Time Interpolation of Forcing Fields in Ocean Models

PETER D. KILLWORTH*

NERC Oceanography Unit, Clarendon Laboratory, Oxford, England

3 January 1995 and 12 May 1995

ABSTRACT

It is shown that the usual practice of forcing ocean models by linear interpolations of monthly mean data values does not produce a forcing whose mean over a month is the data value required. For wind stress data this can yield monthly mean errors, coherent over a basin, of as much as 0.4×10^{-1} N m^{-2}, with 30%–40% of values being in error by more than 10%. A simple method is given to avoid the difficulty, which involves no change to model computer code and no increase in the amount of data stored internally.

1. Introduction

Ocean model simulations run in realistic geometries need accurate forcing by surface fluxes (winds, heating, evaporation, etc.) if they are to give accurate answers. These forcing fields are provided either from observations or from atmospheric general circulation model output. In both cases, these fields are almost invariably given as averages over some timescale ranging from annual, in the case of poorly known fields, to monthly or better for reasonably well observed fields. Challenor and Carter (1994) discuss the accuracy of these average fields as a representation of reality, and a complete discussion can be found in Weller and Taylor (1993). There are many potential errors involved in obtaining forcing fields, for example, in estimating fluxes by using bulk formulas on time averages of data fields, rather than time averages of bulk formulas applied to data fields. Spatial averaging also causes difficulties. However, ocean modelers perforce ignore these issues and must assume their data source to be, in some sense, "correct." The entirely separate question addressed here is how accurately these given fields are used within models.

The ocean model driven by these fields requires their values at each time step, typically some small fraction of a day in length. Thus, a degree of interpolation of the forcing fields must be made in time (and, typically in space as well, but that does not concern us here). The simplest interpolation is to use the monthly mean data value as the forcing for that month. This has the obvious tendency to remove extreme heating and cooling at the surface, which can be a difficulty when bottom water should be produced by the model. However, the main difficulty occurs with wind stress, in that the abrupt change in stress at month ends causes unwanted inertial oscillations to occur. Thus, most workers prefer a smoother method of interpolation. Because storage or input/output limitations on many fine-resolution ocean models often preclude the simultaneous storage of the full time series of the two-dimensional forcing fields, the interpolation often involves only a few of the data values, with the most popular solution being a linear interpolation between two successive values. This is the strategy taken by, for example, Latif (1987), New et al. (1993), the FRAM Group (1991), Ezer and Mellor (1992), Gerdes and Wübber (1991), and many others. Sometimes Hermite interpolation between data values is employed (Bleck and Smith 1990). It is relatively infrequent to find, for example, actual 3-day mean wind stresses used with no interpolation (cf. Semtner and Chervin 1988).

The important thing to note about these methods is that consistently they involve some interpolative technique used between mean values of the data over some interval. This note raises the point that such interpolation does not actually generate forcing data whose mean over the relevant interval is indeed the mean originally supplied in the data. Then we ask, does the difference matter, and if so, what can be done to improve matters?

Epstein (1991) has introduced the problem, noting that it has of course been recognized for decades, but as the examples above prove, the problem is usually ignored. Epstein's straightforward solution (fitting annual and semiannual harmonics through the means) requires many data values at each spatial point and is

* Additional affiliation: Institute of Oceanographic Sciences, Deacon Laboratory, Wormley, Godalming, Surrey, England.

Corresponding author address: Dr. Peter D. Killworth, NERC Oceanography Unit, c/o Dept. of Atmos., Oceanic and Planetary Physics, Clarendon Laboratory, Parks Road, Oxford OX1 3PU, United Kingdom.
E-mail: atmpk@vax.ox.ac.uk

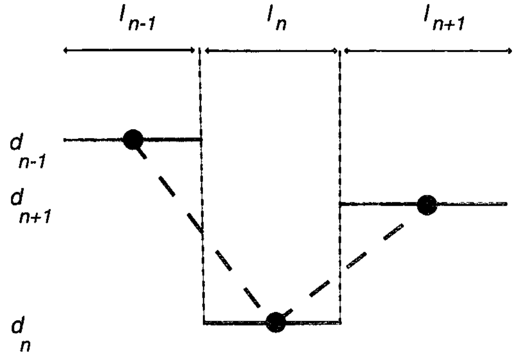


FIG. 1. Schematic of linear interpolation. Mean data values are denoted by d_n, and time intervals by l_n. Heavy dots indicate the month centers, and the heavy dashed lines show the interpolation normally used.

often not viable for high-resolution models where storage, as noted, is at a premium, and so some other method needs to be found.

This note examines the question of interpolation for monthly means, with an example from the European Centre for Medium-Range Forecasting global wind stresses, and finds that the errors are significant. We then provide a simple solution to the errors induced that involves no modification to existing codes and no increase in the amount of data to be stored.

2. Interpolation of data between temporal means

Consider the simple scheme shown schematically in Fig. 1. Month n in the data has length l_n days, and a data value d_n (wind stress, surface temperature, etc.), which can be represented as a vector $\mathbf{d} = (d_1, d_2, \cdots, d_N)$. There are N months in the year.[1] The simplest, and most common, interpolation scheme is a linear interpolation between the data values applied at midmonth, as shown. This yields a mean value over month n, which can be represented in matrix form as

$$\mathbf{a} = \mathbf{Ad}, \tag{1}$$

where the vector $\mathbf{a}$ represents the monthly average of the forcing applied, and $\mathbf{A}$ is given by a tridiagonal matrix (with fringes) of the form

$$\mathbf{A} = \begin{pmatrix} f_1 & g_1 & 0 & 0 & 0 & & 0 & e_1 \\ e_2 & f_2 & g_2 & 0 & 0 & & 0 & 0 \\ 0 & e_3 & f_3 & g_3 & 0 & & 0 & 0 \\ & & & & & \ddots & & \\ g_N & 0 & 0 & 0 & 0 & & e_N & f_N \end{pmatrix}, \tag{2}$$

where, with due annual cyclicity in suffices,

$$e_n = \frac{l_n}{4(l_{n-1} + l_n)}$$

$$f_n = \frac{l_{n-1}}{4(l_{n-1} + l_n)} + \frac{1}{2} + \frac{l_{n+1}}{4(l_n + l_{n+1})} \equiv 1 - e_n - g_n$$

$$g_n = \frac{l_n}{4(l_n + l_{n+1})}. \tag{3}$$

In the special case of months of equal lengths (often assumed for simplicity in data processing), the coefficients become 1/8, 3/4, and 1/8, respectively.

The matrix $\mathbb{A}$ takes a more complicated form for other interpolation schemes, but remains in general a smoothing operator (except possibly for cubic spline interpolation, which can yield interpolants beyond the range of the data; this is seldom used). In any case, operating with $\mathbb{A}$ on the mean monthly data cannot produce the (identical) mean monthly data because $\mathbb{A}$ is not the identity matrix. It is also clear from Fig. 1 that extrema in the mean data are eroded by this process. More generally, the error in the monthly means can be represented approximately as $-(1/8)l_n^2(d^2\mathbb{d}/dt^2)$ for linear interpolation on equal length months, and one-third of this for Hermite interpolation.

3. A global wind stress dataset

How large is the error induced by the practice of incorrect interpolation? This depends on the degree of smoothness of the dataset in general. Here we consider the 12-month annual mean ECMWF global wind stress data (ECMWF 1993).

The error induced by incorrect interpolation can be computed by examining the inferred monthly means as either an absolute error or a fraction of the actual

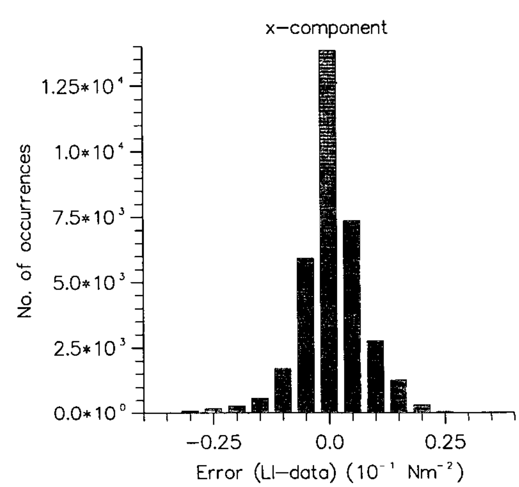


FIG. 2. Histograms of the error in estimates of monthly mean eastward wind stresses produced by interpolation, binned by 0.05×10^{-1} N m^{-2}, for January.

[1] The suggestive notation "month" and "year" are used in what follows, but clearly any other intervals could be used.

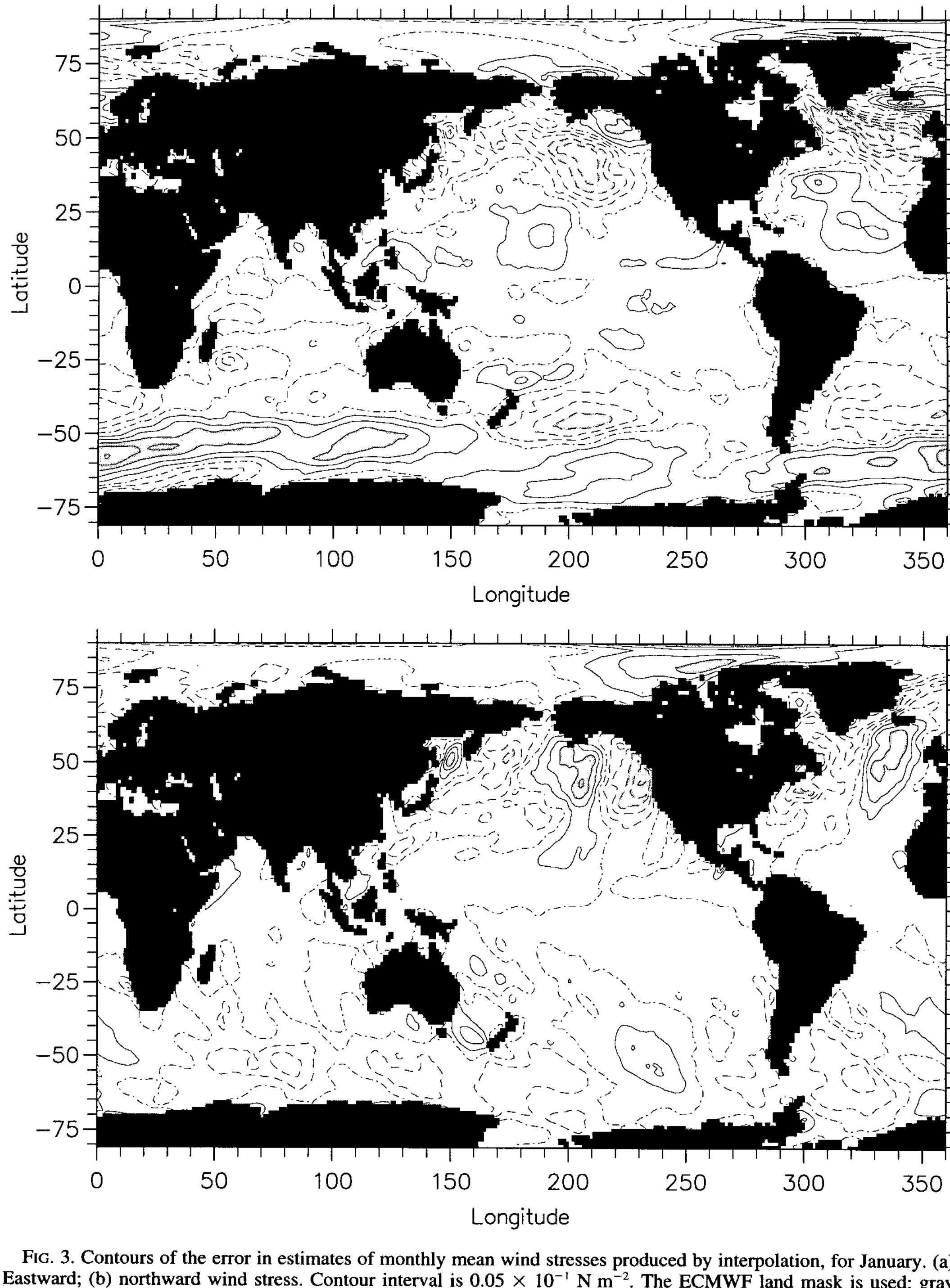
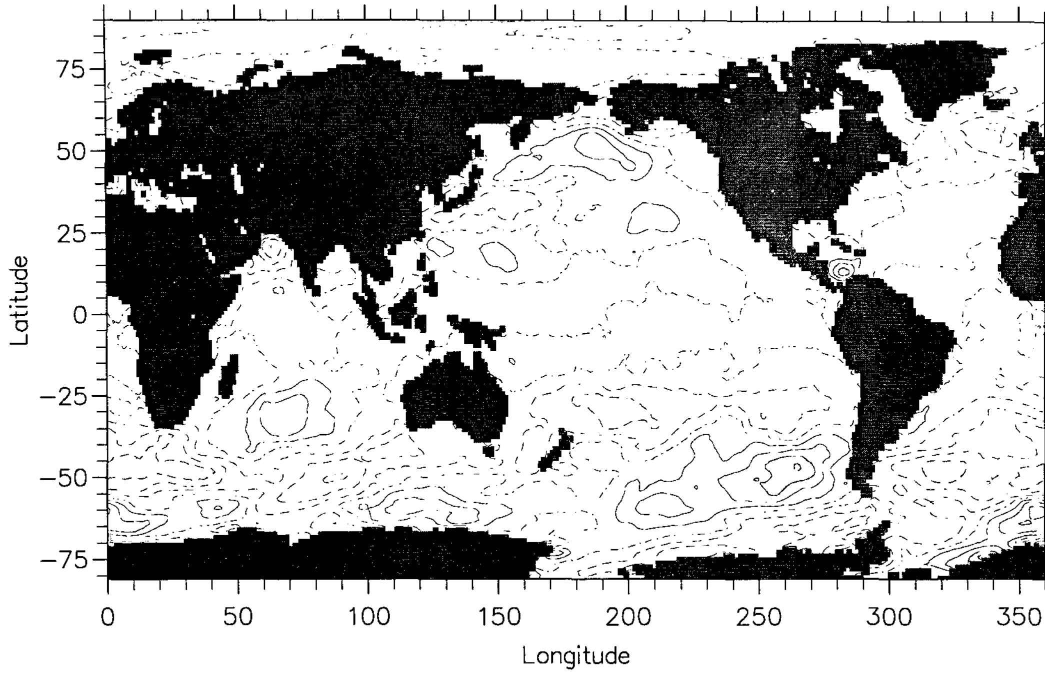


FIG. 3. Contours of the error in estimates of monthly mean wind stresses produced by interpolation, for January. (a) Eastward; (b) northward wind stress. Contour interval is 0.05×10^{-1} N m^{-2}. The ECMWF land mask is used; gray shading shows absolute errors above 0.15×10^{-1} N m^{-2}. Firm lines show positive values; dash–dotted the zero contour, and dashed lines show negative values.

monthly means. Of the eastward stresses, 21.1% are more than 10% lower than the correct values, and 17.2% are more than 10% higher than correct.[2] For the northward stresses, the figures are 28.7% and 19.3%, respectively.

Probably more relevant are the absolute errors. Figure 2 shows histograms of the errors in x component of the monthly means, centered on bins of width 0.05×10^{-1} N m^{-2} for January. Other months and components

[2] The figures quoted are consistently computed solely over ocean points.



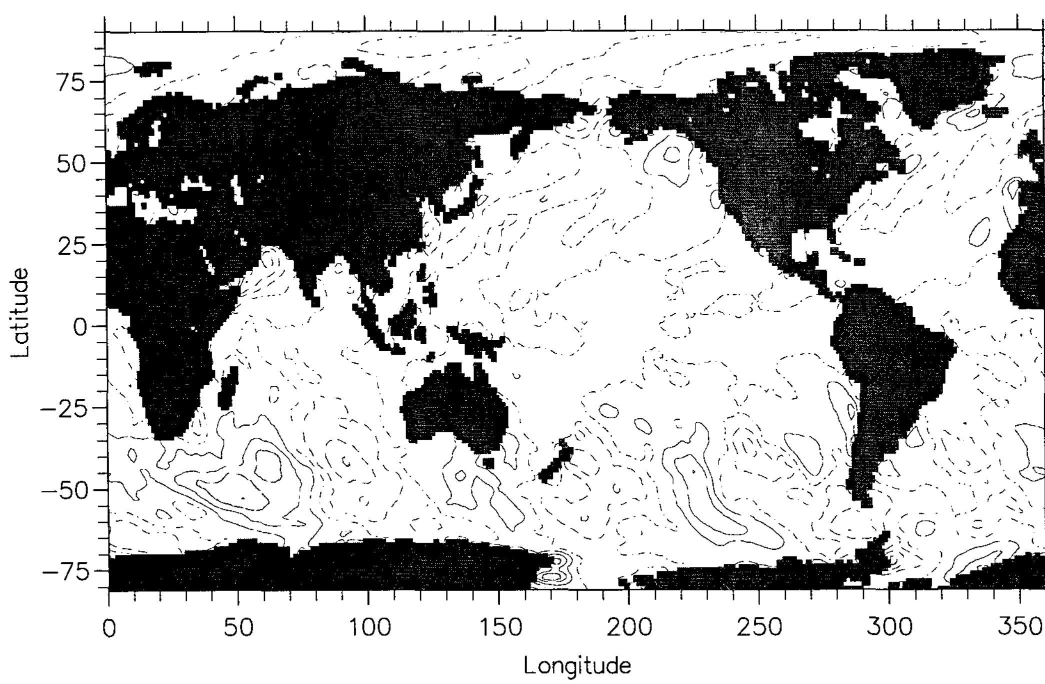


FIG. 4. As for Fig. 3 but for July.

are similar. The majority of the errors are seen to be small. However, the spatial distribution of these errors shows that the extreme errors (of order 0.25 $\times$ 10^{-1} N m^{-2}, roughly) are concentrated in single ocean basins (Figs. 3 and 4) and of uniform sign over much of the basin. An example is the North Atlantic in January, when the eastward stress average is in error across most of the subpolar gyre. A rough estimate based on Fig. 3a leads to an erroneous Sverdrup flux of up to 5 $\times$ 10^6 m^3 s^{-1} for that month. If concentrated in a western boundary layer of width 60 km and depth 5 km, this would imply a barotropic western boundary current of order 1.6 cm s^{-1}, which would move tracers 0.4° of latitude during the month. If the current were baroclinic and concentrated in the top 500 m, the figure would be 4° instead.

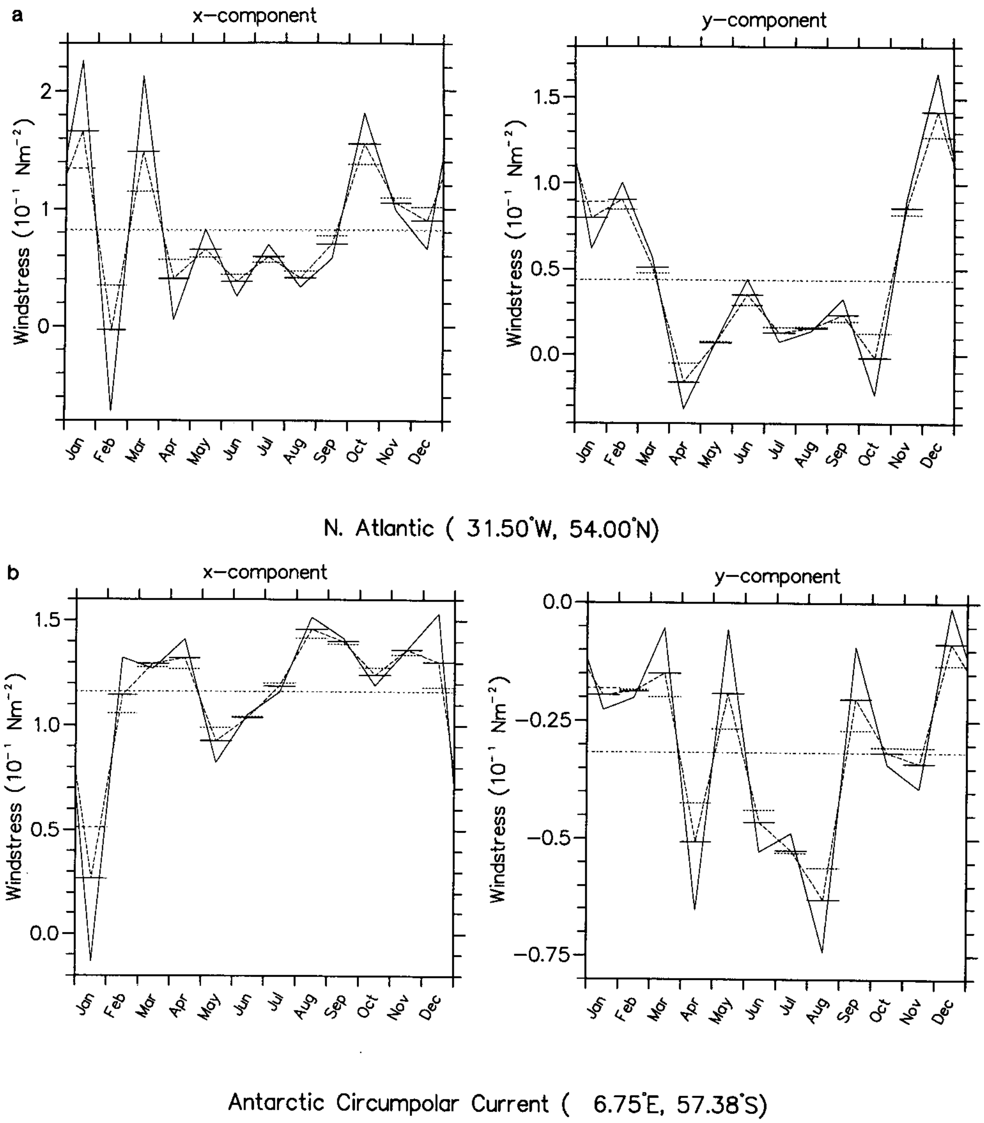


FIG. 5. The 12-month cycle of errors in the monthly mean produced by interpolation, for sites in (a) the North Atlantic (54°N, 31.5°W), (b) the Antarctic Circumpolar Current (57.38°N, 6.75°E), (c) the equatorial Pacific (9°N, 157.5°W), (d) the northwest Indian Ocean (11.25°N, 57.38°E), and (e) the northeast Pacific (42.75°N, 129.38°W). The firm bars show the actual monthly mean; the dashed bars the incorrect value obtained by interpolation between monthly means. The firm line shows the interpolation between values that gives the correct monthly means [computed from (5)]; the dashed line shows the incorrect interpolation for comparison. The long horizontal dotted line shows both the annual mean and the incorrect annual mean, which are indistinguishable on this scale.

Figure 5 shows the 12-month cycle of these errors in the mean at selected locations in the World Ocean. Errors of 0.2 to 0.4 ($\times 10^{-1}$ N m^{-2}) are found in most cases at some stage in the annual cycle. These occur, as predicted, at extrema in the cycle; at other times the errors are small (certainly smaller than the uncertainty in the data). Errors tend to be smaller in the Tropics, due to the smoothness of the variation, although monsoonal changes cause some errors in the Indian Ocean. Even the annual mean of the forcing is in error, if the months are of different lengths, although this error is very small. It is straightforward to show that the variance of the interpolated data is strictly less than the variance of the 12 monthly means (applied as constants over a month). It is tempting to wonder whether ad hoc fixes to wind stress data (e.g., multiplication by a constant factor to obtain a "better" simulation) might be related to this issue, although nominally undertaken to recover errors in the averaging process.

Thus, for wind stresses at least, interpolation between mean monthly data leads to errors that can be

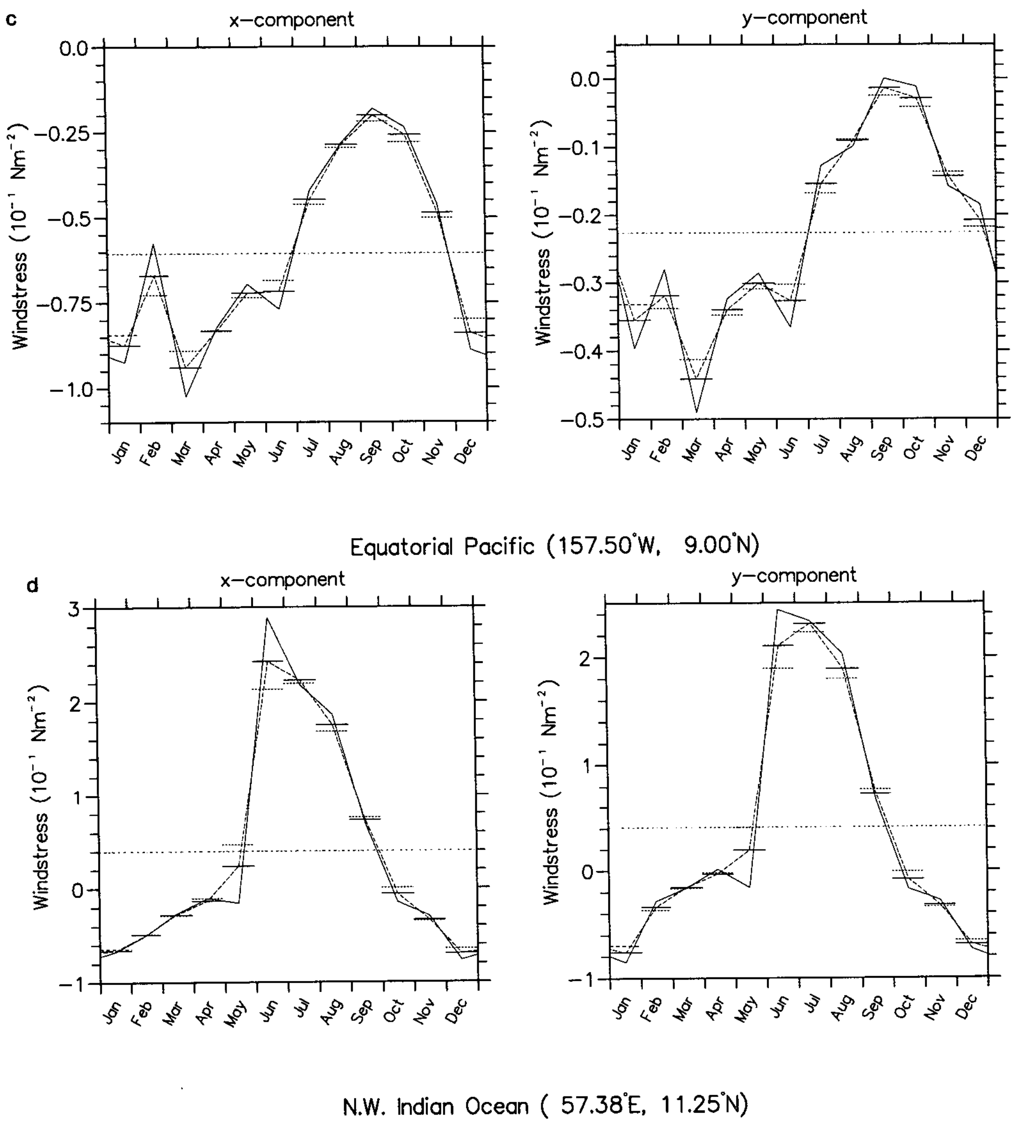


FIG. 5. (*Continued*)

large over a basin, persist for at least a month, and are potentially damaging to a simulation.

4. A simple solution

The above discussion showed that any interpolation scheme operating on the monthly mean data fails to produce a forcing whose mean is that observed. The solution is straightforward: all that is necessary is to replace the data by new pseudodata $\mathbf{d}'$, such that when this is interpolated, the monthly means are computed correctly. Thus, we require

$$\mathbf{A}\mathbf{d}' = \mathbf{d} \quad (4)$$

or, equivalently,

$$\mathbf{d}' = \mathbf{A}^{-1}\mathbf{d}. \quad (5)$$

The inverse matrix $\mathbf{A}^{-1}$ need only be calculated once, and applied to all relevant datasets. The computer code, and storage, used for running the model thus remains unchanged; the only difference is that the data have been changed to the pseudodata.

Figure 5 also shows examples of this modified data and the resulting interpolation and demonstrates clearly that the extrema in the interpolants must be rather stronger than those of the monthly means themselves. For equal length months it is simple to show that the inverse matrix for linear interpolation is diagonally dominant, with diagonal entry $2^{1/2} = 1.41$, with other entries falling off approximately as powers of $(8^{1/2} - 3 = -0.17)$ times this. The 1.41 factor must then lead to stronger extrema than originally. For Hermite interpolation, the diagonal entry is 1.23, with the first

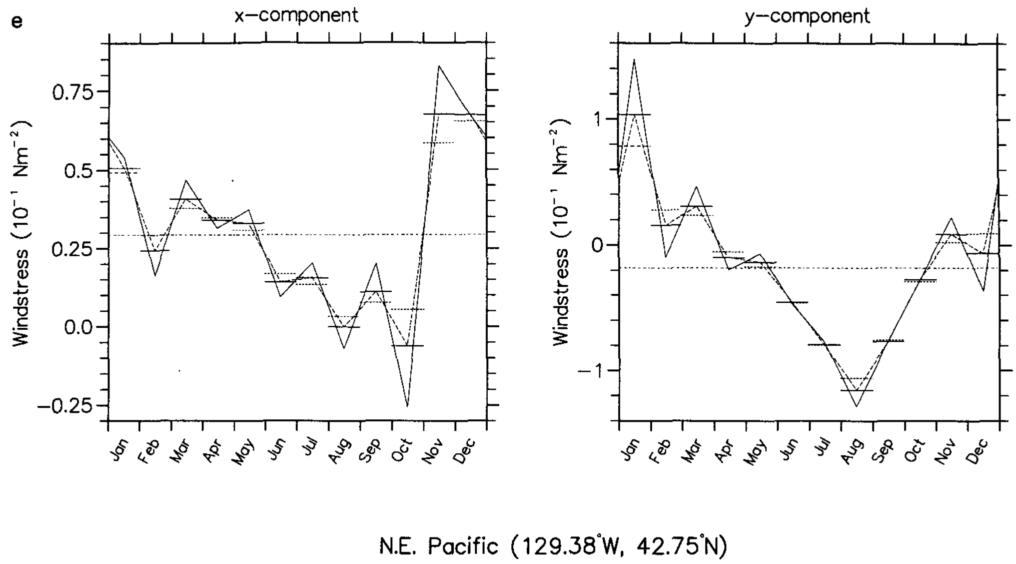


FIG. 5. (*Continued*)

off-diagonal entry being −0.14, so that the same result holds for extrema.

This would be expected anyway, since the original signal *must* have values lying both above and below the mean data. Thus, (even ignoring data errors themselves) the variance of the original signal must be larger than that obtained from the standard linear interpolation, and this solution has the virtue of at least reproducing this feature. Similarly, the variance of the modified interpolation is now strictly larger than the variance of the 12 monthly values applied as constants over the month; again, this must certainly apply to the original data. Whether ocean models react well to the (now more extreme) values they are forced with depends on the model.

5. Discussion

The example above was posed in terms of monthly wind stresses, but obviously applies to any linear interpolative scheme. The choice of action becomes more difficult when nonlinear quantities are involved. For example, consider a surface heat flux posed by

$$Q = Q_1 + Q_2(\mathrm{SST} - T) \equiv Q_2(T^* - T), \quad (6)$$

where SST is the observed sea surface temperature and Q_1, Q_2, and T^* are all functions of horizontal position and time and have been averaged into monthly bins. How best should the interpolation proceed? No simple linear interpolation of pseudodata can recapture the entire variability represented by (6). It may well be best to use the monthly mean values of Q_1 and Q_2 for the entire month (i.e., to undertake no interpolation) and to use pseudodata for T^*. The small jumps at month ends in the forcing will not induce awkward oscillations, unlike similar changes in wind stress, although they may cause difficulties for some mixed layer formulations.

Analysis of North Atlantic ECMWF SST data shows similar behavior to wind stress. We limit attention to the same pair of months. The January data are surprising. It might be assumed that the largest errors in SST interpolation would occur around the minimum surface temperature, which is usually in March or April, the month of maximal deep convection. Thus, January should be fairly innocuous. However, the errors are skewed toward positive values. While most errors are of order 0.1°, extreme values reach 2.1° and −1.5° in the Labrador Sea. In July, errors are skewed toward negative values. Again, typical values are of order 0.1°, but the range is of order ± 0.6°. The typical errors would yield erroneous heat fluxes of order 3–5 W m^{-2} for standard ocean model vertical resolution.

Acknowledgments. My thanks to Jeff Blundell for analyzing the data errors under severe time pressure, to Alex Megann for deriving the Hermite interpolation error, and to Sally Barnard for examining SST errors. Bernard Barnier provided the ECMWF monthly mean data. This paper is a contribution to the U.K. OCCAM Community Research Project, and to the MAST DYNAMO project, under Contract MAS2-CT93-0060. Referee comments were helpful.

REFERENCES

Bleck, R., and L. T. Smith, 1990: A wind-driven isopycnic coordinate model of the north and equatorial Atlantic Ocean. Part 1: Model development and supporting experiments. *J. Geophys. Res.*, **95**, 3273–3285.

Challenor, P. G., and D. J. T. Carter, 1994: On the accuracy of monthly means. *J. Atmos. Oceanic Technol.,* **11,** 1425–1430.

ECMWF, 1993: The description of the ECMWF/WCRP level III-A atmospheric data archive, Technical Attachment, ECMWF Shinfield Park, Reading, UK, 49 pp.

Epstein, E. S., 1991: On obtaining daily climatological values from monthly means. *J. Climate,* **4,** 365–368.

Ezer, T., and G. L. Mellor, 1992: A numerical study of the variability and separation of the Gulf Stream, induced by surface atmospheric and lateral boundary flows. *J. Phys. Oceanogr.,* **22,** 660–682.

The FRAM Group (D. J. Webb, P. D. Killworth, D. Beckles, B. A. de Cuevas, R. Offiler and M. Rowe) 1991: An eddy-resolving model of the Southern Ocean. *Eos, Trans., Amer. Geophys. Union,* **72,** 169–175.

Gerdes, R., and C. Wübber, 1991: Seasonal variability of the North Atlantic Ocean—A model intercomparison. *J. Phys. Oceanogr.,* **21,** 1300–1322.

Latif, M., 1987: Tropical ocean circulation experiments. *J. Phys. Oceanogr.,* **17,** 246–263.

New, A. L., G. Nurser, and R. Bleck, 1993: The effect of the Gulf Stream in an isopycnic-coordinate model simulation of the Atlantic Ocean. (Abstract). *Ann. Geophys.,* **11** (Suppl. II), p. C140.

Semtner, A. J., and R. M. Chervin, 1988: A simulation of the global ocean circulation with resolved eddies. *J. Geophys. Res.,* **93,** 15 502–15 522 and plates, 15 767–15 775.

Weller, R. A., and P. K. Taylor, 1993: Surface conditions and air–sea fluxes. CCCO-JSC Ocean Observing System Development Panel, Texas A&M University, College Station, TX, 131 pp.