

To test the simulations of the mean seasonal variability, denoted by the superscript s, a time compression is first performed on the mean monthly differences $(\langle \mathbf{T}^s \rangle - \mathbf{T}_o^s)$ by considering four seasons (weighted averages, 1/4 1/2 1/4, centered on January, April, July, and October) instead of 12 months. A data compression is then performed in the spatial domain,

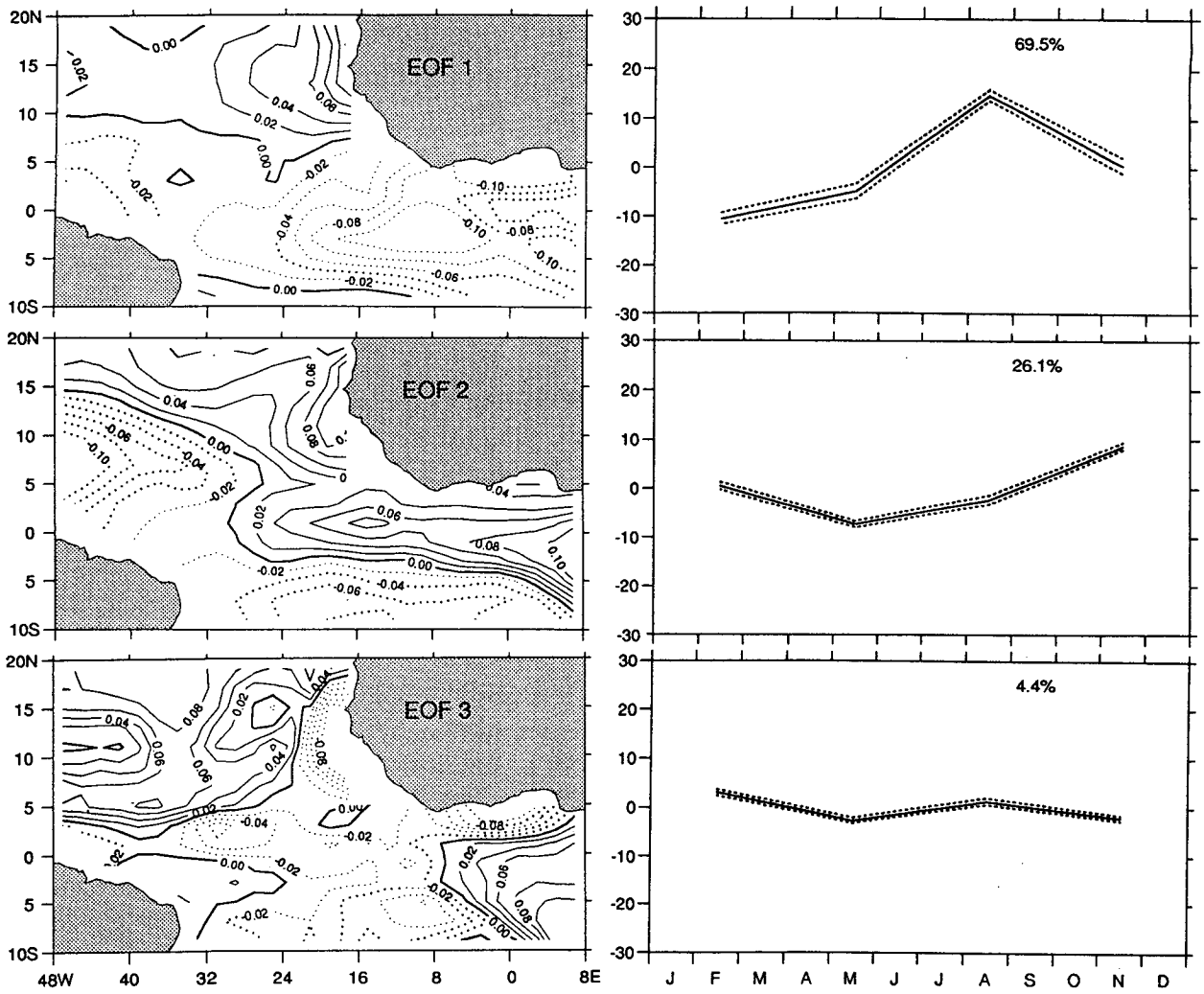

FIG. A1. (Left) EOFs of the SST differences used in the multivariate test for the mean seasonal variability (tuned model). (Right) Corresponding principal components (continuous line) with 95% error bars (dotted line).

using principal component analysis, which provides three basis vectors (EOFs) that represent 100% of the variance (the annual mean is not included). This efficient method leads to more precise model intercomparisons than in the previous uses of the model testing method, since there is no truncation during the spatial reduction, and recent results have shown that the method is sensitive to the level of truncation (Sennéchael-Scoffier 1994). Figure A1 illustrates the data compression for the tuned model by representing the three spatial EOFs of the seasonal model–observation differences (left), and the corresponding principal components together with univariate 95% confidence intervals derived from the square root of the diagonal elements of the error covariance matrix in reduced space, using the t distribution (right). The model–observation differences have large space–time correlation scales, with a primarily annual component, and are much larger than expected from the data uncertainties considered. The misfit T^2 is computed in a 12-dimensional space.

REFERENCES

Blanke, B., and P. Delecluse, 1993: Variability of the tropical Atlantic Ocean simulated by a general circulation model with mixed layer physics. *J. Phys. Oceanogr.*, **23**, 1363–1388.

Blumenthal, M. B., and M. A. Cane, 1989: Accounting for parameter uncertainties in model verification: An illustration with tropical sea surface temperature. *J. Phys. Oceanogr.*, **19**, 815–830.

Braconnot, P., and C. Frankignoul, 1993: Testing model simulations of the thermocline depth variability in the tropical Atlantic from 1982 through 1984. *J. Phys. Oceanogr.*, **23**, 626–647.

Cane, M. A., 1984: Modeling sea level during El Niño. *J. Phys. Oceanogr.*, **14**, 1864–1874.

——, S. E. Zebiak, and S. C. Dolan, 1986: Experimental forecasts of El Niño. *Nature*, **321**, 827–832.

Cardone, V. J., J. F. Greenwood, and M. A. Cane, 1990: On the trends in marine surface wind observations. *J. Climate*, **3**, 113–127.

Duchêne, C., 1989: Test statistique de la validité des modèles océaniques équatoriaux à l'aide d'observations. Thèse de doctorat. Université Pierre et Marie Curie, Paris, 162 pp.

——, and C. Frankignoul, 1990: Sensitivity and realism of wind-driven tropical ocean models. *J. Mar. Syst.*, **1**, 97–117.

——, and ——, 1991: Seasonal variations of surface dynamic topography in the tropical Atlantic: Observational uncertainties and model testing. *J. Mar. Res.*, **49**, 223–247.

Esbensen, S. K., and Y. Kushnir, 1981: The heat budget of the global ocean: An atlas based on estimates from marine surface observations. Climatic Research Institution, Report 29, Oregon State University, Corvallis, 27 pp.

Février, S., 1993: Modèle à deux couches de l'océan atlantique tropical. Thèse de doctorat. Université Pierre et Marie Curie, Paris, 178 pp.

Frankignoul, C., C. Duchêne, and M. Cane, 1989: A statistical approach to testing equatorial ocean models with observed data. *J. Phys. Oceanogr.*, **19**, 1191–1208.

——, N. Scoffier, and M. Cane, 1993: An adaptive inverse method for model tuning and testing. *Probability Concepts in Physical Oceanography, Proc. "Aha Huliko'a Hawaiian Winter Workshop,"* P. Müller, D. Henderson, Eds., University of Hawaii, 331–349.

Harrison, D. E., B. S. Giese, and E. S. Sarachik, 1990: Mechanisms of SST change in the equatorial waveguide during the 1982–83 ENSO. *J. Climate*, **3**, 173–188.

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

Isemer, H.-J., and L. Hasse, 1987: *The Bunker Climate Atlas of the North Atlantic Ocean*, Vol. 2: *Air–Sea Interactions*. Springer-Verlag, 252 pp.

Judge, G. G., R. C. Hill, W. E. Griffiths, H. Lutkepohl, and T-C. Lee, 1988: *Introduction to the Theory and Practice of Econometrics*. 2d ed. John Wiley, 1024 pp.

Large, W. G., and S. Pond, 1981: Open ocean momentum flux measurements in moderate to strong winds. *J. Phys. Oceanogr.*, **11**, 324–336.

Latif, M., A. Sterl, E. Maier-Raimer, and M. M. Junge, 1993: Climate variability in a coupled GCM. Part 1: The tropical Pacific. *J. Climate*, **6**, 5–21.

Levitus, S., 1982: *Climatological Atlas of the World Ocean*. NOAA Prof. Paper No. 13, U.S. Govt. Printing Office, Washington, DC, 173 pp.

Liu, W. T., and P. P. Niiler, 1990: The sensitivity of latent heat flux to the air humidity approximation used in ocean circulation models. *J. Geophys. Res.*, **95**, 9745–9753.

——, K. B. Katsaros, and J. A. Businger, 1979: Bulk parameterization of air–sea exchanges of heat and water vapor including the molecular constraints at the interface. *Atmos. Sci.*, **36**, 1722–1735.

Mardia, K. V., J. T. Kent, and J. M. Biddy, 1979: *Multivariate Analysis*. Academic Press, 518 pp.

Miller, A. J., T. P. Barnett, and N. E. Graham, 1993: A comparison of some tropical ocean models: Hindcast skill and El Niño evolution. *J. Phys. Oceanogr.*, **23**, 1567–1591.

Morrison, D. F., 1976: *Multivariate Statistical Methods*. McGraw-Hill, 415 pp.

Reynolds, R. W., 1982: A monthly averaged climatology of sea surface temperature. NOAA Tech. Rep. NWS 31, National Weather Service, Silver Spring, MD, 35 pp.

——, 1988: A real-time global sea surface temperature analysis. *J. Geophys. Res.*, **90**, 11 587–11 601.

Seager, R., S. E. Zebiak, and M. A. Cane, 1988: A model of the tropical Pacific sea surface temperature climatology. *J. Geophys. Res.*, **93**, 1265–1280.

Sennéchael-Scoffier, N., 1994: Optimisation d'un modèle de l'océan atlantique tropical par méthode inverse adaptative. Thèse de doctorat. Université Pierre et Marie Curie, Paris, 171 pp.

Servain, J., J. Picaut, and A. J. Busalacchi, 1985: Interannual and seasonal variability of the tropical Atlantic Ocean depicted by 16 years of sea surface temperature and wind stress. *Coupled Ocean-Atmosphere Models*, J. C. J. Nihoul, Ed., Elsevier, 211–237.

Smith, S. D., 1988: Coefficient for sea surface wind stress, heat fluxes, and wind profiles as a function of wind speed and temperature. *J. Geophys. Res.*, **93**, 15 467–15 472.

Tarantola, A., 1987: *Inverse Problem Theory*. Elsevier, 613 pp.

Zebiak, S. E., and M. A. Cane, 1987: A model El Niño–Southern Oscillation. *Mon. Wea. Rev.*, **115**, 2262–2278.