

OCEANIC MODEL TESTING

Claude Frankignoul

Laboratoire d' Océanographie Dynamique et de Climatologie,
Université Pierre et Marie Curie, 4 Place Jussieu, 75005 Paris, France

1. Introduction

In general, it is difficult to objectively compare oceanic model results to observations, as observations are usually noisy and inaccurate, and the sample too limited, resulting in a small signal-to-noise ratio. The noise of the observed signal results from both measurement uncertainties and real disturbances, like mesoscale eddies when studying the general circulation, vortical modes when studying internal waves, or interannual fluctuations when considering the seasonal cycle; this noise mostly has large correlation scales, so that simple point-by-point comparisons with model predictions are difficult to interpret, and a multivariate viewpoint is needed. Moreover, and this is specific to the oceanic case, most motions are *externally forced* but precise information is lacking on the forcing and on the initial conditions. Thus, even a perfect oceanic model will not provide predictions that are fully consistent with the observations, i.e., within their uncertainties, and it is necessary to distinguish between model inadequacies and the model response uncertainties caused by poor knowledge of the input data.

This applies in particular to tropical motions, which are primarily wind forced and could be simulated deterministically for the most part if the wind stress were accurately known. However, information on surface wind is sparse and noisy, and the bulk formulae used to estimate the wind stress are rather inadequate. As equatorial model simulations have become increasingly realistic, visual comparisons, which reveal obvious differences, are unable to unambiguously identify model inadequacies, and more refined validation procedures are needed. Frankignoul et al. (1989) have thus developed a model testing method based on multivariate statistical analysis that could take into account explicitly all the observational uncertainties. So far, the method has been used to test and intercompare simulations of the tropical ocean. However, the approach is general and could be used to test dynamical models of the internal wave field, as suggested below.

In the next section, the model testing method is described in a general manner. A possible application to internal wave studies is then briefly discussed. In section 4, the method is illustrated by summarizing the testing and intercomparison of model simulations of the tropical Atlantic.

2. The multivariate model testing method

The simulation of oceanic motions can often be represented as an input-output problem. A model is driven by a prescribed forcing field $f(x,t)$, which could include the initial conditions, and it predicts in particular the space/time behavior of a variable that is also observed. We denote the model prediction by m , with $m = L(f)$. To test the model, m is considered in the whole x - t domain where the model is believed to be realistic and observations, denoted by d , are reliable, which is normally a highly dimensional space. The observations are usually inaccurate, so a probability region more properly describes the true oceanic state (Fig.1). As the forcing field is generally not

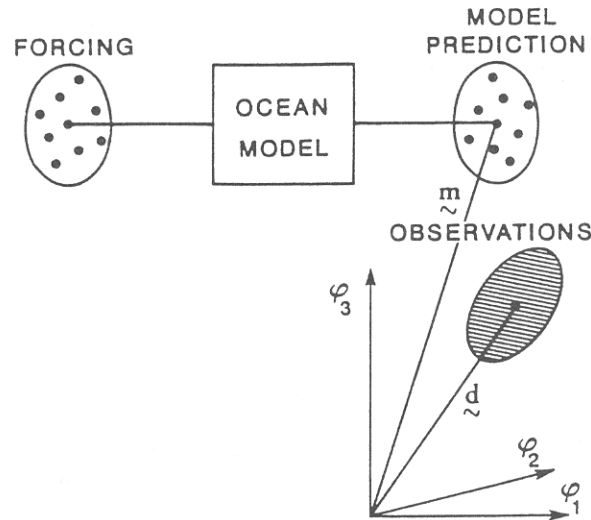


Fig. 1 Sketch of the effects of data uncertainties on the comparison between model predictions and observed data, from a multivariate statistical viewpoint.

known accurately, it should be considered as a (noisy) realisation of the true forcing field, and other equally plausible input data often are available (e.g., different wind stress products) or could be generated by Monte Carlo techniques. This is represented in Fig.1 by a probability region for the true forcing $\langle f \rangle$. Thus, several equally plausible model responses can be generated, and the model prediction \mathbf{m} really should be considered as the sum of the (unknown) model response to the true forcing field, say $\langle \mathbf{m} \rangle$, and that resulting from the input errors, which is again represented by a probability region in Fig. 2. If the multinormal assumption holds, these uncertainties can be described by their error covariance matrix, denoted by \mathbf{D} for the observations and \mathbf{M} for the model predictions. The model testing problem can now be viewed as that of comparing two noisy vectors with unequal error covariance matrices. It should be stressed that a crucial, and often cumbersome, step in the procedure is the evaluation of the two error covariance matrices; this may require much data analysis and many simulations.

The agreement between the observations and the simulations is characterized by the misfit

$$T^2 = (\mathbf{m} - \mathbf{d})' (\mathbf{M} + \mathbf{D})^{-1} (\mathbf{m} - \mathbf{d}) \quad (1)$$

which provides a measure of the differences between the two fields, weighted by the data uncertainties. If the null hypothesis that $\langle \mathbf{m} \rangle = \langle \mathbf{d} \rangle$ holds, (1) is the appropriate test statistic which is distributed as a χ^2 variable (or a Hotelling T^2 variable if sample estimates of the error covariance matrices are used), and the usual acceptance and rejection rules can be applied. If the null hypothesis does not hold, T^2 is approximately distributed as a non-central χ^2 variable, and confidence intervals can be constructed to compare different models or model versions. Details are given in Frankignoul et al. (1989).

In practice, however, this cannot be done unless the dimensionality of the system is first strongly reduced. Indeed, the error models normally are very approximate (they are either estimated

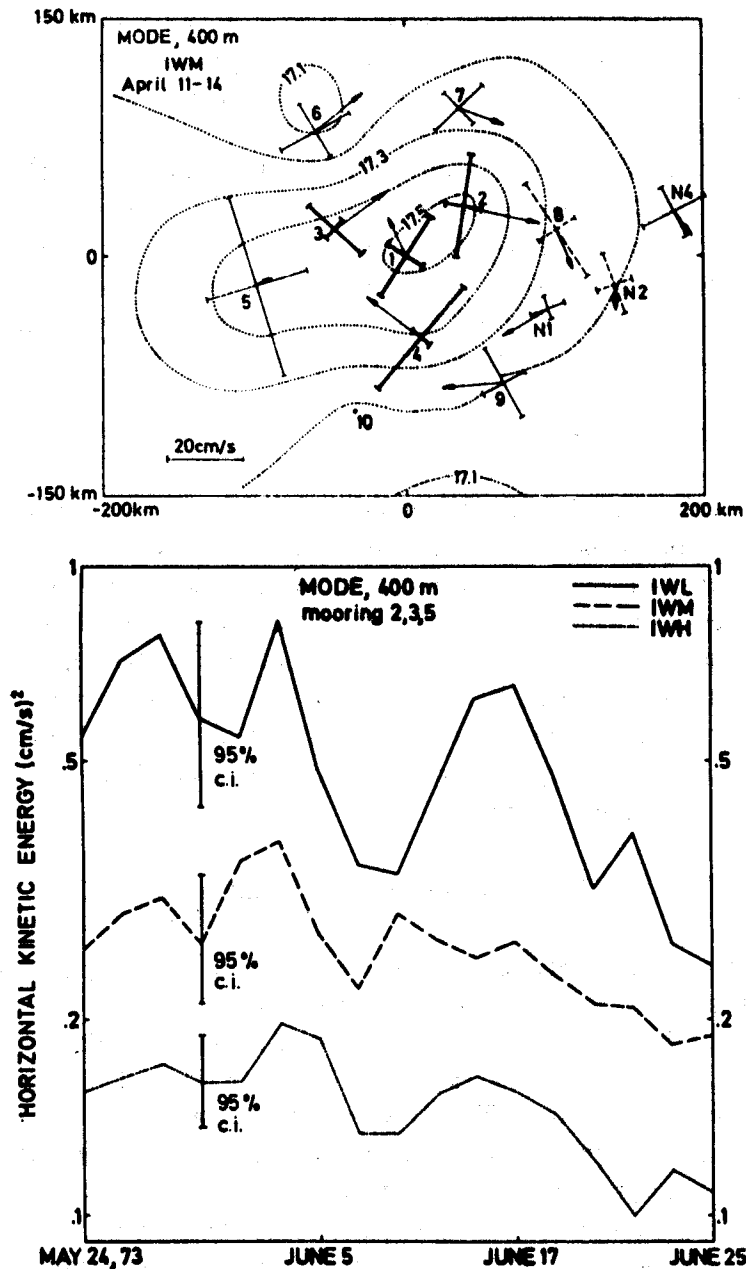


Fig. 2. *Top*: Ellipse principal axes of the internal wave horizontal kinetic energy (HKE) at 400 m depth for the internal wave frequency band (0.2-0.4 cph) in a 4-day period during MODE (the central mooring is at 28°N, 69°40'W). Arrows show 4-day mean current, and dotted lines corresponding isotherms. *Bottom*: Time series of HKE averaged over the line of moorings 2, 3 and 5 for three frequency bands in the internal wave continuum. From Frankignoul (1976).

from a limited sample or from a necessarily oversimplified error analysis), so that the inverse in (1), even if it exists, is dominated by unreliable information. A strong data compression is thus needed for the problem to be well-posed. This can be done efficiently using principal component analysis, as illustrated below. The main point is to perform the model-reality comparison in a sufficiently small orthonormal space where the main "trustable" features of both model simulations and observations are well represented, and the error models sufficiently reliable. Noisy details are filtered out, but of course so is a small part of the signal. Note that the test is very stringent as no model is expected to be perfect. However, much progress may be expected from an understanding of the discrepancies between models and observations, and the method is very efficient at distinguishing the true differences from those due to data uncertainties.

3. Application to internal wave studies

These concepts could be applied to the internal wave case. Suppose for example that we want to test a model of the dynamics of the internal wave field, which requires understanding their sources, sinks, and main interactions. Theoretical models of internal waves do not yet predict how the various energy sources control their spectral distribution, but plausible generation mechanisms have been suggested (wind forcing, topographic scattering, interaction with the mesoscale shear flow, forcing by the barotropic tides, etc.), and predictions of the spatial and temporal evolution of the averaged internal wave properties could be obtained. In particular, the variations of the total internal wave energy $E(\mathbf{x},t)$, should obey the radiation balance equation (Muller and Olbers, 1975) which can be written in the approximate form

$$(d_t + \mathbf{U} \cdot \nabla) E = S^i(\mathbf{U}) + S^i(\text{atmosphere}) + S^i(\text{bottom}) + \dots + S^d \quad (2)$$

where $\mathbf{U}(\mathbf{x},t)$ is the mean current, S^i denotes the source terms, and S^d the dissipation. If models are available for the source terms, they can be tested by comparing the predictions from (2) with corresponding observations. Internal wave spectra have been shown to be modulated both on the seasonal scale (Fig. 2), and on short time and space scales (Fig. 3); also, evidence of energy propagation and/or transfer has been found (Fig. 4). These observations contain critical clues on internal wave energy sources and sinks, but their interpretation has been disappointing, as the dynamics is complex and the signal-to-noise ratio very low.

We believe that the key to a successful interpretation will be a multivariate model testing approach where predicted changes are compared to the observed ones, as this strongly enhances statistical significance in model-observation comparisons (see, e.g., Hasselmann, 1979). However, the input in (2) will be poorly known: the wind stress is difficult to observe at the internal wave scales, the mean shear can only be coarsely estimated from current-meter data, the bottom topography and bottom currents are inaccurately known, and little information will be available on the boundary conditions for the region of interest. Thus, there will be a large uncertainty in the forcing data that will have to be considered in addition to that of the internal wave spectral estimations, when testing the theoretical predictions. The method of section 2 should then be applicable, even if some adaptation to the problem will be needed. On the other hand, neglecting the effects of data uncertainties or the multidimensionality of the fields could result in erroneous conclusions.

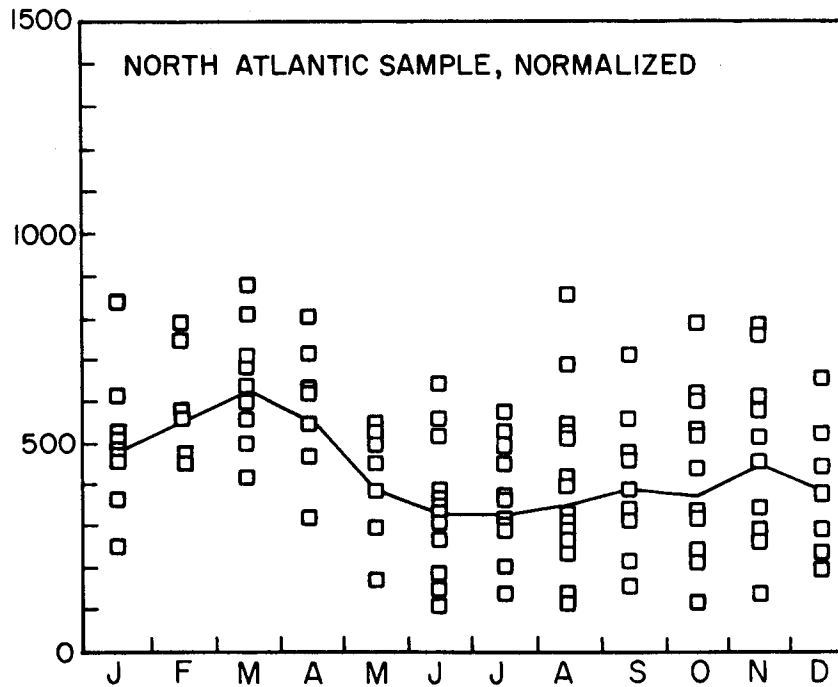


Fig. 3. Monthly high-frequency (0.1-2 cph) HKE from 17 data sets in the North Atlantic, normalized by the Brunt-Väisälä frequency, versus time of year. The solid line connects the monthly averages. From Briscoe (1984).

One remark is in order. Dynamical models of the internal wave field will unavoidably contain parameterizations and adjustable parameters (e.g., relaxation or dissipation time). A poor choice in the values of the arbitrary parameters may also be responsible for model-observation discrepancies, without impairing the model validity. Thus, it is of interest to include model tuning in the statistical method. If there is only one or two parameters, their influence on the misfit (1) can be easily found, with the optimal choice corresponding somehow to its minimum (see an example in Frankignoul et al., 1989). However, if there are many parameters, the problem becomes tedious. Recent efforts at combining model testing and parameter optimization by inverse methods are underway, and they may also prove useful in the internal wave case.

4. Application to the seasonal variability of the tropical Atlantic

To illustrate the model testing method, let us briefly consider its application to the numerical modeling of the tropical circulation. Specifically, we want to verify whether the OPA general circulation model of LODYC is able to simulate the mean seasonal changes in the surface dynamic topography of the tropical Atlantic, approximated by the 0-400 db dynamic height. The observed changes have been recently estimated from historical data by Duchêne and Frankignoul (1991), who performed an extensive error analysis to also estimate their error covariance matrix. The observations are compared in Fig. 5 (left) to the prediction of the OPA model (right). The simulation is the mean of a run forced by 1982-1988 monthly observed winds (Morlière and

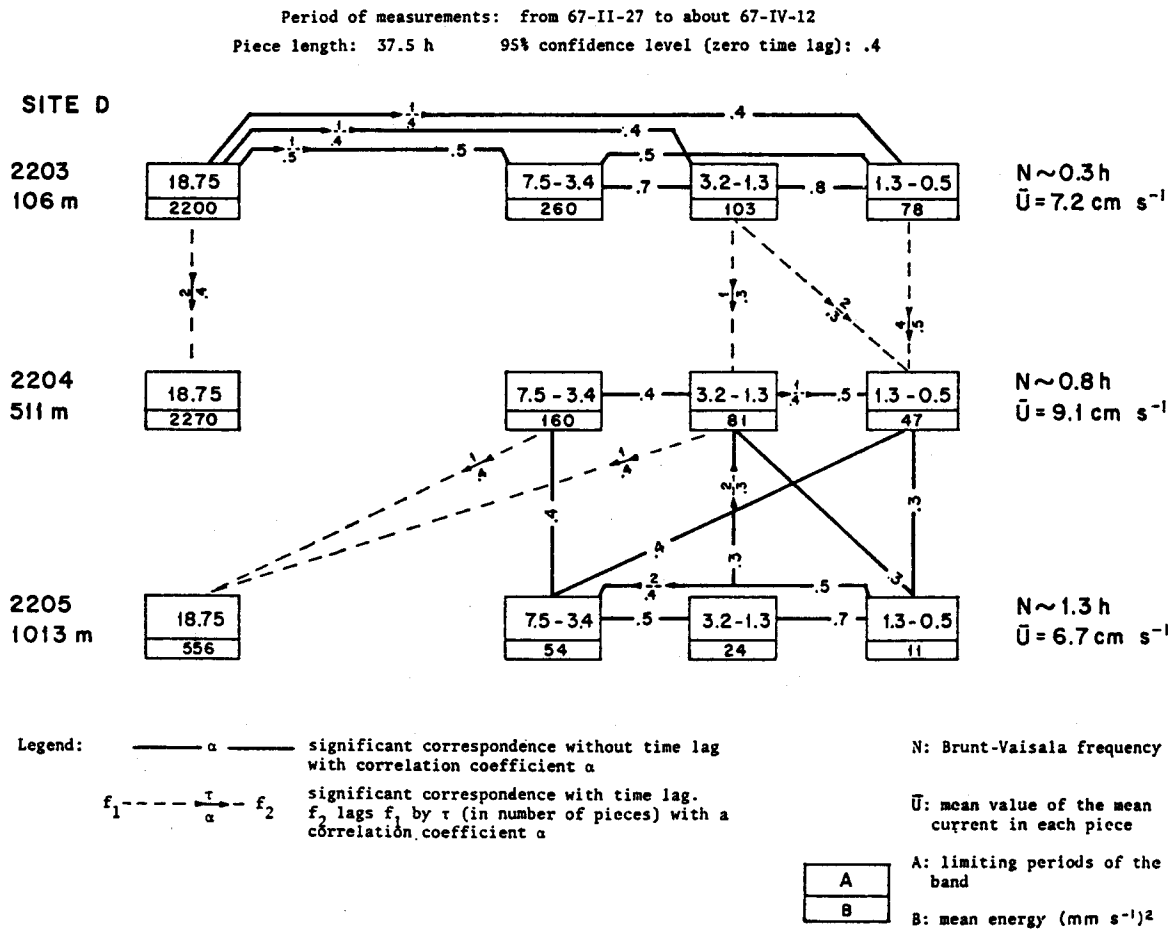


Fig. 4. Schematic representation of energy correspondences in frequency-depth space observed at Site D ($39^\circ 20' \text{N}$, 70°W) during a two-month period. From Frankignoul (1974).

Duchêne, 1991), so that the effects of the random wind stress errors and the interannual variability are automatically represented in the error covariance matrix that is calculated from the 7-year model response. The error associated with our lack of knowledge of the drag coefficient for the wind stress is also taken into account in a simplified manner by assuming a 20% random uncertainty in its value, as in Frankignoul et al. (1989).

The problem is to establish whether the two series of 12 monthly maps that describe the mean seasonal changes in space and time are consistent with each other, i.e., are within error bars. The overall dimension is high (number of grid points times number of months), and the noise correlated (due to data interpolation, forcing errors, etc.). Also, the sample used to estimate \mathbf{M} is small, and \mathbf{D} only is an idealized error model. Thus, the details of \mathbf{D} and \mathbf{M} are unreliable and a strong data compression is required. It is efficiently done by using common principal component analysis: the first four common empirical orthogonal functions (EOFs) (Fig. 6) are sufficient to represent about 90% of the mean simulated changes around the annual mean and 80% of the (more noisy) observed ones. Only two additional orthonormal vectors are needed to also represent the annual averages,

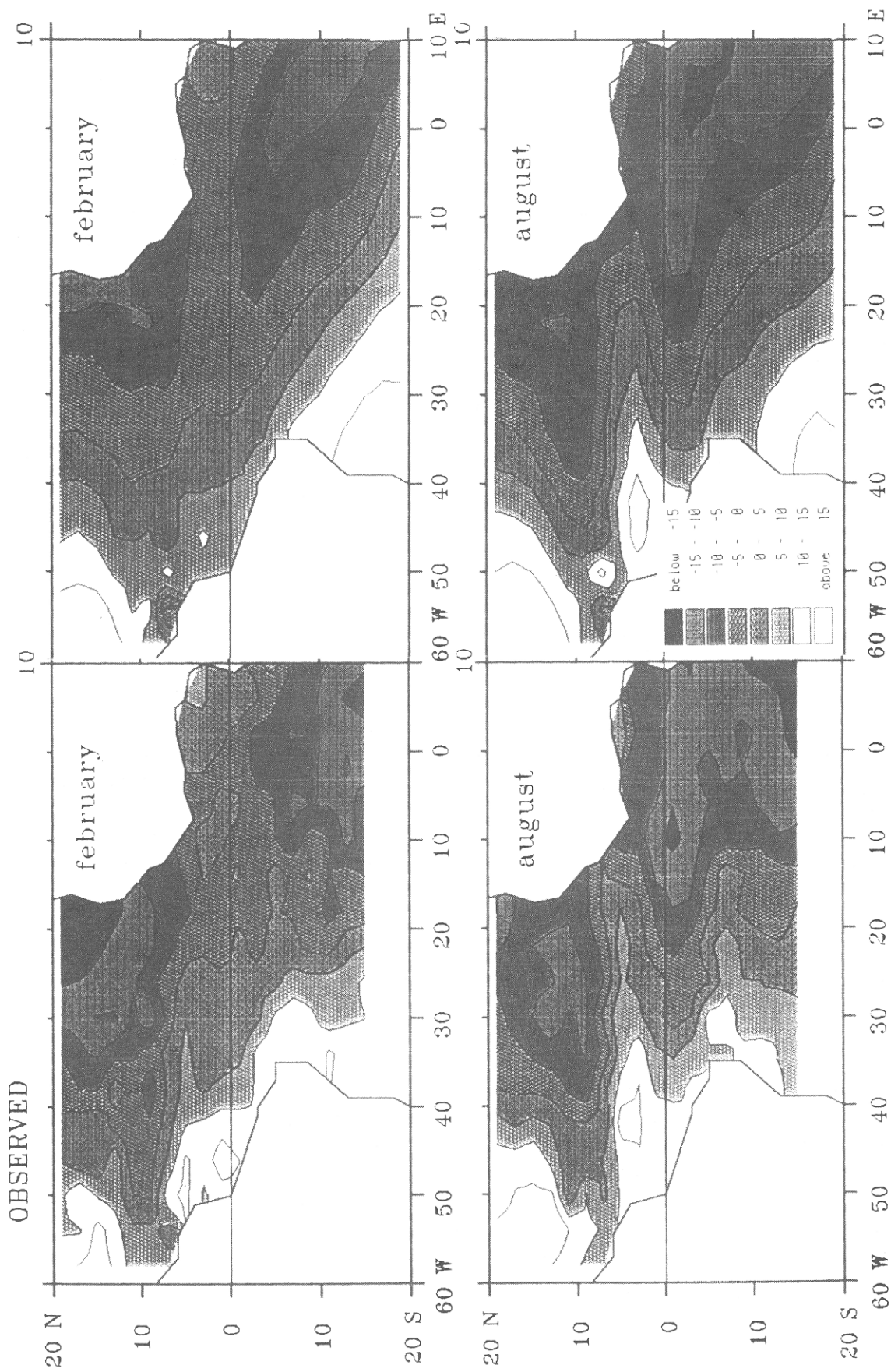


Fig. 5. *Left:* mean observed 0-400 db dynamic topography in dyn. cm. for February (top) and August (bottom). *Right:* Corresponding prediction using the OPA general circulation model. In both cases, the space-time mean has been subtracted.

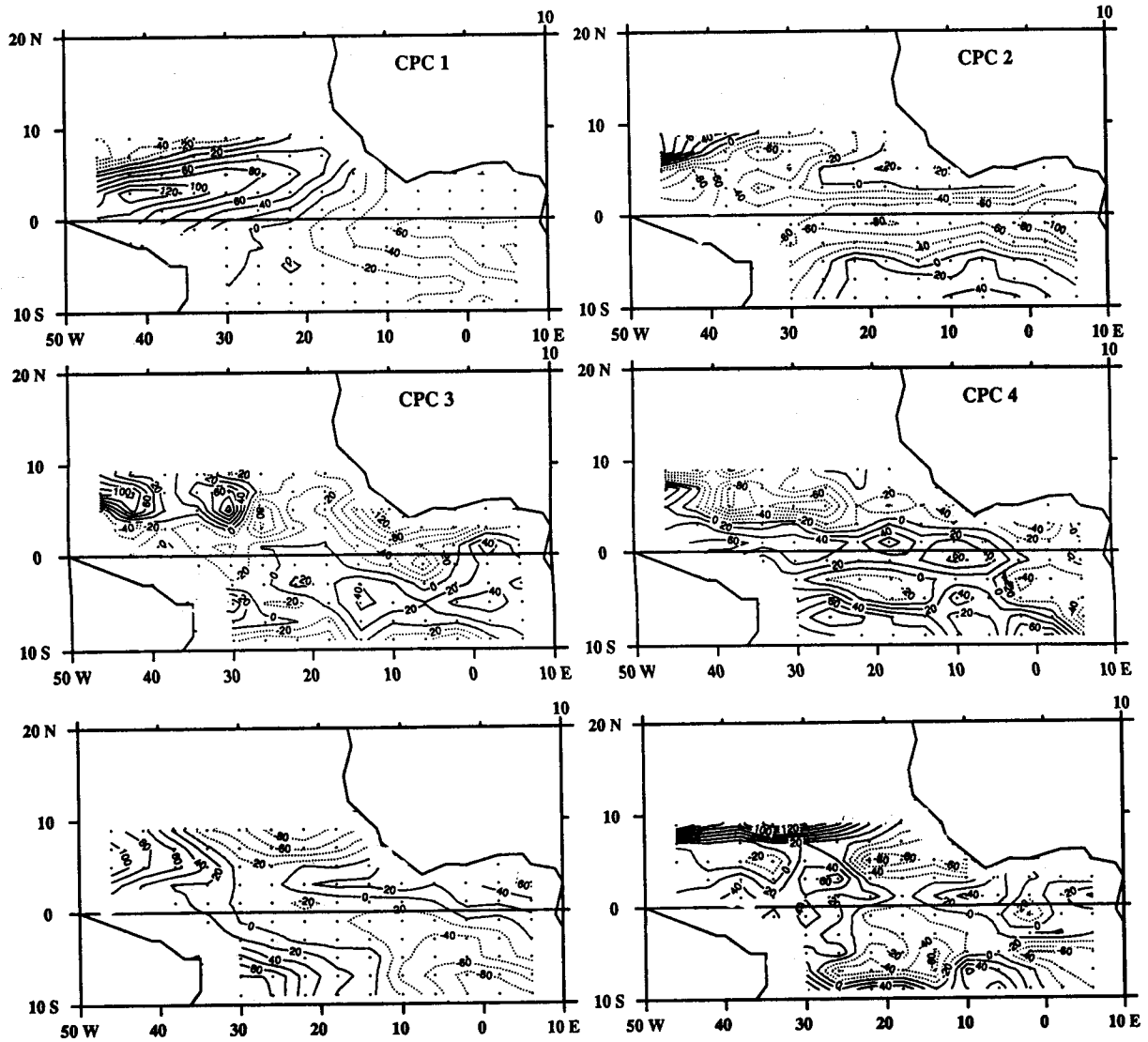


Fig. 6. The six orthonormal basis vector for the model-observation comparison. The top four vectors are the common EOFs of the seasonal variations.

which project poorly onto the common EOFs. In Fig. 7, the observed (dashed line) and simulated (continuous line) seasonal variations are represented in the reduced space, together with 95% confidence intervals estimated from the diagonal terms of the error covariance matrices, assuming normality. Note that the latter are based on univariate statistics and do not represent error correlation. Nonetheless, Figure 7 suggests that the differences between the observations and the simulations in Figure 5 cannot be entirely explained by the data uncertainties. This is confirmed by the results of the statistical test in Figure 8 with (left) and without (right) the annual mean. Note that a further data compression was done in the time domain by considering only four seasons (the dimension is thus 4×4 and 6×4 , respectively).

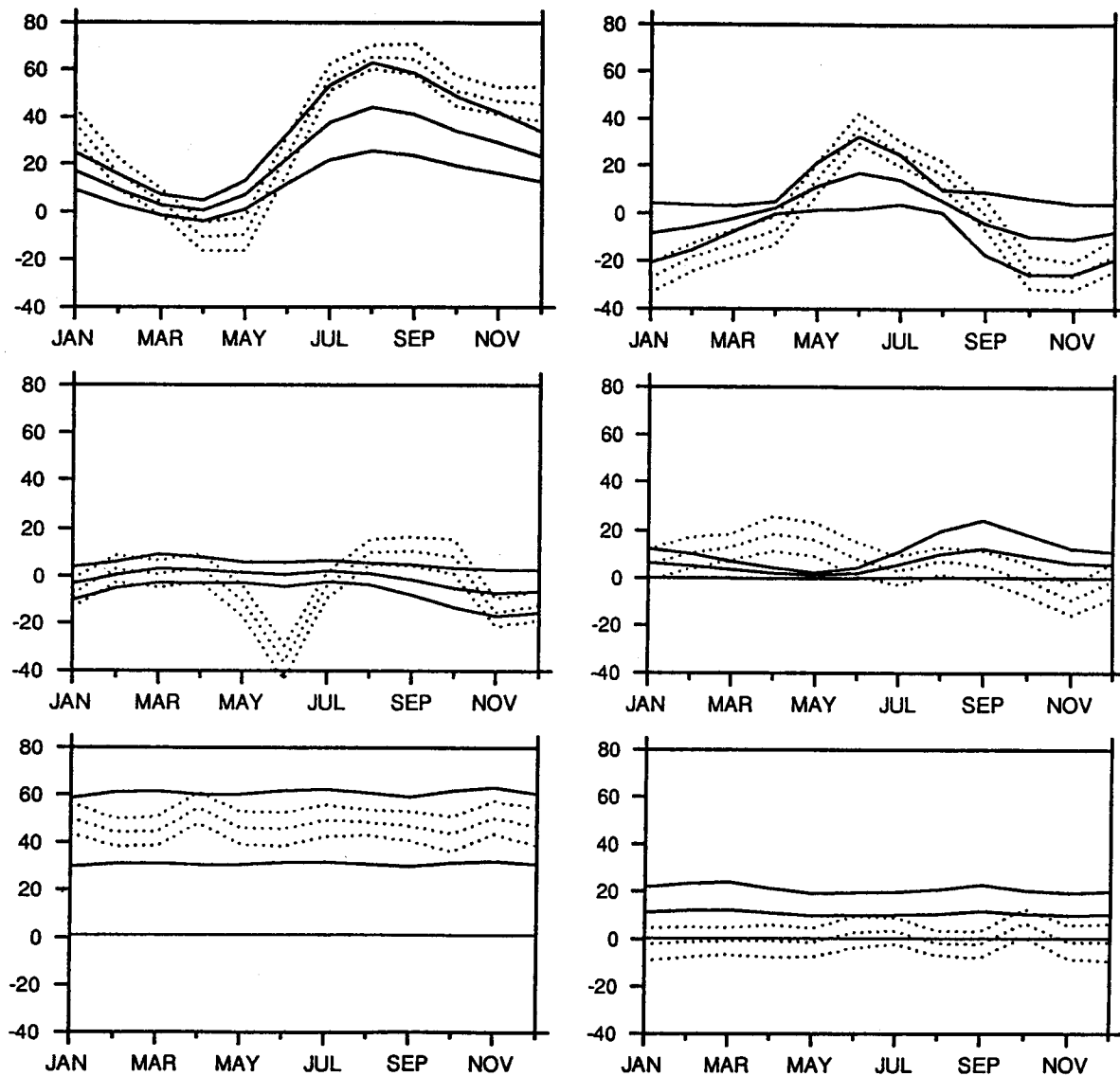


Fig. 7. Observed (dotted line) and modeled (continuous line) seasonal variations in the reduced space. The error bars are univariate estimates of the 95% confidence intervals.

The results show that the model is not consistent with the observations, as the misfit is much larger than the critical value T^2_{crit} for perfect consistency at the 5% level. Recall that (1) is a measure of the (square) difference between simulations and observations, normalized by the data uncertainties, and not an absolute measure of the fit; if the data uncertainties had been larger, T^2 would be smaller but the model would not perform better. The test only shows that the model-observation differences are about three (no mean) or four (with mean) times (i.e. $(T^2/T^2_{crit})^{1/2}$) larger than expected from the data uncertainties, at the 5% level. This indicates that there remains substantial room for model improvements. The results also show that the GCM performs better for

the mean seasonal changes around the annual mean than for the whole signal. This has also been found for all other models and variables that have been considered, and presumably reflects the fact that the long-term mean depends more on the representation of the dissipation processes, which is generally very approximate.

Figure 8 also represents for comparison the test results for two simpler models, the linear, multi-mode model of Cane (1984) and the 2-layer nonlinear LODYC model (Février, in preparation). The test values show that the tropical Atlantic data are accurate enough to distinguish between the performances of different oceanic models, and the OPA model performs *significantly* better than the two other models, consistently with its much higher sophistication. This illustrates the usefulness of the model testing approach, and the interest of quantifying as realistically as possible all the data uncertainties.

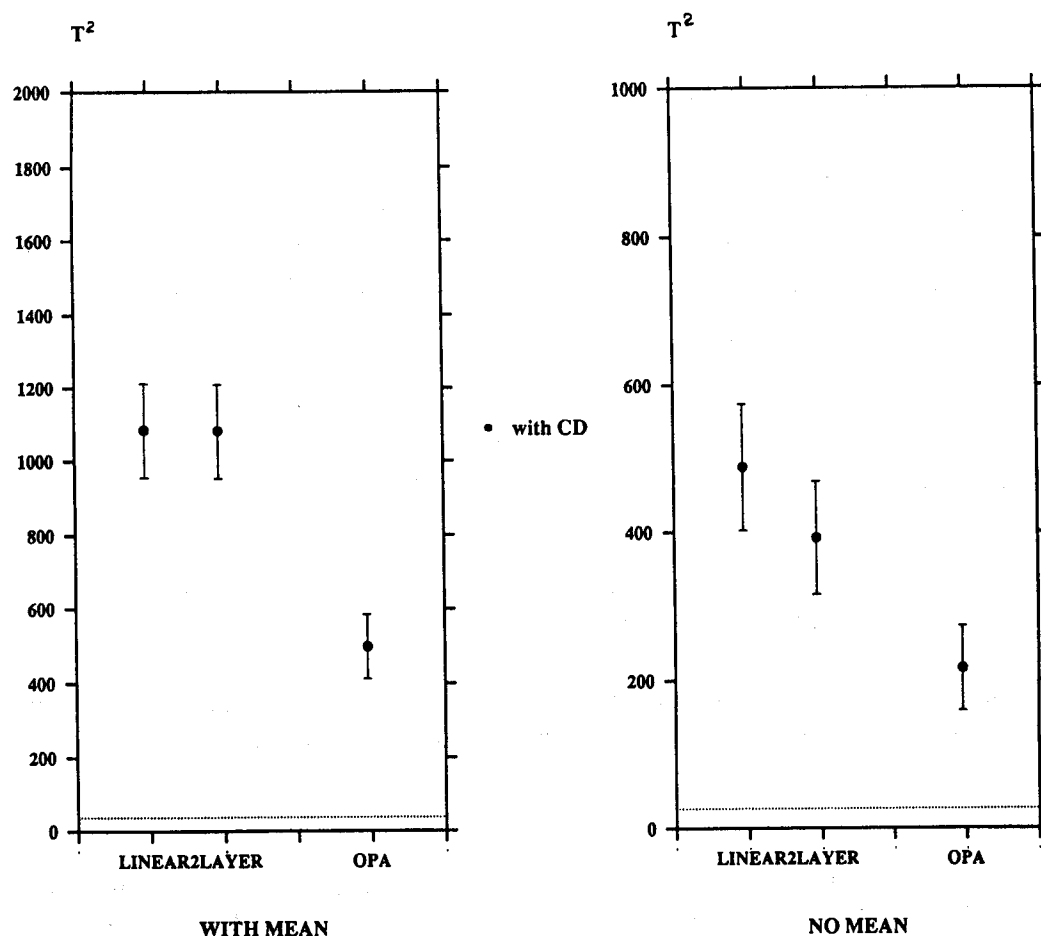


Fig. 8. Value of T^2 for three models (see text) used to simulate the seasonal cycle of surface dynamic topography in the tropical Atlantic with (left) and without (right) annual mean. The error bars represent approximate 95% confidence intervals, the dashed line the critical value for T^2 at the 5% level.

Acknowledgments

Some of the results reported here come from unpublished work performed at LODYC by Christine Duchene, Alain Morlière and Sabine Février who are gratefully acknowledged. I would also like to thank Peter Müller for stimulating discussions.

References

- Briscoe, M.G., 1984: The monthly variability of the upper ocean internal wave energy: a progress report on the correspondence with wind stress. *Internal Gravity Waves and Small-scale Turbulence*, Proceedings of the 'Aha Hulikoa'a Winter Workshop, P. Muller and R. Pujalet, eds. Hawaii Institute of Geophysics, Honolulu, po. 129-150.
- Cane, M. A., 1984: Modeling sea level during El Niño. *J. Phys. Oceanogr.*, **14**, 1864-1874.
- Duchêne, C., and C. Frankignoul, 1991: Seasonal variations of surface dynamic topography in the tropical Atlantic: Observational uncertainties and model testing. *J. Marine Res.*, to appear.
- Frankignoul, C., 1974: Preliminary observations of internal wave energy flux in frequency, depth-space. *Deep-Sea Res.*, **21**, 895-909.
- Frankignoul, C., 1976: Observed interaction between oceanic internal waves and mesoscale eddies. *Deep-Sea Res.*, **23**, 805-820.
- 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.
- Hasselmann K., 1979: On the signal-to-noise problem in atmospheric response studies. *Meteorology of the Tropical Ocean*, D.B. Shaw, ed., Roy. Meteor. Soc., 251-259.
- Morlière, A. and C. Duchêne, 1991: Progress in simulating and objectively evaluating the surface currents in the tropical Atlantic. *J. Geophys. Res.*, submitted.
- Muller, P. and D.J. Olbers, 1975: On the dynamics of internal waves. *J. Geophys. Res.*, **80**, 3848-3860.