How Predictable are the Geostrophic Currents in the Recirculation Zone of the North Atlantic?

JAMES A. CARTON

Center for Ocean–Land–Atmosphere Interactions, Department of Meteorology, University of Maryland, College Park, MD 20742

ABSTRACT

POLYMODE data is used to produce a new gridded analysis of multilevel streamfunctions in a $(500 \text{ km})^2$ domain during July 1977–August 1978. The analysis is used to define initial and lateral boundary conditions, and verification fields for six predictability experiments. These experiments are designed to determine the accuracy of current techniques of forecasting and hindcasting the circulation in a limited domain and to define the sources of error.

A forecast based on persistence is shown to be reasonably accurate only for ten days. A quasi-geostrophic model with persistence boundary conditions can maintain the same accuracy in the inner $(125 \text{ km})^2$ to 23 days. If the boundary information is updated, the root-mean-square error can be maintained below 60% throughout the $(500 \text{ km})^2$ region for at least 120 days. The initial state of the circulation only influences the first 30 days of a hindcast.

The instabilities in the circulation and inaccuracies in the specification of the boundary conditions place a lower bound on the error of streamfunction predictions of 35%. Thus a doubling of the accuracy of hindcasts and even greater improvements for forecasts appear to be possible if the accuracy of the model and boundary data can be improved.

1. Introduction

While little is known about the predictability of the synoptic/mesoscale ocean, many studies have addressed the predictability of the atmosphere. We know that small errors in General Circulation Model height fields double every 2–3 days. This error growth is due both to limitations in the ability of models to imitate the atmosphere, and to instabilities inherent in the circulation (see Shukla, 1985, for a recent review). The presence of instabilities implies that there are limits to the potential accuracy of any prediction scheme, and this has led to studies to define these limits.

A prediction loses all validity when the error, for example the difference between the prediction and observations, is as large as the difference between randomly chosen observations. A more severe upper bound on forecast error growth is defined by the accuracy of the current best models. We can expect future improvements in model design and data handling to improve the predictions, and thus lower this bound.

To isolate the fraction of the model error due to instabilities of the fluid, Charney et al. (1966) and several others compared the growth of differences between pairs of model solutions, or pairs of observed circulation patterns (Lorenz, 1969) with similar initial states. Since the model induced errors are similar for each solution, the growth of differences between the model solutions is assumed to be relatively model independent.

In a variation of this approach, Lorenz (1982) showed that a succession of operational forecasts form a series of forecast pairs. By comparing forecasts separated by one, two, or more days, he determined the system sensitivity to a variety of initial error sizes.

Limited area models introduce the additional question of the sensitivity of forecasts to the lateral boundary conditions. Anthes et al. (1985) found that the control exerted by the boundaries rapidly propagates toward the center of the region so that proper specification of boundary conditions is crucial to accurate predictions. The control exerted by the boundary conditions dominates the physical processes controlling the circulation.

A strong case can be made for using similar limited area models surrounded by open (water–water) boundaries in the midlatitude ocean. In situ measurements are time consuming, difficult, and expensive to make, so synoptic surveys must be fairly limited. Also the scales of interest are much smaller than the size of an ocean basin (50 km vs 5000 km).

Robinson and Leslie (1985) present preliminary ocean prediction experiments using some of the POLYMODE data from the recirculation zone of the northwest Atlantic to verify limited area forecasts, predictions which use no information about the future, and hindcasts, predictions which have continuously updated boundary conditions. They introduce the issue of the growth of error one may expect in a limited area ocean model, and the importance of the boundaries in controlling the interior flow versus internal dynamics.

This study presents a comprehensive objective analysis of the POLYMODE data, the most extensive dataset available today for midlatitude forecasting and verification. The analysis is used to define initial con-

ditions, boundary conditions and to provide verification for six prediction experiments. I estimate the growth rate of the streamfunction error at 100 m depth which may be currently expected in a $(500 \text{ km})^2$ region of the recirculation zone of the western North Atlantic. I explore the dependence of the error on boundary updating, and determine a lower bound on the error expected in predictions as the circulation models improve. An emphasis in this article is on the statistical reliability of the results. For this reason, the errors estimated in each experiment represent the average of 34 separate model integrations.

2. Data analysis

POLYMODE was an intensive effort by the United States and the Soviet Union to sample the mesoscale variability from July 1977 to July 1978 in a region centered on 29°N, 70°W. The measurements of ocean circulation were acquired using a variety of different instrument types including extensive hydrographic surveys, fixed current moorings, subsurface floats, as well as satellite altimetry. McWilliams et al. (1983) surveys the results from the Local Dynamics Experiment (LDE) subprogram. This was a high resolution study in a limited area. The XBT temperature maps from the large scale (500 km × 700 km) Synoptic Dynamics Experiment (SDE) are presented by Harrison and Heinmiller (1983).

Data from 19 Soviet current meter moorings are presented by Polloni et al. (1981) and Carter (1983). The float tracks are described in Rossby et al. (1983), which includes references to data reports and analysis papers. The Soviet hydrography is described in Nelepo (1983), and the POLYMODE Atlas (Kamenkovich et al., 1986), which reviews most of the other data collected as well.

The datasets used in the present analysis (essentially those listed above), are given in Table 1. To use this data for initialization and verification of models, it is converted to uniformly gridded streamfunction. The methods used are briefly described here, and in more detail in Carton and McGillicudy (1985).

Initially the daily averaged current data (from meters and floats) was collected at nominal depths of 700 and 1400 m. All the Soviet current meter velocities were reduced by an empirical factor of 1.75 to correct for the effects of wave pumping of the surface moorings,

TABLE 1. POLYMODE observations used in the gridded data analysis.

Dataset	Number of measurements
XBT	4141
Soviet Hydrography	1488
LDE CTDs	554
Soviet current moorings	19 (700, 1400 m)
LDE current moorings	7 (600–700 m)
SOFAR floats	22 (590–900 m)
	29 (1264–1501 m)

and to bring those measurements into agreement with the float data, following Polloni et al. (1981). Variations in the depth of the SOFAR floats is treated as an additional source of error. Because clustering of data poses a problem for the objective analysis, the data was sorted into $(10\text{-km})^2$, 10-day bins, retaining only the first measurement in each bin.
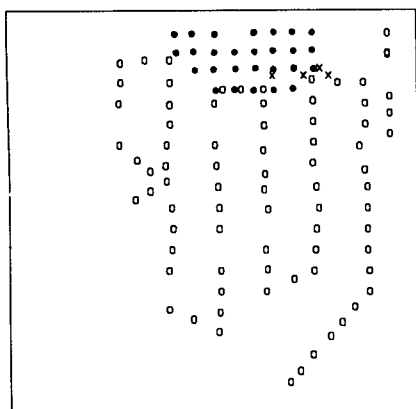
The prediction model described later requires uniformly gridded fields of streamfunction for initial conditions. To produce these from the nonuniform observations the objective analysis procedure of Bretherton et al. (1976) is used. The technique assumes that the data correlations are homogeneous and the instrument errors are uncorrelated. First, the streamfunction fields at 700 and 1400 m are constructed using the velocity measurements.

The correlations between observations of velocity and streamfunction are assumed to be of the Gaussian form suggested by Freeland and Gould (1976), although this was modified to allow the peak of the correlation function to propagate westward at 5.5 km day^{-1} at 700 m and 2.5 km day^{-1} at 1400 m (these speeds were estimated from the current meter data by Carter, 1983). The objective analysis scheme linearly estimates the streamfunction at each gridpoint by adding all the velocity measurements within ±10 days of the analysis date, weighted by their cross correlation with the streamfunction and their cross correlations with the other measurements. Calculation of the cross correlations is made simpler by assuming the flow is geostrophic and nondivergent. The result is a series of maps of streamfunction every 10 days beginning on Julian Day 3340 (15 July 1977) and extending through 3750 (29 August 1978), in a $(500 \text{ km})^2$ domain centered at 29°N, 70°W, with a grid resolution of 15.635 km. The data coverage for day 3700, Fig. 1, is better than average.
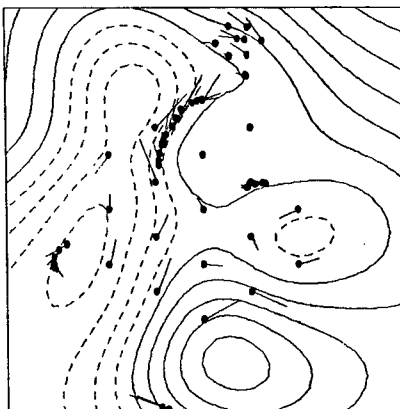
Like the velocity data, the hydrographic data, collected by XBTs, CTDs and several Soviet instruments, is first binned into $(10\text{-km})^2$, 10-day bins. In addition, casts more than 300 km outside our domain and those with anomalous data values are discarded. These procedures result in the elimination of 2276 casts to bring the number of casts used down to 3916 (the totals shown in Table 1 are before binning but after elimination of casts outside our domain). The mean temperature–salinity relationship for the earlier MODE experiment is used to estimate density from the XBT data, while density is estimated from the CTD data directly. Then all the density profiles are integrated upward from 700 m to produce dynamic height at 400 and 100 m. Typical coverage is shown in Fig. 1. The density data below 700 m has been ignored since too little of it was available to provide a complete set of synoptic maps. The true dynamic height at the measurement location is assumed to be the observed dynamic height plus an uncorrelated random error with an amplitude of 10% of the dynamic height variance.
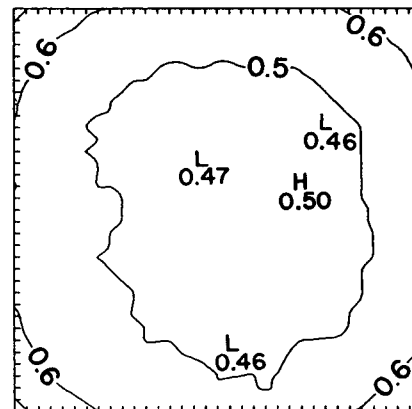
HYDROGRAPHIC STATIONS

700 m VELOCITY

AVERAGE ERROR

CI = 0.02   CI = 0.1

FIG. 1. Hydrographic station locations for days 3700–3710 are shown in the left panel. The symbols ●, O, × mark LDE CTDs, XBTs, and Soviet hydrography. Streamfunction at 700 m, day 3700, with the velocity vectors within ±5 days of the analysis date included, is shown in the central panel. Data comes from Soviet and American current meters and floats. The 100-m normalized error variance map is shown at right. This is an average of the error variance maps for the streamfunction analysis shown in Fig. 3.

The auto correlation of 100 m dynamic height estimated directly from the data (after removing the spatial mean) is fit to an analytical form (Fig. 2)

100 m DYNAMIC HEIGHT CORRELATION

MIN = -0.13 MAX = 0.6
CI = 0.1

MIN = -0.34 MAX = 0.69
CI = 0.1

CI = 0.1

CI = 0.1
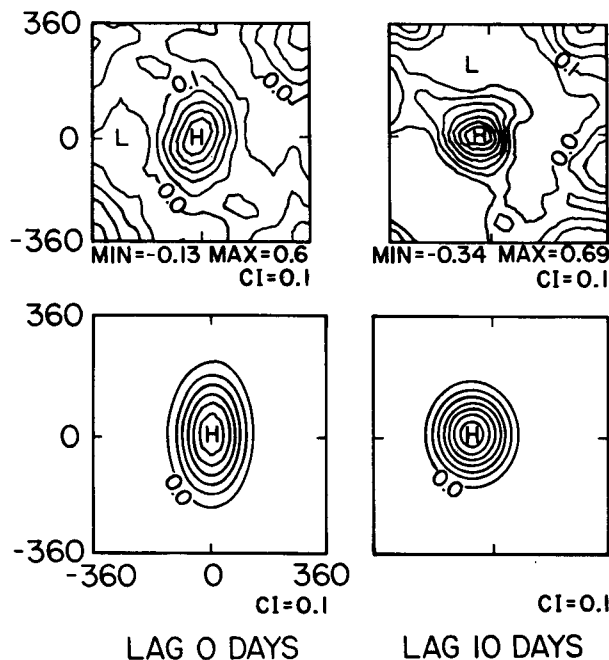
LAG 0 DAYS   LAG 10 DAYS

FIG. 2. Correlation of 100 m dynamic height as a function of horizontal lag for temporal lags of 0 and 10 days, computed directly from the raw data, above, and fit to an analytical function, below.

$$\gamma\left(1 - \frac{\Delta x^2}{r_1^2} - \frac{\Delta y^2}{r_2^2}\right) \exp\left(-\frac{\Delta x^2}{r_1^2} - \frac{\Delta y^2}{r_2^2} - \frac{\Delta t^2}{\beta^2}\right),$$

where Δx, Δy are the principal coordinates of the correlation function. The analytical correlation at 100 m and a similar function at 400 m is then used to objectively estimate the dynamic height on the same uniform grid as the velocity data.

When its zonal, meridional, time average is removed, the dynamic height analysis represents the vertical change in streamfunction (dynamic pressure) in the upper ocean. So adding it to the streamfunction analysis at 700 m gives the complete streamfunction in the upper ocean. The maps of 100 m streamfunction, shown in Fig. 2, have been scaled by $fU_0L/g = (7 \times 10^{-5} \times 0.1 \times 5 \times 10^4)/9.8 = 1/28$ m. The mean zonal and meridional velocity components for the 40 maps are −0.27 and −0.13 cm s⁻¹. The rms velocity components are 10.1 and 11.2 cm s⁻¹. For comparison, Owens et al. (1982) find current variability of 11.4 and 13.2 cm s⁻¹ in the 445 day record of the 250 m depth current meter from the LDE central mooring.

These fields contain evidence of many dynamical features discussed in the POLYMODE literature and provide a helpful way of relating the individual datasets. For example, the barotropic wave observed by Price and Rossby (1982) in the 700 and 1400 m float data is associated with the reversal of the 100 m flow in the north, from southwestward on 3660 to northeastward by 3690.

However small, intense features which are apparent in the LDE data and some float tracks (float 52, for
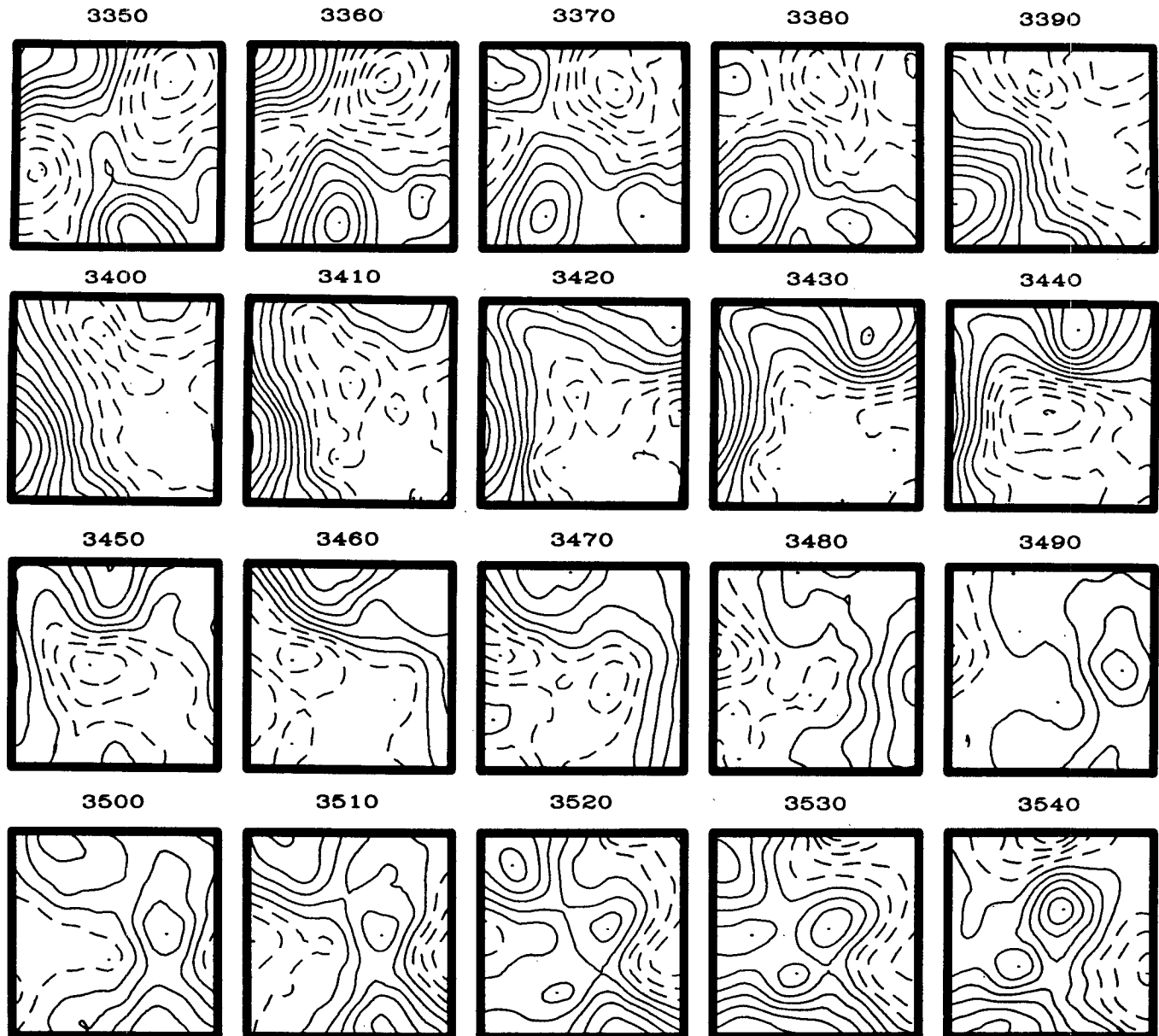
FIG. 3. Streamfunction at 100 m in a $(500 \text{ km})^2$ domain centered at 29°N, 70°W. The streamfunction has been normalized by $g/(fU_0L) = 28 \text{ m}^{-1}$. The contour interval is 1.0. Dashed contours are negative.

example, McWilliams et al., 1983), have been eliminated in this analysis either because of coarse sampling or smoothing implicit in the objective analysis. In addition, the analysis fields shown in Fig. 3 contain significant errors due to uneven sampling. The objective analysis procedure which created the dynamic height and 700 and 1400 m streamfunction analyses produces estimates of mean square error due to these effects.

Assuming that the errors in the different analyses are uncorrelated, the mean square error for 100 m streamfunction is the sum of the dynamic height and 700 m streamfunction errors. The average error for the 100 m maps (Fig. 1, right-hand panel) shows somewhat lower values toward the center reflecting the fact that measurements are more heavily concentrated there. The size of the average error makes it evident that very large datasets are required for any reasonable synoptic mapping.

An independent check on the accuracy of the analyses is provided by De Mey and Menard (personal communication, 1986) who have produced an analysis of the surface pressure field during 3660–3740 using the GEOS-3 and Seasat satellite altimetry in the same domain as this analysis. Their fields compare remark-

3550   3560   3570   3580   3590

3600   3610   3620   3630   3640

3650   3660   3670   3680   3690
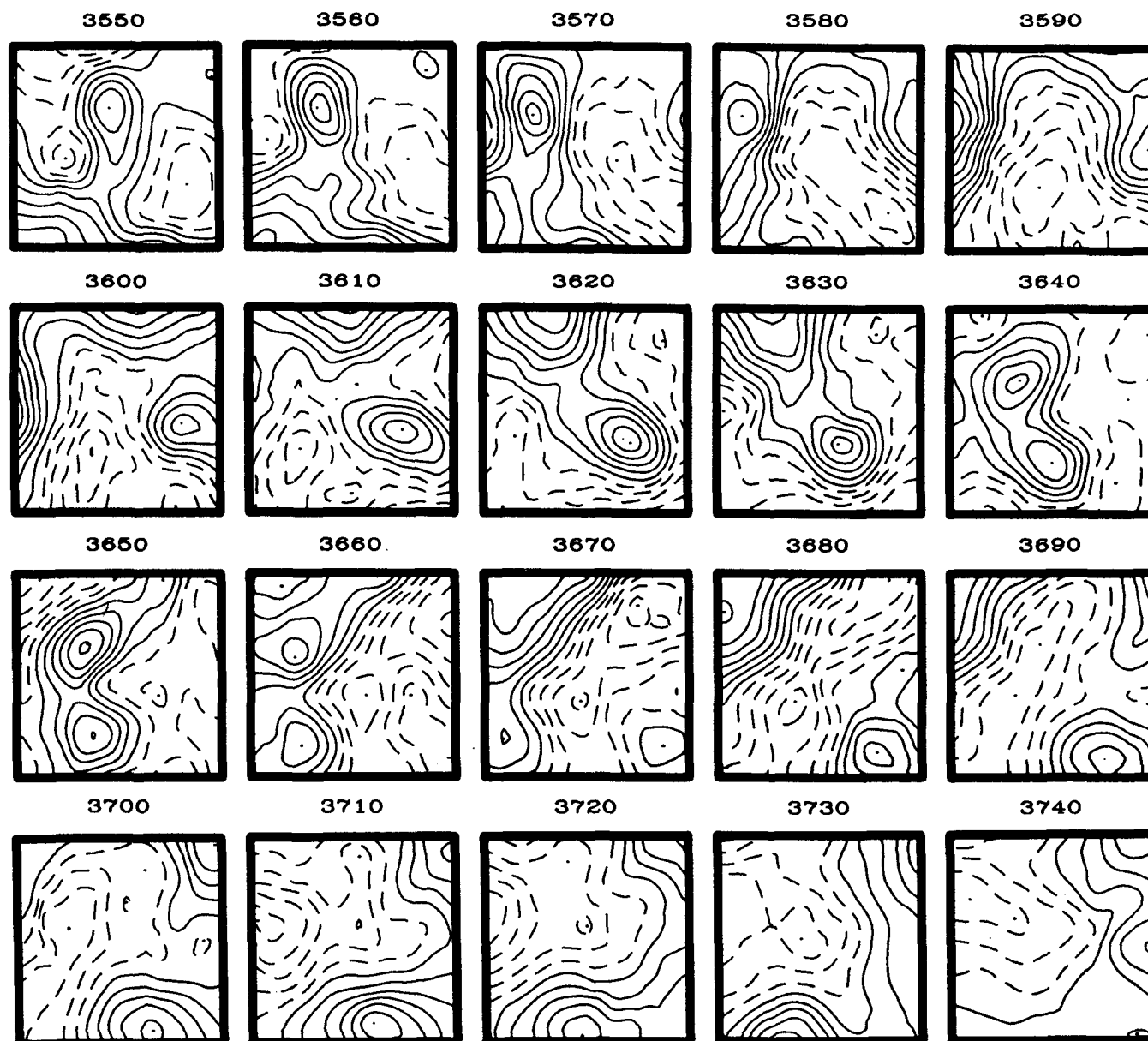
3700   3710   3720   3730   3740

FIG. 3. (Continued)

ably well with the analysis in Fig. 3 despite considerable error known to be in the altimeter data. Correlations between the gridded analyses during this period range from 0.47 to 0.74.

Below 1400 m, little direct information is available about the variability. Two methods are used here for determining the streamfunction in the deep ocean, information which is needed to initialize the model. Both methods are based on the observation by Richman et al. (1977) that motions of density surfaces in the MODE data are strongly vertically coupled and can be described by a smooth function in the vertical. To de-

termine such a function which is appropriate during POLYMODE, the seven current meters of the LDE central mooring were decomposed into empirical orthogonal eigenfunctions (EOFs), during the first 180 days when seven meters were functioning.

The analysis is performed on the individual velocity components separately, and the speed. No allowance is made for the possibility that the velocity components are coupled. In the upper ocean, the speeds are similar for the three first EOFs, but in the deep ocean, the meridional velocity EOF is significantly weaker. In each case the first EOF describes much of

the variance of the data, but is a poor estimator of the deep current.

The three first EOFs of each analysis are subjectively averaged giving amplitudes of (1.0, 0.9, 0.85, 0.7, 0.5, 0.2, 0.2) at depths (269, 394, 516, 616, 839, 2008, 5250 m). These amplitudes are linearly interpolated to the model level depths. For each analysis date, the strength of the averaged EOF is estimated by least squares fit to the streamfunction in the upper ocean at each horizontal grid point, and then used to estimate the streamfunction at 2400 and 3950 m. Generally this means that the deep currents are similar to the surface currents but their amplitudes are reduced by a factor of five.

In two experiments, a second method of creating the deep ocean current is tried in which the deep streamfunction is assumed to be proportional to the streamfunction at 1400 m, the deepest level at which data is available, with proportionality factors of $\frac{1}{3}$ and $\frac{1}{6}$ at 2400 and 3950 m.

3. Prediction experiments

We need an objective measure of prediction accuracy in order to compare different hindcasting and forecasting strategies. Two general types are used in the meteorological literature—correlations of predicted fields with the gridded analysis data, and root-mean-square (rms) differences between the two. The former is most sensitive to whether the model reproduces the pattern of observed highs and lows, while the latter is sensitive to whether the model reproduces the correct magnitudes. Here we chose to use the horizontal average of the rms difference of the 100 m streamfunction as our measure.

The rms difference between prognoses of length j-days and k-days for the streamfunction field at the same day is defined as

$$E_{jk} = \left[\frac{1}{S} \frac{1}{N} \sum_{i=1}^{N} \int_S [(\psi_{ij} - \bar{\psi}_{ij}) - (\psi_{ik} - \bar{\psi}_{ik})]^2 ds \right]^{1/2} \quad (1)$$

where the average is taken over $N = 34$ difference fields if model prognoses are compared with data, and over $N = 38$ if data prognoses are intercompared. The overbar represents a horizontal spatial average over the (500 km)^2 domain and usually $S = (500 \text{ km})^2$. All rms error estimates are normalized by the square root of the nondimensional variance of the 400-day 100 m streamfunction data, 2.318 (corresponding to 0.0828 m).

Many different prediction strategies are available. The simplest strategy is to assume that the ocean is persistent. The accuracy of this persistence forecast can be evaluated by comparing successive maps of the gridded data. Figure 4 shows the average persistence forecast error, E_{j0}, as a function of prediction time j. The error grows rapidly, reaching a value of 0.6 when the time lag is only 10 days. When persistence forecasts
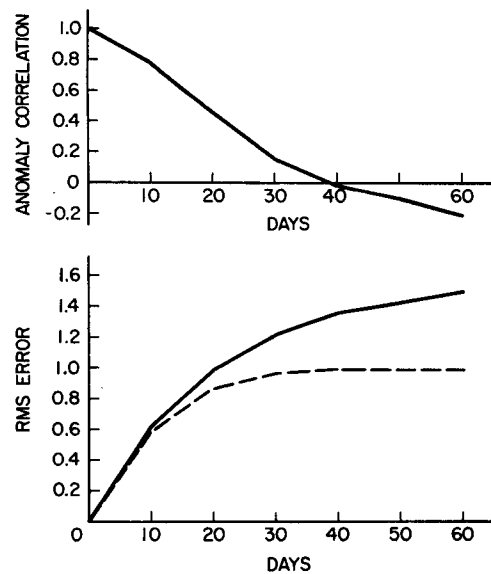

FIG. 4. Average 100-m streamfunction error statistics for persistence forecasts (solid curves) and damped persistence forecasts (dashed curve), with a damping time of 30 days. (a) The anomaly correlation, and (b) the rms difference. The rms difference is normalized by the square root of the nondimensional variance of the 100 m streamfunction data presented in Fig. 1, which is 2.318. The anomaly correlations of the persistence and damped persistence forecasts are the same.

reach 40 days the error begins to level off as the prediction and observation fields become uncorrelated. From Eq. (1) we can estimate the limiting expected error under these circumstances. If the forecast and observation variances are σ_1^2 and σ_2^2 then the difference field has variance

$$E^2 = (\sigma_1^2 + \sigma_2^2). \quad (2)$$

For the persistence forecast the normalized limiting error, $E_{\infty 0}$ is $2^{1/2} \approx 1.41$. The persistence forecast error exceeds 1.41 from 40 to 60 days because the data is actually negatively correlated in that interval.

Persistence alone is quite a bad prediction strategy. A simple way of improving forecast accuracy is to multiply the persistence forecast by a damping factor which depends on the duration of the forecast, $G(t)$, which has $G(0) = 1$ and which approaches zero for large forecast times. The limiting error in (1) is then 1.0. The improvement in error obtained for $G(t) = \exp(-t/30$ days) is shown in Fig. 4. The anomaly correlation is independent of G since only the amplitudes of forecast eddies have been affected, not their positions. To further improve our prediction strategy we turn to a model of ocean dynamics.

Numerical Ocean Prediction. The model chosen here solves the quasigeostrophic equations—no attempt is made to evaluate ageostrophic terms since these are expected to be generally small. The model uses a fourth order finite element scheme for interior horizontal de-

rivatives and second order finite differencing in the vertical. The horizontal resolution is 15.625 km, with six levels in the vertical at (100, 400, 700, 1400, 2400, 3950 m) at which the streamfunction is specified. A time step of two hours is used, with a fourth-order horizontal Shapiro filter applied every time step to prevent numerical instability caused by the cascade of vorticity to small scales. The filtering causes the model variance to decrease by 20% in the course of a 60-day integration.

The lateral open boundary conditions are the widely used conditions of Charney et al. (1950) which require the streamfunction to be specified along the boundaries at all times, and vorticity to be specified at inflow gridpoints. In addition to initial conditions, the model requires the mean oceanic stratification to be specified at seven levels (the surface, 200, 600, 800, 2000, 2800 and 5120 m). This is obtained by averaging the mean POLYMODE density profile between the streamfunction levels. The first baroclinic radius of deformation scale (43 km) and first internal model eigenfunction shape computed by the six-level model compare well with those computed from the continuous stratification. The model details are provided in Miller et al., (1983).

Observed bottom topography is included in all experiments. The major feature in the topography is the base of the Blake Plateau which extends 200 km into the domain from the west. The domain average depth is 5306 m, while the maximum deviation from the mean is 600 m. The 100 m streamfunction analyses

show no evidence of topographic steering in the first or second order statistics. Wind forcing has not been included in this study because global wind analyses could not provide accurate wind curl on scales smaller than our basin scale (500 km) and because of the mismatch between the scales of the wind forcing and ocean circulation. Thus the domain is being driven entirely by the lateral boundaries.

In Fig. 5 an example of a prediction of 100 m streamfunction is presented along with the difference between the model and gridded analysis fields. For this example the boundary conditions have been specified by the observed data so the difference field at the boundaries is zero. The error grows in the interior with time until most of the interior model field is depressed. This is indicated by the development of a negative streamfunction difference field with time (Fig. 5). The model hindcast does show an anticyclone (a local maximum in streamfunction) in the southeast which is a striking feature of the observed circulation on 3620 (Fig. 3). The 60-day error is 0.520, somewhat less than the average hindcast error discussed later. The growth of error differs widely in different predictions, and in fact may begin decreasing with time.

Six prediction experiments are described here, Table 2. Each experiment consists of 34 60-day integrations, cf., Fig. 5, beginning at 10-day intervals from 3350 through 3680. For each integration, streamfunction fields are sampled every 10 days. Thus the basic model dataset for the $33 \times 33 \times 6$ point model consists of
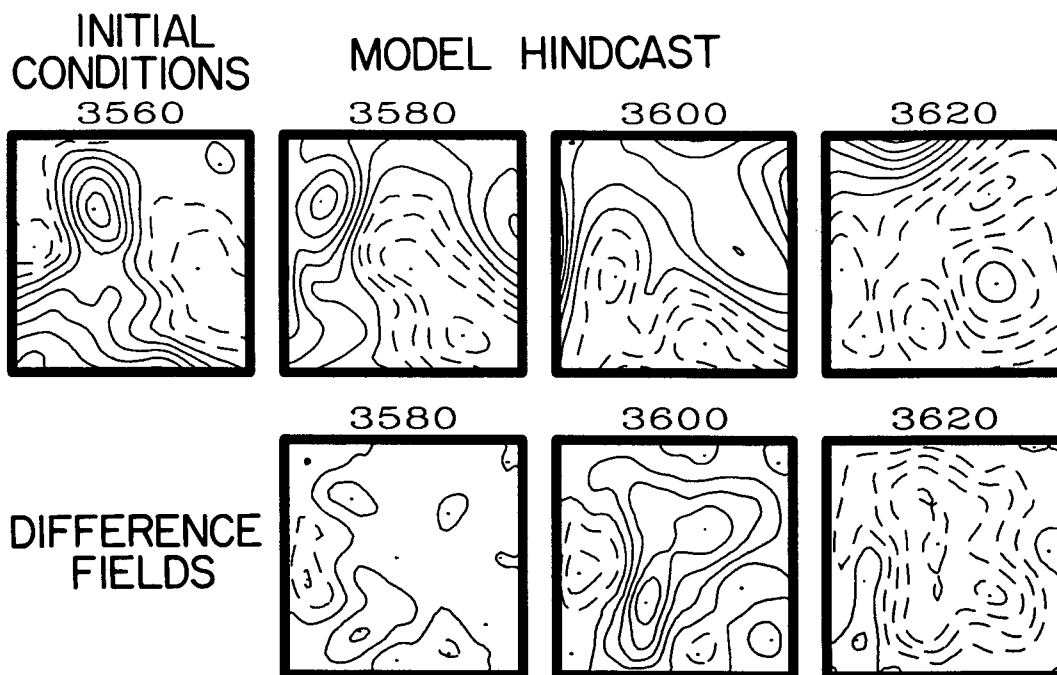

FIG. 5. A 60-day model hindcast beginning Julian Day 3560 (from Exp. 3) shown every 20 days, and the difference between the model hindcast and the gridded analysis. Exp. 3 consists of 34 such hindcasts. The contour interval is 1.0, where the streamfunction is normalized as in Fig. 1.

TABLE 2. Numerical experiments.

Exp.	Number of runs	Length (days)	Initial conditions	Boundary specification	Comments
1	34	60	Data	Persistence	
2	34	60	Data	Damped Persistence	
3	34	60 (5 to 120)	Data	Data	
4	34	60	Data	Data	second extension to deep levels
5	34	60	Data	Data	linear
6	34	60	zero	Data	second extension to deep levels

$6534 \times 34 \times 6 = 1\ 332\ 936$ numbers. From this archived data, a second file has been created containing the 100 m streamfunction predictions at 10-day intervals for each experiment. The error statistics from the individual integrations are averaged to provide statistical reliability, and are compared in Fig. 6.

When persistent boundary conditions are used (Exp. 1), the initial error growth is nearly as large as the persistence forecast. However, the error statistics are a strong function of the size of the domain in which the error is computed. In the central $(125\ km)^2$ domain the error continues to grow to 40 days and reaches a lower error of 0.75 (Fig. 7). The forecast error is unlikely to be less than the intrinsic error in the analysis fields. The low error in the inner domain at 10 and 20 days implies that the formal error, shown in Fig. 1, of the analysis fields is an overestimate. The strong domain dependence of the error suggests that errors are propagating into the interior from boundaries, although some effect may be due to the larger error in the analysis fields near the boundaries.
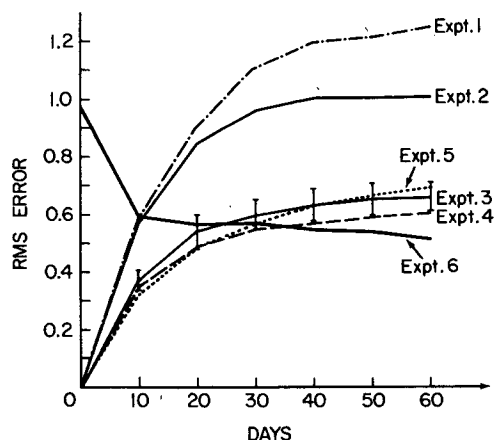

FIG. 6. Average 100 m rms error statistics for the experiments listed in Table 2. The rms error has been normalized by 2.318 as in Fig. 2.

Figure 8 shows the spatial distribution of the rms difference between the forecasts and gridded data at 10-day intervals. As the forecasts progress, boundary induced errors propagate from gridpoints near the eastern boundary in toward the center of the domain at roughly 4 km day^{-1}. In the west errors are larger and tend to collect within a few gridpoints of the boundary. The forecast error is improved significantly by using damped persistence for the boundaries (Exp. 2); the average error in the $(500\ km)^2$ domain reaches a limiting error of 1.0 by 40 days (Fig. 6).

Since much of the error in Exps. 1 and 2 is due to inaccurate boundary conditions, the experiment is rerun with the boundary conditions specified directly by the gridded data (Exp. 5). Integrations of this type are called hindcasts since updated information is continually being applied to the boundaries. While this reduces the boundary error, there is still some error introduced because the boundary conditions themselves are not perfect.

The average error, shown in Fig. 6, grows rapidly initially although still slower than the persistence boundary condition forecast. Because the error levels off by 40 days, more than half of the 60-day error is already present at 10 days. Improvements to the 10-day hindcast, either by improving the quality of the initial conditions, or improving the model, will have a large impact on model accuracy during the interval when the prediction error is controlled by the initial conditions. The error curve levels out after 50 days, well below the limiting error of 1.41. The first five hindcasts were extended beyond 60 days to verify the long term error. The average error of these at 120 days is 0.545.

Calculations shown in the Appendix suggest that an error with long wavelengths in the 2400 m streamfunction will contaminate predictions of 100 m streamfunction to some extent in 30 days. To test the dependence of the streamfunction predictions on the deep flow, the set of hindcasts in Exp. 3 are repeated with the second method of determining the deep velocity, which gives reduced magnitudes, as in Exp. 4.
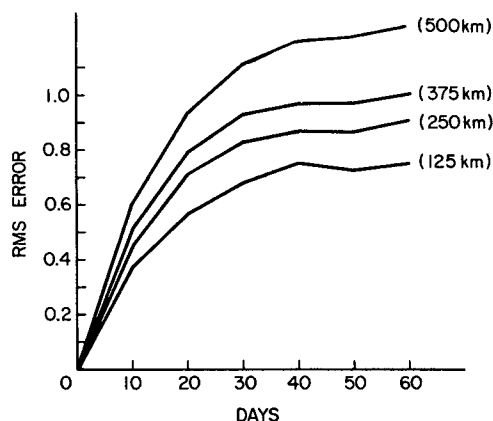
FIG. 7. Average 100 m rms error statistics for the persistence boundary Exp. 1 as a function of area of averaging. The error grows much less rapidly near the center of the domain than near the boundaries.

The error statistics are improved, however, the improvement is not quite statistically significant. A possible conclusion from this improvement is that when nothing is known about the deep velocity field, it is
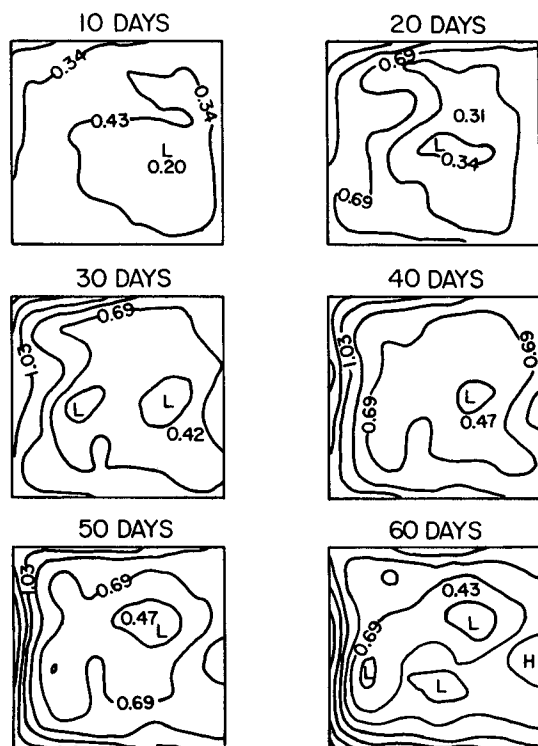

FIG. 8. The square root of the average square of the difference between the persistence boundary forecast (Exp. 1) and the gridded data for each point in the grid, plotted at 10-day intervals. The boundary induced error spreads rapidly into the interior, mainly from the east while large gradients of error are maintained along the western boundary. The contour interval is 0.173.
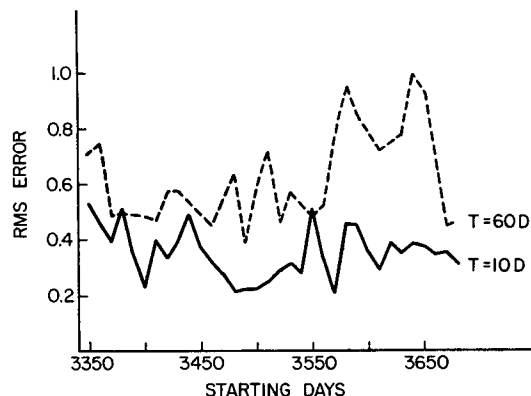

FIG. 9. The rms error for 10-day and 60-day prognoses for Exp. 3 plotted as a function of the starting date of the prognosis.

better to assume a weak flow rather than creating a deep motion field with erroneous velocities of the correct magnitude.

The error statistics for individual hindcasts vary substantially with time. To define the variability of the average error curves. (Fig. 6), 95% confidence limits have been included on the Exp. 3 error (the confidence limits for the other curves are similar). To compute these, the error estimates for individual hindcasts are assumed to be independent, giving 33 degrees of freedom. This assumption is based on the lack of correlation between successive predictions evident in Fig. 9, except in the interval 3570–3660, a period in which the kinetic energy of the gridded 100 m fields is highest. The variability of the error estimated from successive predictions underlines the need for many forecasts and verification analyses to draw meaningful conclusions about model error. It is also evident that prediction error does not increase uniformly with prognosis time, and in fact may decrease with prognosis time in individual cases.

When primitive equation atmospheric models are used for extended forecasting they develop a substantial systematic error independent of initial conditions. One method of correcting for systematic error is to remove the case-average difference between the model and the gridded data, from each model run before computing (1). This connection was attempted for Exp. 3, and led to a substantial reduction in the average 60-day error of 24% to 0.501. The systematic error in this model may be due to an intrinsic characteristic of the model, to errors in the initial data, or in the way the initial data interacts with the boundary conditions.

The low error which can be maintained by the model implies that instabilities may not dominate the changes in circulation, since such instabilities should be sensitive to small errors in the initial conditions. The typical Rossby number of the flow, U_o/fL, choosing the rms velocity as the measure of speed, and the internal deformation radius as a length scale, is $0.1/(7 \times 10^{-5}$

$\times 5 \times 10^4) \sim 0.03$. These facts suggest that, operationally, the advective terms in the vorticity equation may not be critical to the model dynamics in this limited domain. Experiment 5 shows (Fig. 6) that removing these terms does increase the error, but only by 6% over 60 days.

One way to isolate the fraction of the error growth due to instabilities, which I adopt here, is to compare integrations with perturbed initial conditions (identical twin experiments). The success of this method depends on the near cancelation of errors introduced by the model into both runs. The difficulty with this approach is to pick pairs of initial conditions which are consistent with the same boundary conditions. But in fact, as Lorenz (1982) points out, we already have a series of such hindcast pairs in the experiments presented above.

Call ψ_n^i the streamfunction for the ith day for a prognosis beginning on day n. Two successive forecasts, ψ_n^i and ψ_{n+10}^{i-10}, $i = 10$ to 60 days have the same boundary conditions, since the second prediction begins 10 days after the first but is 10 days shorter. However, the interior fields of the two predictions are different even when the second prediction begins ($i = 10$ days). The beginning of the second prediction, ψ_{n+10}^0 is just the observation field, at day $n + 10$, so the rms error for $i = 10$ (averaged over $n = 1, \cdots, 34$) is the model error at 10 days, which was previously displayed as Exp. 4 in Fig. 6, and which is the upper heavy curve in Fig. 10. The average rms difference between hindcast pairs beginning 10 days apart, ψ_n^i and ψ_{n+10}^{i-10}, is the lowest thin curve in Fig. 10, plotted as a function of i. The upper thin curves marked $k = 20, \cdots, 50$ are the rms differences between hindcasts starting k days apart. Comparison of these curves shows the sensitivity of the difference growth to the strength of the initial difference.

In a global atmospheric model, the presence of dynamic instabilities causes initial differences between the predictions to continue to amplify with time. In

the present oceanic case, as i increases, the rms difference for $k = 10$ days approaches a limit of 0.35. This limit is due both to a balance between the growth of instabilities and the control exerted by the boundary conditions, and to the boundary induced error that results from different interior circulations at $i = 10$ days. Even when the initial difference is larger ($k = 20, \cdots, 50$) the difference between hindcasts seems to approach this limit. The decrease in 60-day error from 0.61 in Exp. 4 to 0.35 is because we are now comparing hindcasts with hindcasts instead of hindcasts with data. The implication is that the limiting hindcast error in this $(500 \text{ km})^2$ domain can be reduced by half to 0.35 by improving the model interior physics and the quality of the analysis fields, so that the model and data are more consistent. The only hope for holding the error below 0.35 is to improve the specifications of the boundary conditions themselves.

When accurate initial conditions are combined with inaccurate boundary data the inaccurate boundary information eventually overwhelms the initial data, leading to large prediction errors. In Exp. 6, the initial conditions are taken to be zero (no motion) but the gridded analysis is used to define the boundary conditions. The initial error rapidly decreases as the boundary information propagates into the interior. By 30 to 40 days the zero initial condition hindcasts are as good as those begun with accurate initial conditions. This is the timescale during which the initial conditions influence the solution. For longer times the area average circulation is controlled by the boundaries.

Picture-by-picture comparisons of Exps. 4 and 6 show that not only do they both differ from the analysis by the same amount after 30 days, but they differ in the same ways, in other words, they are very similar to each other. This is further evidence that the boundary conditions, combined with model physics, force the model to develop certain kinds of errors independent of the initial conditions. The errors are therefore not the result of an instability of the flow.

By 60 days the rms error for hindcasts beginning with zero initial conditions (Exp. 6) is less than the hindcast errors obtained when using the analysis for initial conditions (Exp. 4) after 30 days. By 60 days it is 15% lower. This surprising result may indicate that the initial conditions contain variability which is not able to leave the domain smoothly, and so eventually contributes to the error field. An attempt was made to find a scale dependence to the eddies left behind in the domain erroneously. The individual hindcast and analysis maps were Fourier decomposed and the error in various wavenumber bands were compared. There did not appear to be a significant scale dependence for the error in either Exp. 4 or 6.

4. Discussion

Limited area models are appealing in the midlatitude ocean because of the small scale of variability observed
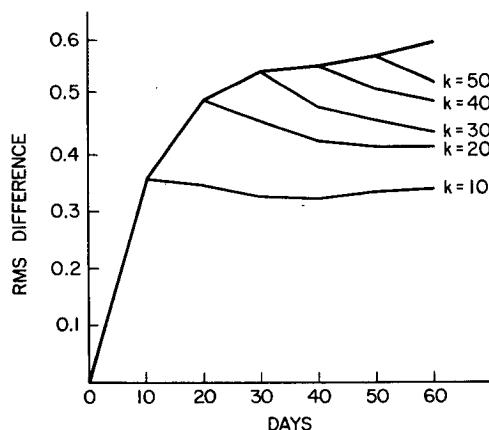
FIG. 10. Average rms differences between successive pairs of model hindcasts $(\psi_n^i - \psi_{n+k}^{i-k})$ from Exp. 4, plotted as a function of i. The bold curve connects points where $i = k$.

(50 km), and the expense and difficulty of acquiring sufficient in situ data to define the initial state of the ocean. To evaluate their usefulness for assimilation and prediction of the synoptic/mesoscale, a gridded dataset has been assembled using most of the POLYMODE data available for a $(500 \text{ km})^2$ region of the western North Atlantic. This data is used here to initialize and verify model integrations to determine the basic 100 m error statistics for a limited area model, and to define the sources of the error.

The rms error obtained when the ocean is assumed to be persistent exceeds 0.6 in 10 days. When the quasigeostrophic model is used to predict flow evolution given persistence boundary conditions, the $(500 \text{ km})^2$ domain averaged forecast error still reaches 0.6 in 10 days, but the error in the inner $(125 \text{ km})^2$ region doesn't reach 0.6 until 23 days. The boundary conditions themselves can be predicted using a statistical model, as suggested by Robinson and Leslie (1985), or by embedding the limited region in a large domain model.

Accurate boundary information allows the error in hindcasts to be maintained below 0.6 apparently indefinitely. Over half of this error develops in the first 10 days. Comparison between hindcasts with accurate initial conditions and those with zero circulation initial conditions shows that the initial information influences only the first 30 days of the hindcasts. Hence, to improve hindcasts in the 10 to 30 day range, improvements in initial conditions, model physics, and initialization procedures are necessary.

A lower bound on prediction error of 0.35 is determined by the growth of error in hindcasts with perturbed initial conditions. This error is due both to the balance between growing instabilities and the control exerted by the boundaries, and to errors caused by inaccurate boundary conditions. Thus, it should be possible to reduce the hindcast prediction error by half by increasing the model accuracy and the quality of the data specified along the boundaries. For further decreases in prediction error it will be necessary to improve the way the boundary conditions are specified.

Several limitations of this study should be mentioned. These results are domain dependent. Instabilities will play a greater role in regions of strong current such as the Gulf Stream. The analysis data is degraded by sparse sampling and the lack of information about deep ocean circulation. A 50% error at 2400 m may cause a 10% error in the 100 m streamfunction. The results presented here indicate that the domain size is an important parameter. Experiments in different physical and dynamical regimes need to be performed. Finally, physical processes such as direct wind forcing and mixed layer dynamics have been neglected.

Limited area prediction of the development and movement of mesoscale variability in the midlatitude oceans, on timescales of months, is possible. In the studies described here, only statistical information is used to create the analysis fields from data. Improved analyses require an interaction between the data, its statistics, and a dynamical model, such as is carried out in atmospheric numerical weather prediction under the name of four dimensional data assimilation. Exploitation of the potential of this strategy for oceanography will depend partly on understanding the sensitivity of model prognoses to initial and boundary error such as described in this study. This will become an increasingly critical area of research as attention is focussed on data assimilation.

APPENDIX

Deep-Ocean Errors

Measurements of the circulation in the deep ocean are difficult to obtain. One important issue is how errors in assumptions about the circulation in the deep ocean influence predictions in the circulation of the upper ocean. Suppose that an error is introduced into the fifth level (at 2400 m) of a six-level quasigeostrophic model which is linear, flat bottom, and in a horizontally unbounded domain, but otherwise like the numerical model used here. Assume that the error at time $t = 0$ is $\psi (z = 2400 \text{ m}) = \cos(kx)$, for some wavenumber k. The streamfunction error satisfies

$$\nabla^2 \psi_t + \left(\frac{f^2}{N^2(z)} \psi_z \right)_{zt} + \beta \psi_x = 0 \qquad \text{(A1a)}$$

$$\psi_z(z = 0) = \psi_z(z = H) = 0$$

$$\psi(t = 0) = \begin{cases} \cos(kx), & \text{if } z = 2400 \text{ m} \\ 0, & \text{otherwise} \end{cases} \qquad \text{(A1b)}$$

where $(\ )_z$ represents a discrete vertical derivative. Expand (A1) into dynamical modes that have the form $\phi_n(z) \cos(kx - \sigma_n t)$ satisfying

$$\left(\frac{f^2}{N^2(z)} \phi_{nz} \right)_z + \lambda_n^2 \phi_n = 0$$

$$\omega_{nz}(z = 0) = \phi_{nz}(z = H) = 0.$$

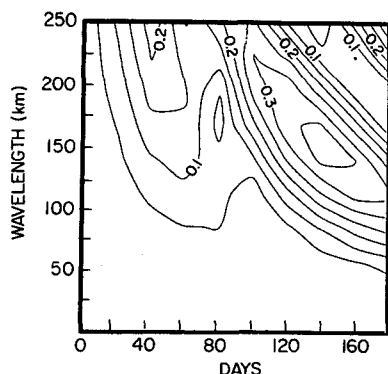Substituting $\psi = \sum_{n=1}^{6} a_n \phi_n(z) \cos(kx - \sigma_n t)$ into (A1a,

FIG. 11. The spatial average rms streamfunction error at 100 m given an initial error of the form cos(kx) at 2400 m, plotted as a function of wavelength $2\pi/k$, and time. Initially the error at 100 m is zero.

b) and using the orthogonality of ϕ_n gives

$$\sigma_n = -\beta k/(k^2 + \lambda_n^2)$$

$$a_n = \phi_n(z = 2400 \text{ m})\Delta Z_5/[\sum_{i=1}^{6} \phi_n^2(z_i)\Delta Z_i],$$

where ΔZ_i is the thickness of the layer surrounding the ith streamfunction level. The root square mean of the streamfunction error,

$$E = \left(\int_{-L}^{L} \frac{\psi^2(x, t)dx}{4L^2}\right)^{1/2}, \quad L \to \infty,$$

is plotted in Fig. 11 as a function of time and wavelength. For 60-day forecasts only the waves longer than 150 km can influence the 100 m streamfunction, and they can only cause errors of 20% of the initial 2400 m error. However somewhat larger errors are possible in longer forecasts.

REFERENCES

Anthes, R. A., Y.-H. Kuo, D. P. Baumhefner, R. M. Errico and T. W. Bettge, 1985: Predictability of mesoscale atmospheric motions. Advances in Geophysics, Vol. 28b, Academic Press, 159–202.

Bretherton, F. P., R. E. Davis and C. B. Fandry, 1976: A technique for objective analysis and design of experiments applied to MODE-73. Deep-Sea Res., 23, 559–582.

Carter, E. F., 1983: The statistics and dynamics of ocean eddies. Ph.D. dissertation, Harvard University.

Carton, J. A., and D. McGillicudy, 1985: Comprehensive objective maps of POLYMODE streamfunction. Reports of Meteorology and Oceanography, Harvard University, 26 pp.

Charney, J. G., R. Fjortoft and J. Von Neumann, 1950: Numerical integration of the barotropic vorticity equation. Tellus, 2, 237–254.

——, R. G. Fleagle, V. E. Lally, H. Riehl and D. Q. Wark, 1966: The feasibility of a global observation and analysis experiment. Bull. Amer. Meteor. Soc., 47, 200–220.

Freeland, H. F., and J. Gould, 1976: Objective analysis of mesoscale ocean circulation features. Deep-Sea Res., 23, 915–924.

Harrison, D. E., and R. H. Heinmiller, 1983: Upper ocean thermal variability in the Sargasso Sea July 1977–July 1978: The POLYMODE XBT program. J. Phys. Oceanogr., 13, 859–879.

Kamenkovich, V. M., A. S. Monin and A. D. Voorhis, Eds., 1986: The POLYMODE Atlas, Woods Hole Oceanographic Institution, 375 pp.

Lorenz, E. N., 1969: Atmospheric predictability as revealed by naturally occurring analogues. J. Atmos. Sci., 26, 636–646.

——, 1982: Atmospheric predictability experiments with a large numerical model. Tellus, 34, 505–513.

Miller, R. N., A. R. Robinson and D. B. Haidvogel, 1983: A baroclinic quasigeostrophic open ocean model. J. Comput. Phys., 50, 38–70.

McWilliams, J. C., and co-authors, 1983: The local dynamics of eddies in the western North Atlantic. Eddies in Marine Science, A. R. Robinson, Ed., Springer–Verlag, 92–113.

Nelepo, B. A., Ed., 1983: Experimental investigations under the international POLYMODE program. Amerind Publishing, New Delhi (translated from Russian), 143 pp.

Owens, W. B., J. R. Luyten and H. L. Bryden, 1982: Moored velocity measurements on the edge of the Gulf Stream recirculation region. J. Mar. Res., 40, 509–524.

Polloni, C., A. Mariano and H. T. Rossby, 1981: Streamfunction maps of the Soviet POLYMODE current meter array. Tech. Rept. 81-1, University of Rhode Island, 133 pp.

Price, J. F., and T. Rossby, 1982: Observations of a barotropic planetary wave in the western North Atlantic. J. Mar. Res., 40, 543–558.

Richman, J. G., C. Wunsch and N. G. Hogg, 1977: Space and time scales of mesoscale motion in the western North Atlantic. Rev. Geophys. Space Phys., 15, 385–420.

Robinson, A. R., and W. G. Leslie, 1985: Estimation and prediction of oceanic eddy fields. Progress in Oceanography, Vol. 14, Pergamon, 485–510.

Rossby, T. S., C. Riser and A. J. Mariano, 1983: The western North Atlantic—A Lagrangian viewpoint. Eddies in Marine Science, A. R. Robinson, Ed., Springer–Verlag, 66–91.

Shukla, J., 1985: Predictability. Advances in Geophysics, Vol. 28b, Academic Press, 87–121.