A tropical Pacific prediction system of intermediate complexity: role of the vertical structure variability.

Dewitte B.[1], Ganachaud A.[2], Dejoan M.-A.[2], Grimini P.[2] and Y. duPenhoat[1]

1 LEGOS/IRD/CNES, 14 av. Edouard Belin, 31401 Toulouse Cedex, France

2 IRD/LEGOS, Centre de Nouméa, 101 Promenade Roger Laroque - Anse Vata, BP A5 - 98848 NOUMEA Cedex, New Caledonia

Abstract

An intermediate ocean-atmosphere coupled model of the tropical Pacific is used to investigate the sensitivity of the seasonal forecasts to the configuration of the oceanic vertical structure. The models consist in a three baroclinic mode tropical Pacific Ocean model and a Gill (1980)'s tropical atmosphere. The predictive skill of the model using a simple nudging method for the initialization is first presented from 1970 and compared to the results of other prediction system of similar complexity, which emphasizes the modulation of the skill on decadal timescales. It is then demonstrated that the skill is critically dependant on the energy distribution on the baroclinic mode, higher-order mode contributions being favored at some period and not at others. Linear Green's function are used to assimilate satellite observations (SST and wind) and derive the optimized set of parameters that determines the vertical structure in the model for a particular period of time. The scheme is first tested for the period prior to the development of the 1997 El Niño. It is shown that substantial improvement in forecasting the event is realized for an increase of the relative contribution of the higher-order modes through the model parameters setting. Assimilation experiments of satellite data for the September 2003-February 2004 period are carried for producing initial conditions for the coupled model. Results of forecasts runs for 2004 are presented and discussed.

1. Introduction

Despite the fact that the tropical Pacific has never been so well observed, the last 1997/1998 El Niño was surprisingly the most unexpected event of the last 15 years. In particular, the Lamont-Doherty Earth Observatory model (hereinafter referred to as the Lamont model) which used to be skillful in the 80s and early 90s (Cane et al., 1986; see the Climate Diagnostic Bulletins), failed to predict the onset of this strong warm event. Similarly, the 2002 El Niño, although of much weaker amplitude than the 1997/98 event was not well detected with this simple prediction system. Several explanations were offered. First, the Lamont model is very sensitive to the initialization: Chen et al. (1999) noted a large sensitivity of the predictions to the wind products for the 1997/1998 El Niño. Initializing the prediction runs with sea level anomalies derived from observations also results in a large range of behaviors of this model (Périgaud et al. 2000). Second, as an anomaly model with prescribed mean states, it can not grasp climate changes which are likely to impact El Niño occurrence. As the general circulation models (GCMs) with extensive assimilation schemes have done a better job in recent year, this calls for revising the simple prediction systems by including more sophisticated assimilation schemes (Chen et al. 1995, hereafter CZBC95) and/or including more physics. Dewitte (2000) (hereafter D0) has included more vertical modes into the oceanic component of a model similar to Lamont in order to better represent the thermocline depth fluctuations and the dissipation of the waves associated to vertical propagation of energy. This was motivated by recent observational and modeling studies which indicate that more than one baroclinic mode is necessary to capture correctly the sea level and zonal current variability (McPhaden, 1999). Sensitivity experiments with this model to the energy distribution on the baroclinic mode have shown that the predictions were sensitive to the subsurface condition prior to the development of the ENSO event (Dewitte et al., 2002). In particular, the model better grasp the rise in temperature and amplitude at the mature phase of the 1997/98 El Niño. Their results however still leads much space for improvement of the predictions, which can be obtained from data assimilation. In the absence of realistic estimate of the baroclinic contribution to, say, sea level or current anomalies from observations, there is a need to develop assimilation techniques that take into account the specificity of the baroclinic modes. In this paper, we present an assimilation method that allows projecting surface observations on the model baroclinic modes contribution. We show that it provides initial conditions for the model that lead to improvement in forecasting compared to the nudging method. The results of Dewitte et al (2002) are confirmed, namely that the contribution of higher-order baroclinic mode are important for ENSO predictions and that the second baroclinic mode contribution was increased prior to the development for the 1997/98 El

Niño, which has participated to a large extent to its intensity. The study also provides an estimate of the variability that projects on the first 3 baroclinic modes prior to the development of the 1997/1998 El Niño event and analyse prediction for 2004.

The paper is organised as follows. Section 2 is devoted to the data, model and assimilation method description. Section 3 presents the results and section 4 is a discussion.

2. Model and data description

Model:

We use a tropical Pacific ocean-atmosphere model of intermediate complexity. It is an extension of the Zebiak and Cane (1987)'s model (hereafter ZC model) in that it is based on similar physics, i.e. shallow-water for both components. The ocean component includes 3 baroclinic modes with characteristic phase speed c_n , projection coefficient P_n and 'thermocline coefficient' scl_n (used to derive the thermocline fluctuations in a multi-mode context and which depends directly on N^2 and the vertical derivative of the baroclinic mode vertical structures – see D0 for details) derived from the Levitus data set. A mixed layer model is embedded in the ocean model that consists in a thermodynamical budget in a 50m-thick surface layer. The surface heat flux is parameterized as being negatively proportional to local SST anomalies. Subsurface entrainment temperature into the surface mixed layer is parameterized as a function of thermocline depth anomalies and mean thermocline depth (cf. Dewitte and Perigaud, 1996). The atmospheric component is similar to Gill (1980). The reader is invited to refer to D0 and Dewitte et al. (2002) for more details on the model and results on sensitivity tests to parameters and ENSO predictions with this model.

Data: We assimilate two types of data into the Intermediate Coupled Model (ICM): SST and zonal wind stress anomalies.

SST: The SST data used for the assimilation over the full reduced grid are compiled by the NOAA Climate Analysis Center, (http://ingrid.ldgo.columbia.edu/SOURCES). Originally, they were computed on a 1^o by 1^o regular grid, monthly averages. Due to changes in the seasonal cycle, interannual SST anomalies are first obtained by removing the climatological seasonal cycle averaged between January 1982 to December 2001 for the prediction runs of the 1996/97 El Niño event, and then between January 1992 to December 2003 for predictions of 2004. Then, the anomalies are interpolated on the model grid (2^ox5.625^o).

WIND: The wind stress anomalies used to force the ocean model are derived from the pseudo-stress data provided by Florida State University.They are relative to the seasonal cycle computed over the period January 1980-December 1995. We also used ERS-1 and ERS-2 wind data provided by the Centre ERS d'Archivage et de Traitement (CERSAT, Brest). The method for converting the radar backscatter measured by the ERS-1 and ERS-2 scatterometers is detailed in Bentamy at al. (1996). The original product is gridded on a 1^o longitude by 1^o latitude regular grid and is a weekly average wind stress field. Data are monthly averaged and interpolated on a 2^ox2^o regular grid. The anomalies relative to the seasonal cycle over October 1992-September 1996 were derived for the period May 1992- December 1998. Menkes et al. (1998) have highlighted the weak amplitudes of the ERS wind stress as compared to TAO wind data. A multiplicative factor of 1.4 was applied to the data in order to have the same level of variability than the FSU wind anomalies over the period January 93-December 96. From October 1999, Quickscat data are used. ERS data and QuickSCAT data are intercompared over the overlapping period (i.e. October 1999-February 2002) in order to scale the level of variability in the equatorial band (a uniform 1.25 coefficient is used).

Like for SST, the anomalies are relative to the seasonal cycle computed over the periods, January 1982-December 2001 for the predictions of 1997/98, whereas they are obtained by removing the climatological seasonal cycle averaged between January 1992 to December 2003 for predicting the 2004 conditions.

The data were monthly averaged and interpolated on a 2^ox2^o regular grid. Due to the tendency of the Gill model to overestimate the magnitude of the anomalies outside the equatorial band, wind stress is assimilated only in the 7°S-7°N band. Sensitivity tests also suggests that the model physics is not appropriate to deal with the assimilation of observed meridional wind stress anomalies, which exerts too large of a constraint within the model assumptions. They were therefore not assimilated in the model.

3. Assimilation method

The method of linear Green's Functions (Menemelis and Wunsh, 1996) is used to improve ICM initial conditions and parameters with data assimilation.

A linearity hypothesis is done by assuming that large-scale differences between the model's predictions and the data on a reduced grid are sufficiently small for their evolution to be linear. After a perturbation experiment to test the model's stability and prepare the linearization, the Green's Functions are calculated to build a linear model which describes the evolution of these differences.

Then, we use the Gauss-Markov's estimator to minimise a cost function and obtained the best fit for the desired parameters. The formalism is detailed in the Appendix.

As mentioned above, only SST and zonal wind stress anomalies were assimilated. Two critical parameters of ICM were chosen to be optimised: The projection coefficient, P_n and 'thermocline coefficient', scl_n parameters. As shown in D0, these two parameters determine the relative distribution of energy on the baroclinic modes in the model and can have a large impact on the predictions. Dewitte et al. (2002) only tested the sensitivity of the 1997/98 El Niño prediction to prescribed P_n. Using the assimilation method proposed here, it is interesting to evaluate if assimilation of data in ICM leads to changes in P_n similar to what was guessed empirically in Dewitte et al. (2002). We also complement the latter study by providing estimate of change in scl_n.

4. Results

Several assimilation windows, never exceeding 6 months (due to computational constrain), were chosen in order to assess from which date and duration, the assimilation of data is useful in improving the forecasts compared to the classical nudging procedure (CZBC95). Each times, changes in (P_n, scl_n), i.e. (δP_n, δscl_n), resulting from the data assimilation over the 2 to 6 months period is estimated from the minimization procedure of the cost function. They are given in percentage of the mean value. For each experiment, perturbations at the initialisation of the prediction runs are applied to produce an ensemble of 100 runs.

The 1997/98 El Niño:

The NINO3 SST index (SST anomalies average over (150°W-90°W; 5°S-5°N)) are displayed in figure 1 for one experiment (assimilation window: Nov-Apr 1996/1997).

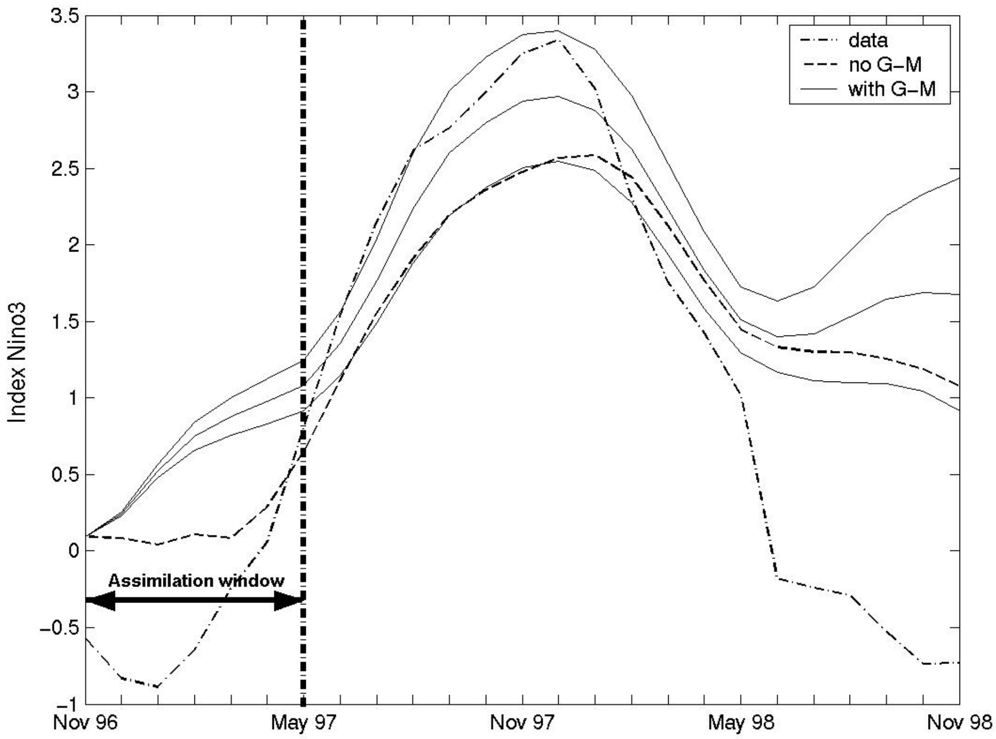

Figure 1: NINO3SST index for the observations (dash-doted line), for the model using the standard initialisation method (thick dashed line) (initial conditions in May 1997) and for the model with the data assimilation procedure starting in November 1996 until April 1997 (middle solid thin line). Prediction with assimilation is the average of one hundred perturbed predictions runs, standard deviation curves are also represented (thin solid lines).

The assimilation of data results in an improved prediction of the 1997/98 El Niño, with a more realistic amplitude of SSTA at the mature phase of the event as long as with a better timing of the peak. The growth phase of the event is also better simulated with less delay between simulation and observation at the early stage of the development of the event. Similar improvements of the prediction are obtained for the assimilation windows of 6 and 4 months spanning December to May 1996/1997, January to June 1997 and March to June 1997. For the experiment with a 2 months windows, May to June 1997, the system is as skilful as the standard predictions with the peak of the event underestimated by ~30% as compared to the observations (not shown).

It is now interesting to analyse the impact of the assimilation on the (P_n, scl_n) parameters. Results for the various experiment are summarized in Table 1.

Assimilation period	δP_1	δP_2	δP_3	δscl_1	δscl_2	δscl_3
Exp1: Nov-Apr 1996/97 (6-month window)	13	29	5	21	13	2
Exp2: Dec-May1996/97 (6-month window)	14	27	3	28	3	-0.2
Exp3: Jan-Jun 1997 (6-month window)	16	25	4	24	7	-0.4
Exp4: Mar-Jun 1997 (4-month window)	2	9	3	4	5	3
Exp5: May-Jun1997 (2-month window)	5	5	0.7	5	6	1

Table 1: Percentage of variation of the (P_n, scl_n) parameters from their mean value for the different assimilation experiments associated to the prediction of the 1997/98 El Niño.

Consistently with Dewitte et al. (2002), the experiments that lead to improve simulation of the peak phase of the 1997/98 El Niño event are associated with an increased relative contribution of the second baroclinic mode. For instance, for Exp1 (figure 1), the increase in P_2 is twice as large as in P_1. Interestingly, there is an opposite behavior for the scl_n coefficient, with an increase in scl_1 much larger than the one in scl_2 for the experiments leading to the best fits between observations and predictions (i.e. Exp 1, 2 and 3). This suggests that for the 1997/98 El Niño to develop, vertical advection had to be more controlled by the first baroclinic mode whereas zonal advection was more controlled by the higher-order baroclinic modes. This is consistent with OGCM results of Dewitte et al. (2003) that show that the baroclinic mode contribution to zonal current anomalies was larger for mode 2 than for mode 1 during the growing phase of the 1997/98 El Niño. Also, the 1997/98 El Niño developed very rapidly and reverse to La Niña abruptly, which is also consistent with vertical advection controlled by the fast first baroclinic mode (which determined the warm phase) and strong eastward zonal current anomalies resulting from the reflection of second baroclinic mode Kelvin wave at the eastern boundary (see Dewitte et al. (2003)).

<u>*2004:*</u>

Similar experiments were performed for predicting SSTAs for the second half of 2004. The results indicate a much larger sensitivity of the results to the period over which the assimilation is performed. In all the runs, the model predicts slightly warm conditions (not shown). However, the impact on the (P_n, scl_n) is more difficult to interpret with a large deviation of (P_n,scl_n) depending on the period over which the data assimilation is performed (see Table 2). For instance, P_1 is increased by 32% for the assimilation starting in October 2003, whereas it is decreased by 10% for the assimilation starting one month later. P_2 is always reduced whereas scl_2 is increased when the assimilation starts in December 2003. scl_1 is slightly increased in all the experiments. The trend in the parameter changes from these few experiments suggests however that, in case of El Niño conditions actually developing in 2004, thermocline feedback would be favoured, leading to SST warm anomalies slowly propagating eastward.

Assimilation period	δP_1	δP_2	δP_3	δscl_1	δscl_2	δscl_3
Exp6: Oct-Mar 2003/04 (6-month window)	+32	-16	-11	+5	-17	+2
Exp7: Nov-Apr 2003/04 (6-month window)	-10	-40	-10	+7	-22	-8
Exp8: Dec-May 2003/04 (6-month window)	-23	-19	-5	+5	+14	+2

Table 2: Percentage of variation of the (P_n, scl_n) parameters from their mean value for the different assimilation experiments associated to the prediction of the 2004 conditions.

Discussion and conclusions

The sensitivity to the relative contribution of the baroclinic modes in a intermediate prediction system is confirmed using an assimilation method. It uses SST and zonal wind stress satellite derived data to optimise the initial conditions corresponding to SSTAs and the baroclinic mode parameters (P_n, scl_n). It is shown that the method is helpful in simulating realistic amplitude of the peak phase of the 1997/98 El Niño. In addition, the variations of the optimised parameters (P_n, scl_n) is consistent with earlier sensitivity experiments (Dewitte et al., 2002), showing that the 1997/98 El Niño is better predicted when the second baroclinic mode contribution is increased compared to its mean value prior to the development of the event. The method further brings some insight on the coupled unstable modes involved for this particular event through the variation of the scl_n, that controlled the thermocline feedback in the model. The results suggests that the zonal advective feedback is enhanced through the high-order baroclinic modes in the model whereas the thermocline feedback is more controlled by the gravest baroclinic mode. According to instability theories, this favors the presence of a fast unstable Rossby mode, consistently with the rapid development of the event and its rather abrupt reversal to La Niña (See D0).

Further sensitivity tests are carried out for year 2004. The system predicts slightly warm condition developing in the second half of 2004. However, the range of deviations of (P_n, scl_n) depending on the period over which the data assimilation is more difficult to interpret with regards to the coupled mechanisms (thermocline or zonal advective feedbacks) favoured during this period. Further sensitivity tests, that include, for instance, constraints on other parameters in the model (coupling efficiency coefficient, friction…) are needed to reach any conclusive statement. This is work under progress.

Appendix: Method of linear Green's Functions

A predictive model can be writing as follows: $\hat{\xi}(t+1) = \Psi(\hat{\xi}(t), w(t))$,

Where $\hat{\xi}(t)$ is the predicted state vector, $w(t)$ represents boundary conditions and model parameters at any time t. We will first define the reduced state vector: $x(t) = \mathrm{B}^*(\xi(t) - \hat{\xi}(t))$,

where $\mathrm{B}^* = \mathrm{B}_s^*\mathrm{B}_t^*$, represents the state reduction operator that projects the reduced state on a reduced grid, for the spatial and temporal component. The full state is found by using a pseudo-inverse operator B, such as $\mathrm{B}^*\mathrm{B} = \mathrm{I}$, where I is the identity matrix: $\xi(t) - \hat{\xi}(t) = \mathrm{B}x(t) + \varepsilon(t)$, $\varepsilon(t)$, represents the high-frequency specific that has been removed by B^*.

The linearity hypothesis is done by assuming that large-scale differences between the model's predictions and the ocean's real state on reduced grid are sufficiently small for their evolution to be linear. Thus, its evolution is governed by an equation whom form is: $x(t+1) = A(t)x(t) + q(t)$, $A(t)$ is the state transition matrix, $q(t)$ is the control variable that accounts for model errors.

A perturbation experiment is done to test the model stability and prepare the linearization. Model Green's function (Wunsch, 1996), is defined here as the linear model response in SST and zonal component of wind stress to either, unit temperature perturbations, unit P_n perturbations or unit scl_n perturbations of the initial reduced state vector $x(t_0)$.

To obtain a sensible linear model, the transition matrix has to be dependent on time and is calculated through the Green Function $G(t_1, t_2)$.

The perturbations are introduced on each element of the reduced grid and projected onto the original grid using B. The full predictive model run and the response in SST and zonal component of wind stress are projected back onto the reduced grid. The resulting vector gives the column of the Green Function corresponding to each perturbed element.

From any initial state $x(t_0)$, the linear Green Model gives an estimate of the state vector at any time: $x(t_0 + t) = G(t_0 + t, t_0)x(t_0)$. Assimilation is done on this reduced grid, minimizing x, the deviation between the ICM prediction and the data.

The data $\eta(t)$ is represented by the linear combination: $\eta(t) = \mathrm{E}\xi(t) + \nu(t)$, where $\xi(t)$ and $\nu(t)$ are respectively the state vector and the data errors.

Let $y(t)$ be the difference between data and model prediction, (both relative to a seasonal cycle and in the reduced space): $y(t) = \mathrm{B}^*\eta(t) - \mathrm{EB}^*\hat{\xi}(t) = \mathrm{E}x(t) + n(t),$ with $n(t) = \mathrm{B}^*\nu(t).$

The total inversion method is to consider whole the data and using Gauss-Markov's estimation to determinate the optimum initial state. In our experiments, the initial state corresponds to SST, projection and 'thermocline' coefficients (P_n and scl_n). At any time, measurement equations take on the form:

$$
\begin{aligned}
y(0) &= \mathrm{E}x(0) + n(0) \\
y(1) &= \mathrm{E}x(1) + n(1) \\
y(1) &= \mathrm{E}G(1,0)x(0) + n(1) \\
&\vdots
\end{aligned}
$$

That can be put on under matrix form (Stammer and Wunsch, 1996):

$$
\begin{bmatrix} y(0) \\ y(1) \\ \vdots \\ y(k) \end{bmatrix} = \begin{bmatrix} \mathrm{E}(0) & 0 & \cdots & 0 \\ 0 & \mathrm{E}(1) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathrm{E}(k) \end{bmatrix} \begin{bmatrix} \mathrm{I}_N & 0 & \cdots & 0 \\ G(1,0) & \mathrm{I}_N & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ G(k,0) & G(k-1,0) & \cdots & \mathrm{I}_N \end{bmatrix} \begin{bmatrix} x(0) \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \begin{bmatrix} n(0) \\ n(1) \\ \vdots \\ n(k) \end{bmatrix}.
$$

This equation can be rewritten: $y' = Gx' + n'$.

Gauss-Markov's estimation (Wunsch, 1996), through a statistical estimation theory, allows to find an estimate $\tilde{x}'$ and its corresponding $\tilde{n}'$ (as well as their uncertainties $\mathrm{P}_{x'}$ and $\mathrm{P}_{n'}$) which deviates as little as possible in the mean square from the true solution. Data, Y and second moments, $R_{x'x'} = \left\langle x' x'^T \right\rangle$ and $R_{n'n'} = \left\langle n' n'^T \right\rangle$, appear into the algorithm given the estimations :

$$
\begin{aligned}
\tilde{x}' &= R_{x'x'}G^T(GR_{x'x'}G^T + R_{n'n'})^{-1}\mathrm{Y} \\
\tilde{n}' &= \left\{\mathrm{I} - GR_{x'x'}G^T(GR_{x'x'}G^T + R_{n'n'})^{-1}\right\}\mathrm{Y}
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{P}_{x'} &= R_{x'x'} - R_{x'x'}G^T(GR_{x'x'}G^T + R_{n'n'})^{-1}GR_{x'x'} \\
\mathrm{P}_{n'} &= \left\{\mathrm{I} - GR_{x'x'}G^T(GR_{x'x'}G^T + R_{n'n'})^{-1}\right\}R_{n'n'}\left\{\mathrm{I} - GR_{x'x'}G^T(GR_{x'x'}G^T + R_{n'n'})^{-1}\right\}.
\end{aligned}
$$

These expressions form the estimator that permits to find the optimum initial state $x(0)$ to start back the predictive model in order to obtain prediction closer to reality. $\delta\mathrm{P}_n$ and δscl_n were included in the $x(0)$ vector.

References

Bentamy, A., Y. Quilfen, F. Gohin, N. Grima, M. Lenaour, and J. Servain, Determination and validation of average wind fields from ERS-1 scatterometer measurements. *The Global Atmosphere and Ocean System*, **4**, 1-29, 1996.

Chen, D., S. E. Zebiak, A. J. Busalacchi, and M. A. Cane, An improved procedure for El Niño forecasting: Implication for predictability. Science, 269, 1699-1702, 1995.

Dewitte B. and C. Périgaud, 1996 : El Niño-La Niña events simulated with the Cane and Zebiak's model and observed with satellite or in situ data. Part I: Model forced with observations. *J. Climate*, **9**, 1188-1207.

Dewitte B., 2000: Sensitivity of an intermediate coupled ocean-atmosphere model of the tropical Pacific to its oceanic vertical structure. *J. Climate*, **13**, 2363-2388.

Dewitte B., D. Gushchina, Y. duPenhoat and S. Lakeev, 2002: On the importance of subsurface variability for ENSO simulation and prediction with intermediate coupled models of the Tropical Pacific: A case study for the 1997-1998 El Niño. *Geoph. Res. Lett.*, 29 (14): Art. No. 1666.

Dewitte B., S. Illig, L. Parent, Y. duPenhoat, L. Gourdeau and J. Verron, 2003: Tropical Pacific baroclinic mode contribution and associated long waves for the 1994-1999 period from an assimilation experiment with altimetric data. *J. Geophys. Research.*, 108 (C4), 3121-3138.

Gill, A., 1980: Some simple solutions heat-induced tropical circulation. *Q. J. R. Meteorol. Soc.*, **106**, 447-462.

Menemenlis, D., and C. Wunsh, 1996: Linearization of an Oceanic General Circulation Model for Data Assimilation and Climate Studies. *J. Atmos*. Oceanic Technol., **14**, 1420-1443.

Menkes C. and coauthors (1998) Impact of TAO vs. ERS wind stresses onto simulations of the tropical Pacific Ocean during the 1993-1998 period by the OPA OGCM. Climate Impact of Scale Interactions for the Tropical Ocean-Atmosphere System, Euroclivar Workshop Report, Eucliv 13, pp. 46-48.

McPhaden M., Genesis and evolution of the 1997-98 El Niño, Science, *283*, 950-954, 1999.

Wunsh, C., 1996: The Ocean Circulation Inverse Problem. *Cambridge Univ. Press*, 442 pp.

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