AD-A035 822 AIR FORCE GEOPHYSICS LAB HANSCOM AFB MASS F/G 4/2
A PROPOSED PROCEDURE FOR DIAGNOSIS AND IMPROVEMENT OF DYNAMICAL--ETC(U)
APR 76 C YANG
UNCLASSIFIED AFGL-TR-76-0079 NL

1 OF 1
ADA035822

END

DATE
FILMED
3-77

AFGL-TR-76-0079
ENVIRONMENTAL RESEARCH PAPERS, NO. 556

A Proposed Procedure for Diagnosis and Improvement of Dynamical Prediction Models

CHIEN-HSIUNG YANG

13 April 1976

METEOROLOGY DIVISION PROJECT 8604
AIR FORCE GEOPHYSICS LABORATORY
HANSCOM AFB, MASSACHUSETTS 01731

AIR FORCE SYSTEMS COMMAND, USAF

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: AFGL-TR-76-[illegible]79, AFGL-ERP-556
2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle): A PROPOSED PROCEDURE FOR DIAGNOSIS AND IMPROVEMENT OF DYNAMICAL PREDICTION MODELS,
5. TYPE OF REPORT & PERIOD COVERED: Scientific. Interim.
6. PERFORMING ORG. REPORT NUMBER: ERP [illegible]. 556

7. AUTHOR(s): Chien-Hsiung Yang
8. CONTRACT OR GRANT NUMBER(s): (12) 31 p.

9. PERFORMING ORGANIZATION NAME AND ADDRESS: Air Force Geophysics Laboratory (LY), Hanscom AFB, Massachusetts 01731
10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: (16) 86041002 (17) 10

11. CONTROLLING OFFICE NAME AND ADDRESS: Air Force Geophysics Laboratory (LY), Hanscom AFB, Massachusetts 01731
12. REPORT DATE: 13 Apr [illegible]76
13. NUMBER OF PAGES: 31

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): (9) Environmental research papers,
15. SECURITY CLASS. (of this report): Unclassified
15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

PE61102F

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Boundary layer forecast model
Air weather service
Model evaluation

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report lays a logical foundation on which deficiencies of dynamical prediction models can be defined and dealt with on a rational basis. Measures of the quality of performance of dynamical prediction models are suggested, and a procedure with which deficiencies of the models can be diagnosed is proposed. A sample study is used to illustrate the use of the proposed procedure, and the results of the study are presented.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

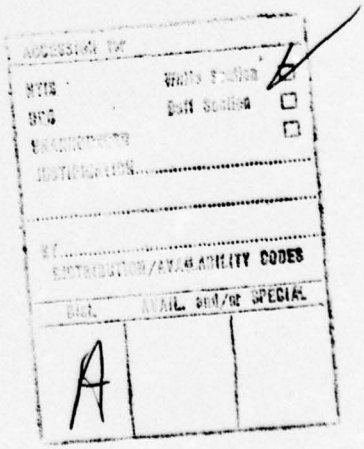
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

4/09578

Preface

The continuing efforts of Capt James E. Kester of the Air Force Global Weather Central (AFGWC) in supplying data, providing counsel and coordinating the work are gratefully appreciated.

Appreciation is also extended to Mrs. Gail Bertolini for expert programming, and Mr. Karekin Agazarian of the Air Force Geophysics Laboratory (AFGL) for his meticulous management of the computations.



Contents

Illustrations

Tables

A Proposed Procedure for Diagnosis and Improvement of Dynamical Prediction Models

1. INTRODUCTION

Since July 1974 a project has been in progress at the Air Force Geophysics Laboratory (AFGL) in cooperation with the Air Force Global Weather Central (AFGWC) to improve the accuracy of AFGWC's boundary layer forecasts within the current operational constraints by identifying and correcting deficiencies of the model. The model, called the AFGWC Boundary Layer Model (BLM), is one of the first to provide routine forecasts in the planetary boundary layer using a numerical model, and has been in operation for more than five years.

In a first report, as this is, we deem it not only important but also necessary to describe in detail some of the more important questions initially encountered and to present the results of a study carried out with the proposed procedure, even though they are based on a small sample. We have chosen to present basic considerations in general terms, since they are applicable to all models of a similar kind, and to focus on the AFGWC-BLM only later when the specifics of a model become relevant.

A dynamical prediction model is defined here to be a totality of logical operations having the following features:

(a) It models the meteorological evolution in a given region over a specific period of time,

(b) its structure is based on the physics of the atmosphere, and

(Received for publication 13 April 1976)

(c) its purpose is to predict future values of observable meteorological variables.

In this view, a dynamical prediction model is a rational attempt to foretell the outcomes of a complex of physical processes in the atmosphere by tracing its evolution. This is accomplished by predicting the future state of pertinent meteorological variables. In this sense, it may be distinguished from a similar endeavor, whose main aim is to investigate the physics of a complex of physical processes by numerical simulation. It is also different in its approach from another type of prediction model whose logic is based on statistical relationships between predictors and predictands.

The quality of performance of a prediction model will be characterized by two attributes of the model, namely accuracy and efficiency. Accuracy describes the closeness of a model to the object it models. In the case of a prediction model this attribute is most pertinently expressed by the forecast error, which is defined as the difference between the predicted and the real outcome of concern.

The second attribute, efficiency, on the other hand, describes the amount of effort expended by a model in producing a prediction. It can be measured in any of many conventional ways, such as the amount of computer time or a combination of the core storage and the machine time required. The importance of efficiency as a factor in the quality of performance arises from the practical consideration that a prediction model is required to deliver its product within the restriction of the operational environment, such as the capacity and capability of the computer in use and the amount of time and labor allocated.

By improving the quality of performance of a prediction model, therefore, we mean implementation of such means that makes the model predict more accurately and/or efficiently. It would be ideal if one means could improve both accuracy and efficiency simultaneously. However, it is as likely, if not more, that one is gained at the expense of the other. A question thus arises as to what combination of changes in the two attributes constitutes improvement in the quality of performance. We believe that such a question should be dealt with individually as it arises under a given circumstance. We are also convinced that it is possible, by taking all relevant factors into consideration, to decide under any circumstance what mix of accuracy and efficiency would be considered an improvement of the quality of performance of the given model.

The aspects in which improvement will be sought involve both the physical assumptions underlying the model and the computational methods employed. We assume that the basic logical framework of the model is sound. We expect as a result of this study to determine whether this latter assumption is justified and also to be able to specify the extent of improvement feasible within the framework of the BLM.

In the rest of this report we shall apply these basic concepts to a specific example of dynamical prediction model. In the context of this example we will introduce measures which express quality of performance, and will also suggest a number of different ways of describing the error or inaccuracy of the model. We will then propose a method of diagnosis which is useful in aiding us to find relationships between various parts of the internal structure of the model and the forecast error.

2. THE AFGWC BOUNDARY LAYER MODEL

A comprehensive description of the model is given in Hadeen[1] and Hadeen and Friend.[2] Only a very brief characterization will suffice for the purpose of this report.

The model is based on the physics of the atmosphere and formulated as an initial-boundary value problem of a spatially-confined body of air, which is assumed to obey a certain set of laws and satisfy a number of empirical rules. The laws as used here designate those relationships among variables that are believed to be universally valid. The empirical rules, on the other hand, are experimental or suppositional. While the laws are well founded in theory and confirmed by measurements, the empirical rules are mostly either supported only by a limited amount of empirical evidence or remain valid only on statistical ground. Introduction of the latter into the model is necessitated by the lack of better alternatives.

The set of mathematical equations expressing these laws and relationships and other pertinent information such as the initial and boundary conditions are solved on a time-space mesh by approximating derivatives and integrals by finite differences and sums. The prediction consists of the values of the wind velocity, temperature and moisture of the air on these grid points.

The logical framework of the model is schematically represented by Figure 1, taken from Hadeen.[1] It clearly shows that the model is an organized, logical structure consisting of a number of interlocking and interacting modules of different specific functions.

The model covers the earth's atmosphere from the ground up to the level 1.6 km above the ground in the vertical and over an area about the size of the North American continent in the horizontal. The horizontal mesh size is the same as the so-called limited-area fine mesh and is half that of the 1977-point octagonal grid of

1. Hadeen, K.D. (1970) AFGWC Boundary Layer Model, AFGWC Technical Memorandum 70-5, published by Air Force Global Weather Central, Air Weather Service (MAC), Offutt AFB, Nebraska 68113.

2. Hadeen, K.D., and Friend, A.L. (1972) The Air Force global weather central operational boundary layer model, Boundary Layer Meteorology, 3:98-112.

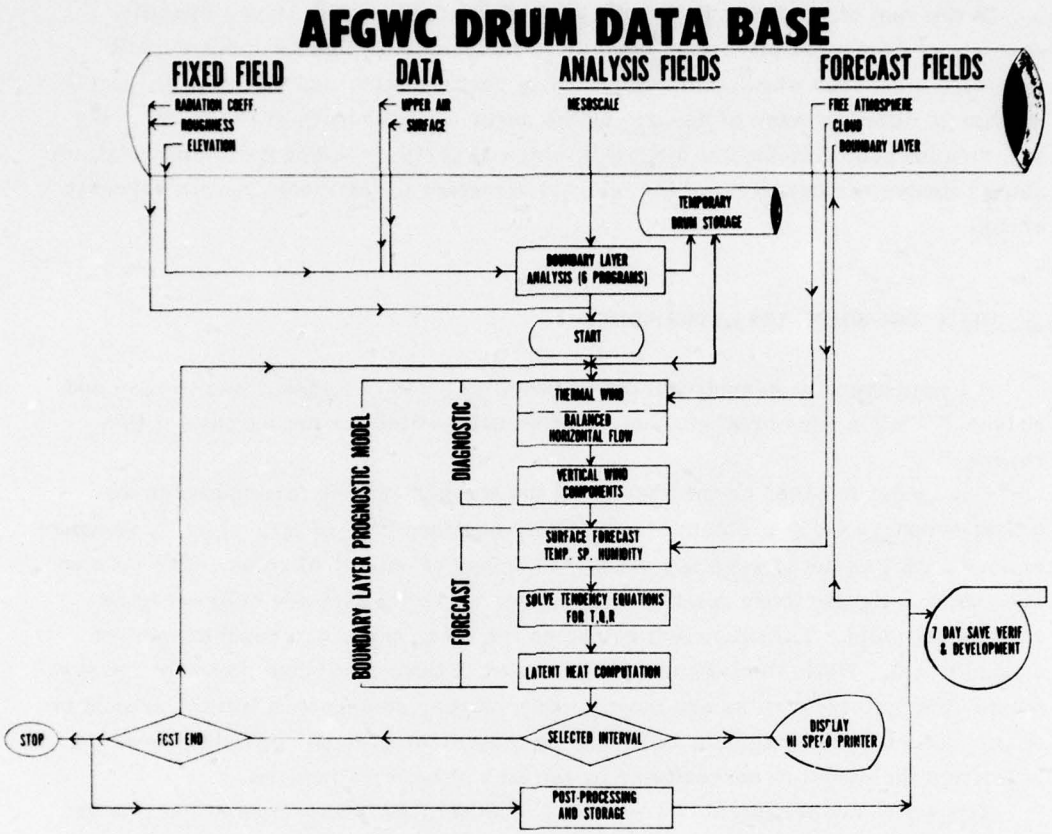


Figure 1. Flow Diagram of AFGWC-BLM (from Hadeen[1])

the National Meteorological Center. The entire depth of the boundary layer is represented by eight discrete levels consisting of the surface, 50-, 150-, 300-, 600-, 900-, 1200-, and 1600-m levels above the ground.

The model's initial state is determined through an objective analysis combining observed and previously predicted values. The boundary values for the wind velocity at the top of the boundary layer during the period of prediction are estimated from forecasts prepared by a separate free-atmosphere prediction model. The nature and amount of cloud cover present in the free atmosphere are obtained in a similar fashion.

Dynamical prediction models differ from the real atmosphere. The prescribed conditions and input information are simplified, inaccurate and incomplete; the physics incorporated within the model are also simplified and incomplete; the

numerical procedures employed in the model are at best gross approximations of processes operating within the atmosphere.

A difference in any area may lead to a difference between the prediction by the model and the real state of the atmosphere, which has been defined to be the forecast error of the model. However, a discrepancy between the model and real atmosphere in any of these areas would not be automatically a defect of the model, unless such a discrepancy led to a forecast error.

3. ROOT-MEAN-SQUARE FORECAST ERROR

Improvement of the accuracy of a model becomes synonymous with reduction of the forecast error of the model. Having so stated, however, we immediately find ourselves asking "What should we mean by the accuracy of a model and by the forecast error of a model?"

As defined in Section 1, the forecast error is the difference between the predicted and the real outcome of the predictand under consideration. Since the Boundary Layer Model (BLM) predicts the values of wind velocity, temperature and moisture of air at a specified set of points in space and time, there is no difficulty in recognizing the predicted state. The matter is, however, not so straightforward with the real state. The value of a parameter assigned to a particular point in time and space may be obtained from an observation or may be inferred from an analysis of a field of observations supplemented with predictions. Both are treated as "real data" but they generally differ from each other. Which is more real?

Such questions may be unimportant and even irrelevant in the case where one could gather as many samples as one desires in calculating forecast errors. They are, however, worth careful consideration when one is concerned with optimizing the benefit of experiments with a limited number of samples. In any event, it is obvious that what the real state should be depends to a large extent on the ultimate objectives of the prediction.

For the present purpose we shall assume that the AFGWC boundary layer prediction is prepared for a general purpose; that is, the model attempts to predict the future state of the physical and dynamical structure of the planetary boundary layer. In view of this assumption we have chosen to equate the "real state" to the objective analysis. We then define the forecast error of the model as the aggregate of the differences at the individual grid points between the predicted values and the values obtained through the objective analysis. This implies that we have chosen to ignore a possible imperfection in the method of objective analysis and abstain from studying its effects on the forecast error.

It is noted here that the above definition is at variance with the standard verification statistics employed by AFGWC, where the difference between the predicted and the observed values of a predictand at each of a number of designated weather stations is defined as the individual forecast error. The difference in choice arises from a difference in purposes.

In the present task where diagnosis and improvement of the model are our main concerns we need measures which are sensitive enough to reflect differences that are present or may be introduced into the internal structure of a model. It is therefore desirable to have measures which are amenable to subdivision or stratification. Thus, the higher density and greater homogeneity of the spatial distribution of the grid points in comparison with that of the weather stations makes the definition adopted here superior to that of the standard verification statistics for our purposes.

For the collective measure of the model forecast error of a predictand we will use a vector of K components, in which each component represents the root-mean-square of the forecast errors on the horizontal array of grid points at a level in the vertical. Each component has a magnitude given by

$$(\mathrm{RMSE}) = \sqrt{\frac{1}{NxM} \sum_{j=1}^{N} \sum_{i=1}^{M} e_{ij}^2} \tag{1}$$

where NxM is the number of the grid points in the horizontal array and e_{ij} stands for the forecast error at grid point (i, j).

It is readily seen that

$$(\mathrm{RMSE})^2 = E^2 + S^2 \tag{2}$$

where E is called the "bias" and defined by

$$E = \frac{1}{NxM} \sum_{j=1}^{N} \sum_{i=1}^{M} e_{ij} \tag{3}$$

and S is the standard deviation of the forecast error, given by

$$S^2 = \frac{1}{NxM} \sum_{j=1}^{N} \sum_{i=1}^{M} (e_{ij} - E)^2, \; S \geqq 0. \tag{4}$$

The geometrical meaning of Eq. (2) is quite obvious; the statistical counterpart is appreciated if it is recalled that in a normal distribution the mean and the standard deviation are the two statistically independent parameters that completely determine

the distribution. For these reasons, we shall use from time to time the ordered-pair representation (E, S) for each component of the root-mean-square forecast error.

4. DISTRIBUTIONS OF FORECAST ERROR

While such parameters as the root-mean-square forecast error or the bias of the forecast error are important and useful, it must be readily admitted that they are neither sufficient nor adequate to describe or characterize the aggregate of the forecast errors on a horizontal array. In an attempt to capture other important characteristics of the aggregate we will employ the following two representations of the distribution.

(1) The contour map of the forecast errors at a level. It depicts the geographical distribution of the forecast errors (Figures 2 and 3). It helps us in recognizing large features of the forecast errors that may originate either in the geographical characteristics or in the synoptic patterns. For example, a cursory examination of the contour maps on different synoptic situations has shown that forecast errors of large magnitude - both positive and negative - in all three predictands tend to cluster in mountainous regions. The same contour maps also show that there is a great deal of coherence in the vertical in the patterns of the geographical distribution.

(2) The partition of the error variance into spectral cells. Another interesting and potentially valuable way of analyzing the forecast error is to map the forecast error at a level into the wave-number domain. We first seek for the aggregate of forecast errors $\{ e_{ij},\ i=1,\ldots,\ M;\ j=1,\ldots,\ N \}$ at a single level the representation

$$e_{ij} = \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} E_{mn} \cos \frac{m\pi x}{L_x} \cos \frac{n\pi y}{L_y} \tag{5}$$

where L_x and L_y are, respectively, the dimensions in the x- and y-directions of the domain of prediction at the level. We thus obtain the transformation from $\{ e_{ij},\ i=1,\ldots,\ M;\ j=1,\ldots,\ N \}$ in the horizontal plane to $\{ E_{mn},\ m=0,\ldots,\ M-1;\ n=0,\ldots,\ N-1 \}$ in the wave-number domain. The variance or error energy in each wave component is then proportional to the square of its amplitude E_{mn}. It is obviously only reasonable and sensible to abstract and summarize this distribution of error energy among a great number of individual wave components into that among a far smaller number of cells by combining many wave components into one single cell.

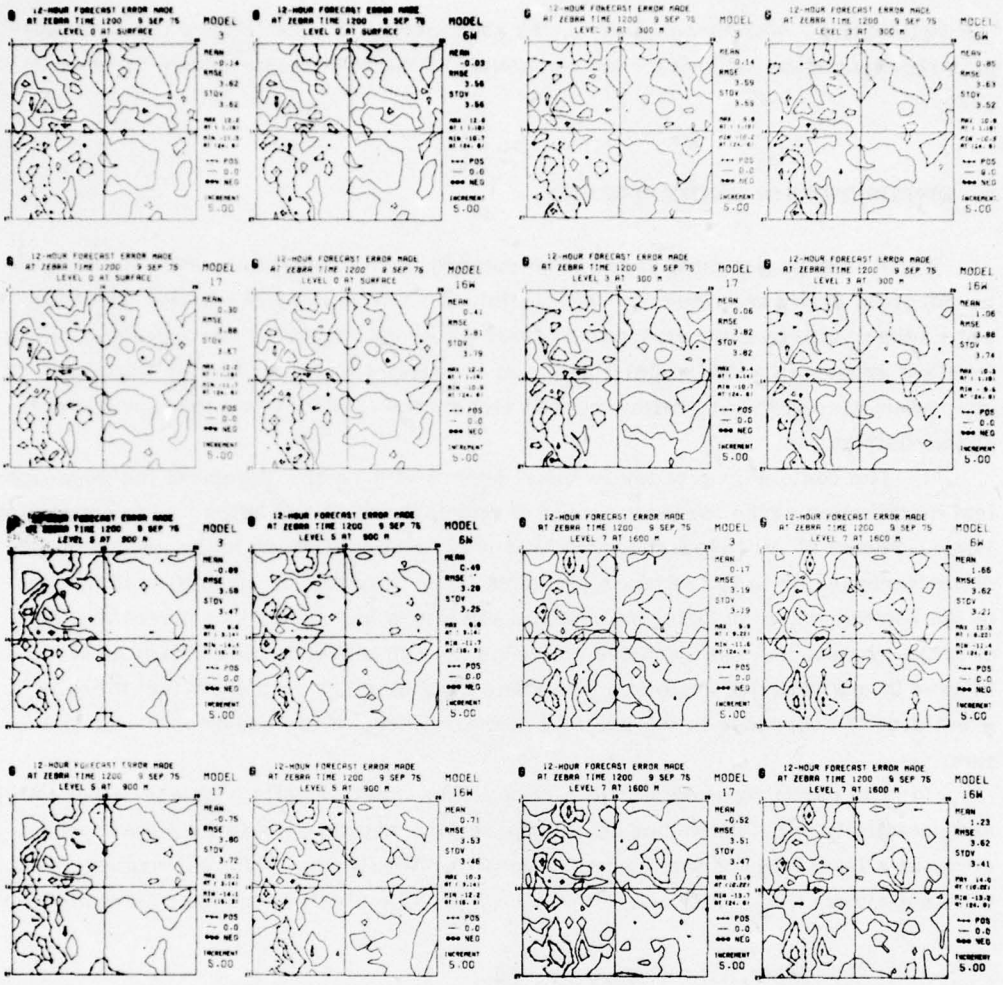


Figure 2. Contour Maps of Forecast Errors, 12-hr Temperature Forecasts

With little knowledge of how the error energy might be likely to be distributed among the wave components, we have tentatively tried the grouping of wave components into cells which is presented in Table 1. Here, U(m, n) represents the union of m and n; therefore, $a < U(m, n) \leqq b$ means that either m or n or both are greater than a and not greater than b. We hope more experience with actual data will eventually enable us to choose such groupings that would shed better light on how the error energy is distributed over the range of spatial scales. Figure 4 shows examples of the partition of error energy into the cells defined in Table 1.

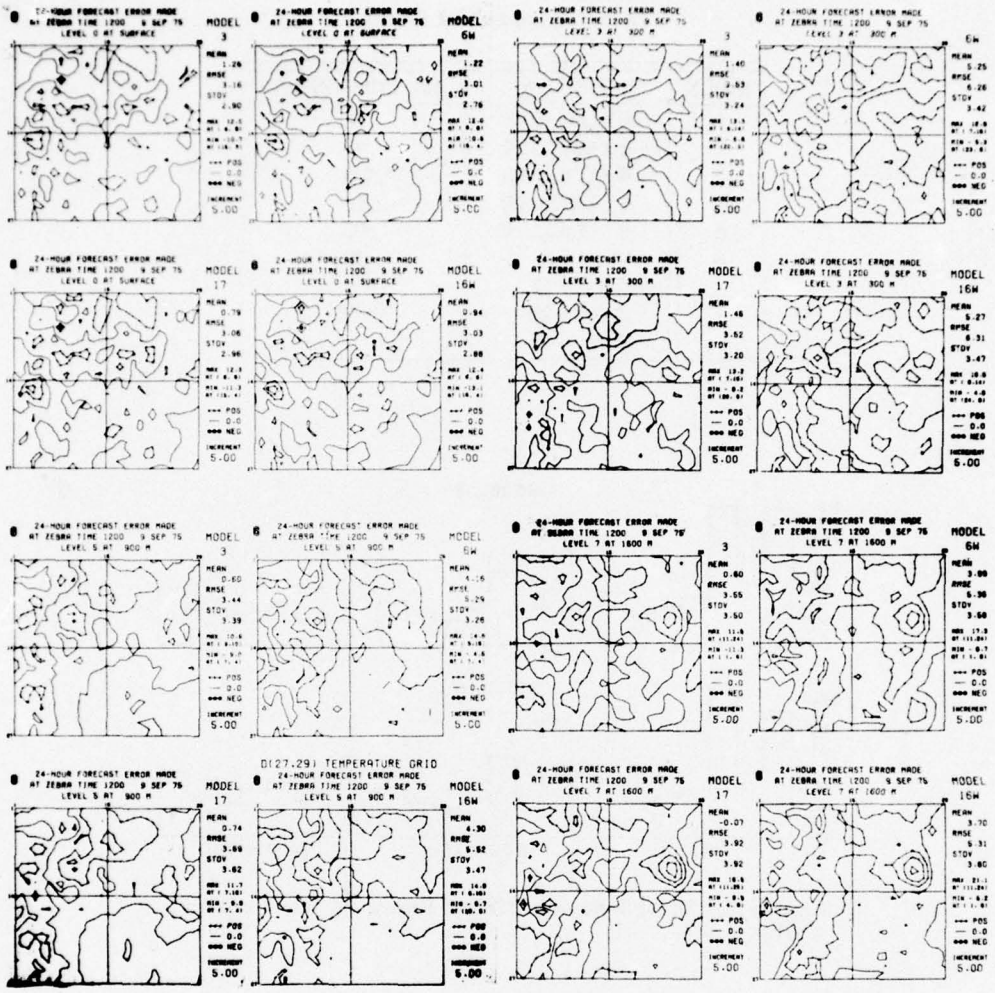


Figure 3. Contour Maps of Forecast Errors, 24-hr Temperature Forecasts

It is important to stress that the purpose of these different representations of the forecast error is to find systems or patterns in the error which may then be ascribed to a defect in a particular part of the internal structure of the model, so that there may be a reasonable expectation of removing or reducing that systematic error.

Table 1. Spectral Cells

INDEX	DEFINITION	NO. COMPONENTS
1	$0 \le U(m,n) \le 2$	9
2	$2 < U(m,n) \le 4$	16
3	$4 < U(m,n) \le 6$	24
4	$6 < U(m,n) \le 8$	32
5	$8 < U(m,n) \le 10$	40
6	$10 < m \le 14$ or $10 < n \le 13$	89
7	$14 < m \le 28$ or $13 < n \le 26$	573

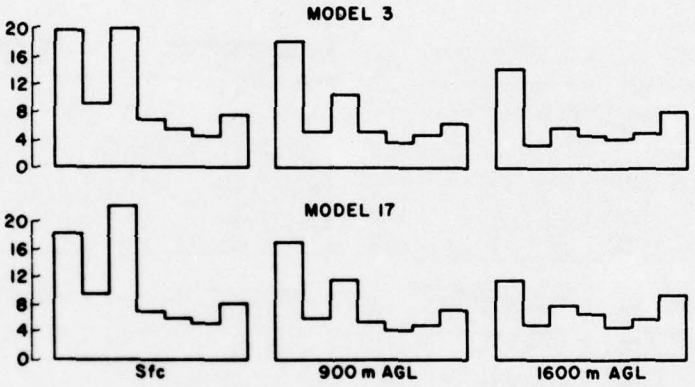


Figure 4. Distribution of Error Energies into Spectral Cells

5. TWIN EXPERIMENTS

The search for means of improvement of the quality of performance must start with establishing cause-and-effect relationships between any defect in the model and its effect as observed on the forecast error. We mentioned earlier in Section 2 that there are in any dynamical prediction model three different aspects in which causes for inaccurate prediction might exist. We also had a glimpse of the difficulty of the problem by noting that imperfection of a prediction model is understood only when it can be measured in terms of the forecast error.

There are two major obstacles in our quest for links between the internal structure of a dynamical prediction model and the forecast error of the model. The

first is the complexity of the models. Since the dynamics of the real atmosphere is highly nonlinear, the models that simulate the processes of the atmosphere tend to be very complicated, as has been illustrated in Figure 1. The interlocking and interacting of various modules make it difficult to isolate any single module and evaluate the effect of any change in such a module on the prediction.

The second obstacle is a lack of adequate observation or monitoring of the real atmosphere. The available measurements of the state of the atmosphere are inadequate both in the variety of variables measured and with regard to their density in space and time. Because of this inadequacy it is hardly possible, for example, to dissect and analyze individual aspects of the AFGWC Boundary Layer Model anywhere between the point of start and the point of stop in Figure 1.

We can thus neither single out the individual modules nor subdivide the process of evolution into substeps for the purpose of analyzing the model in finer detail. We are forced to accept the model as though it were an indivisible, integral unit. Under such a circumstance it appears that the method of twin experiments is the only viable technique of investigation. The method obtains and compares the predictions of two models that are identical in all aspects except the one which is selected for examination. If the difference in the aspect under examination between the two models does not radically change the total nature of the model, then the difference in the results of the two predictions may be interpreted as the response of the model as a whole to the introduced differences.

In order to determine how a prediction model responds to a change in the internal structure, the difference of the forecast errors between the two models is subjected to the same analyses as those described in Sections 3 and 4. The results of such analyses of the differences obtained under various concomitant conditions may reveal idiosyncrasies or orderliness in such responses of the model, and may also provide an estimate of the effects of such a difference in the internal structure on the forecast error.

The use of twin experiments as an assessor among a number of similar prediction models is rather straightforward. By running different models under identical conditions and comparing their forecast errors we can assess and evaluate the relative merits of the different models. It is important to note that while we may employ diagnostic experiments to study the nature of a model and to search for modifications which may improve the quality of performance of the model, we still shall have to rely on the assessing experiments to give a seal of approval to the model with a better performance.

We will consider a series of twin experiments carried out on the AFGWC Boundary Layer Model and illustrate how we may utilize it to achieve our objective.

For ease of experimentation and for clarity in interpretation of the experiments we have divided the entire structure of the boundary layer model as shown in

Figure 1 into two parts at the point of START. This allows us to separate the area of prognostic operation from the area of the prescribed conditions and input information. We can then restrict our attention to the former while recognizing the latter as a vital, but only an accompanying condition.

The series of twin experiments involves four models with slight differences among themselves and from the standard operational model, designated as Model 0. These four models are arranged in a two-by-two matrix in Figure 5 and may be characterized briefly as follows:

Model 3: Model 0 with the GWC packing.

Model 6: Model 0 with the Air Force Geophysics Laboratory (GL) packing.

Model 17: Modified from Model 0 and with the GWC packing.

Model 16: Modified from Model 0 and with the GL packing.

Here, the GWC and GL packings designate two slightly different but parallel procedures of storing the computed values of temperature at each time step. The modifications incorporated in Models 16 and 17are listed in Figure 5. They are: (1) a change in the size of time step from 30 min to 1 hr, and (2) a change in the assumption concerning the temperature at 2 km above ground.

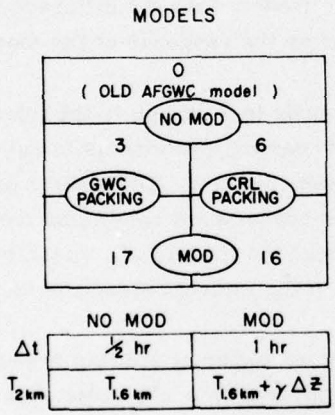


Figure 5. A Schematic Diagram of Combined Twin Experiments

Two issues are involved among the four models: the effect of the difference in packing of temperature values and the effect of the difference in the size of the time step on the forecast error. We are interested in finding out if there is any discernible effect from either of these differences, and, if so, the nature of the effects.

If neither of these differences violates the basic premise of twin experiments, that is, if neither changes the basic nature of the model drastically, we should expect that the difference of the forecast errors between Models 6 and 3 is statistically similar to that between Models 16 and 17. Similarly, we should expect that the

difference of the forecast errors between Models 17 and 3 is statistically similar to that between Models 16 and 16.

The four models were run on a number of sets of initial and boundary conditions which were obtained from actual synoptic cases at various times between April and October of 1975. Each model on each set of prescribed conditions produced one 12-hr and one 24-hr forecast whenever data were available for verification at the end of the 12- or 24-hr period. For each model we obtained 17 cases of the 12-hr forecasts and 12 cases of 24-hr forecasts. For the 12-hr forecasts, 9 belong to the day group and 8 to the night group.

The groups of figures in series Figures 6a through 6d present scatter diagrams of the biases (or average error) of the 12-hr temperature forecasts of the four models plotted against those of Model 0 on four levels. Here, level 0 stands for the surface, level 3 for the level at 300 m above ground, level 5 for the level at 900 m above ground, and level 7 for the level at 1600 m above ground. The two different 12-hr periods of a day - the one starting at 12Z covers most of the daytime, while the one starting at 0Z most of the night - are designated by circles and crosses, respectively. The quantities r_o and r_x refer to the correlation coefficients between the average errors in the respective day and night groups.

We may observe a fair degree of similarity in the pattern of scatter between Models 3 and 6 on one hand and between Models 17 and 16 on the other at level 0. At higher levels, however, the pattern of similarity changes, becoming more pronounced between Models 3 and 17 and between Models 6 and 16. It is also apparent that there is a discernible difference between the day- and night-groups.

Figures 7a through 7d illustrate the vertical profile of the differences between the various models with regard to quadrant for specific cases. Here, Q_i, i=1, 2, 3, 4 represents NW, NE, SW, and SE quadrants of the domain of prediction, respectively. These figures demonstrate first that there is a sufficient degree of uniformity among the four different quadrants to uphold the representativeness of the profiles of the difference on the two average errors over the entire domain. They also show, when Figure 7a is compared with Figures 7d and 7b with Figure 7c, that our expectation on the statistical similarity between the differences of the forecast errors mentioned earlier proves to be substantially true.

Table 2 presents a statistical summary of the differences D_1 and D_2 defined, respectively, by

$$D_1 = E_6 - E_3,$$
$$D_2 = E_{16} - E_{17},$$

where E_I represents the forecast error in Model 1. The two statistics, average, m, and standard deviation, s, over the horizontal array of grid points at each level are calculated for each difference for each case. These individual spatial statistics are grouped together into day- and night-groups. The mean and the standard deviation of each parameter in each group, M and S, are then calculated to furnish the entries of the table.

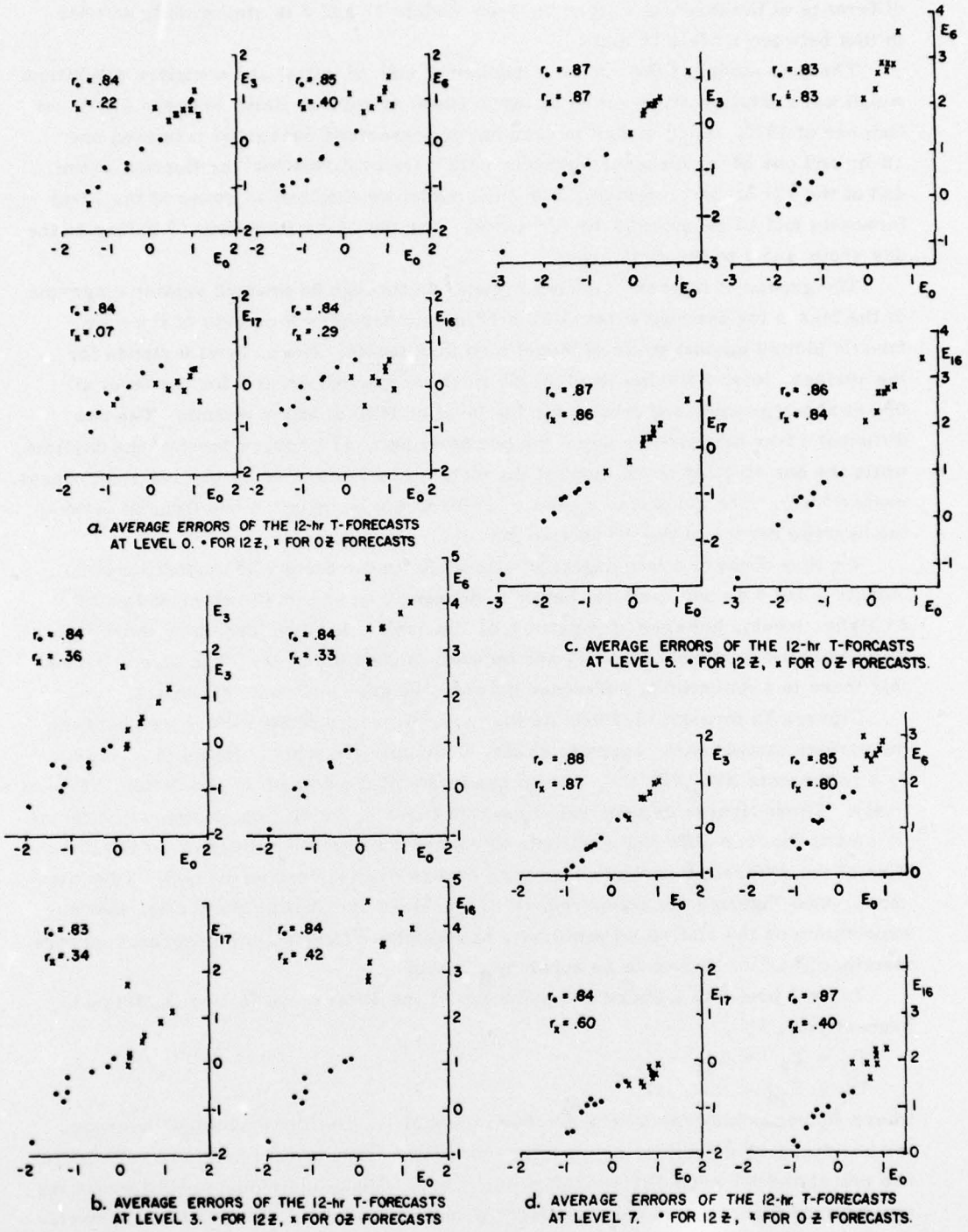
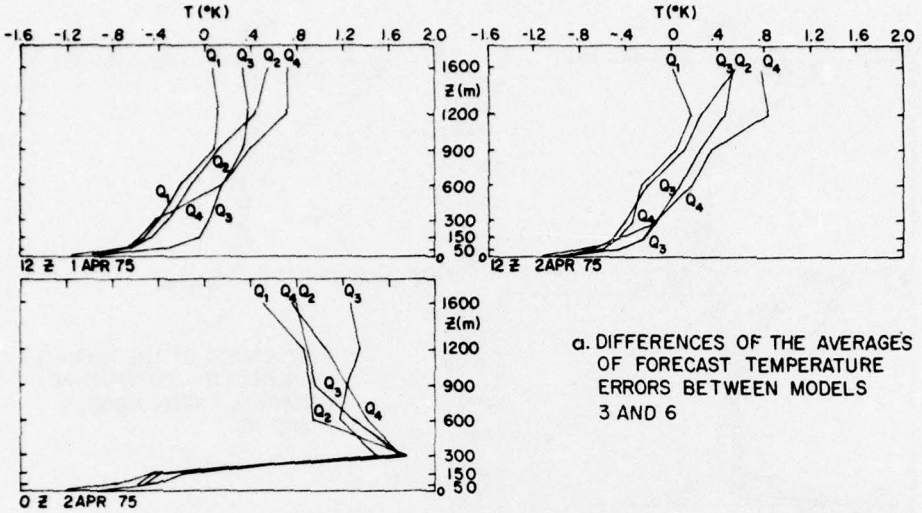
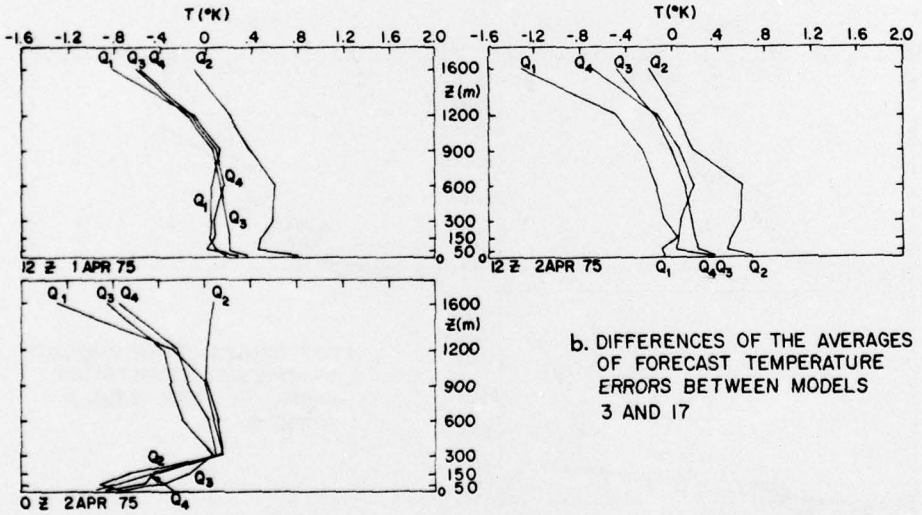


Figure 6. Average Errors of the 12-hr Temperature Forecasts

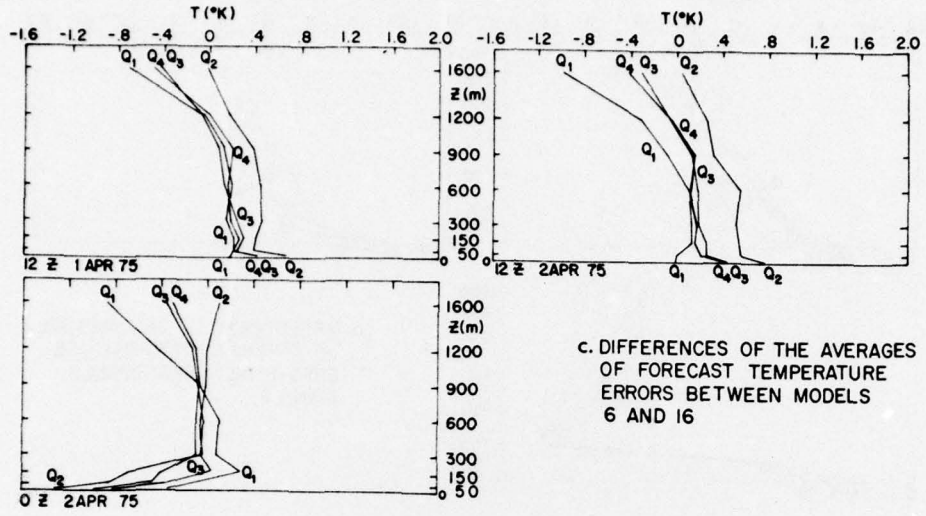


a. DIFFERENCES OF THE AVERAGES OF FORECAST TEMPERATURE ERRORS BETWEEN MODELS 3 AND 6



b. DIFFERENCES OF THE AVERAGES OF FORECAST TEMPERATURE ERRORS BETWEEN MODELS 3 AND 17

Figure 7. Differences of the Average in the Quadrants of Forecast Errors of the 12-hr Temperature Forecasts Between Different Models



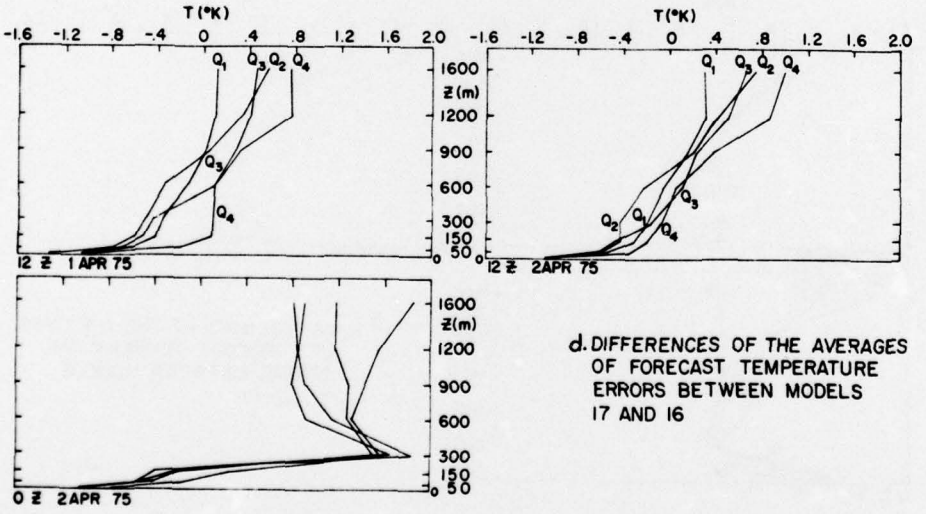


Figure 7. Differences of the Averages in the Quadrants of Forecast Errors of the 12-hr Temperature Forecasts Between Different Models (Cont)

Table 2. Means and Standard Deviations of Differences of Forecast Errors of the 12-hr Temperature Forecasts (°K)

				LEVEL							
				0		3		5		7	
				m	s	m	s	m	s	m	s
TIME	12Z	D_1	M	.15	.43	.96	.72	1.36	.87	1.51	.76
			S	.05	.07	.04	.07	.08	.04	.11	.06
		D_2	M	.16	.47	1.01	.74	1.42	.88	1.73	.83
			S	.04	.06	.03	.06	.08	.05	.08	.09
	0Z	D_1	M	.22	.95	2.93	1.42	2.34	.72	1.88	.71
			S	.07	.28	.12	.09	.12	.07	.10	.05
		D_2	M	.08	.77	2.97	1.32	2.38	.71	2.40	.74
			S	.07	.14	.11	.09	.09	.08	.08	.20

The table clearly exhibits the presence of a high degree of statistical similarity between the two differences of forecast errors, D_1 and D_2. Both parameters, the spatial average and the spatial standard deviation, of the difference of forecast errors between 16 and 17 are found to have values of group mean and variance very close to those for Models 6 and 3. The differences between the day and night groups in D_1 are also found in D_2 in remarkably close resemblance. The smallness of the group variances throughout the table attests to the consistency of the effect irrespective of concomitant conditions.

Table 3, in the identical format as that of Table 2, presents the results of the experiments concerning the differences D_3 and D_4 defined by

$$D_3 = E_{17} - E_3,$$
$$D_4 = E_{16} - E_6,$$

It summarizes the effect of the modifications introduced in Models 16 and 17. The uniformity of statistical similarity between the two differences and the clear distinction between the day and night groups are as well exhibited as they were in Table 2. The consistency of the effect is again reflected in the small values of the group variances except for those at level 7.

Similar statistical characteristics are also observed in the partition of error energies. The group mean and standard deviation, on each of the four levels considered in Tables 2 and 3, of the fraction of the error energy in each of the seven cells defined in Table 1 are presented for each of the four differences D_1, D_2, D_3

and D_4 and for the day and night groups in the series of figures in Figure 8. Resemblance of the profiles between D_1 and D_2 and between D_3 and D_4, and the distinction between the day and night groups are recognized quickly and easily in these figures.

Table 3. Means and Standard Deviations of Differences of Forecast Errors of the 12-hr Temperature Forecasts (°K)

				LEVEL							
				0		3		5		7	
				m	s	m	s	m	s	m	s
TIME	12Z	D_3	M	.42	.65	.22	.59	.15	.58	-.59	1.23
			S	.07	.08	.06	.08	.08	.07	.13	.13
		D_4	M	.42	.61	.27	.64	.20	.56	-.38	1.13
			S	.06	.07	.05	.06	.06	.04	.11	.12
	0Z	D_3	M	-.75	1.16	.01	.96	-.11	.69	-1.10	1.40
			S	.15	.27	.07	.08	.05	.15	.20	.24
		D_4	M	-.89	1.33	.05	1.02	-.07	.73	-.66	1.52
			S	.11	.32	.12	.07	.07	.15	.18	.26

Attention is also called to the differences in distribution of energy produced by the change in packing (Figures 8a and 8b) and that produced by the modifications introduced into Models 16 and 17 (Figures 8c and 8d). The differences between Models 3 and 6 and between 16 and 17 which illustrate the effect of the change in packing are distributed so that cell 1, at the long-wave end, is found to contain most of the energy - more than 40 percent during day and more than 60 percent at night at all levels except the surface. On the other hand, it is definitely the short-wave end, cell 7, that contains a large fraction of the energy in Figures 8c and 8d, which illustrate the effect the effect of the modification. Twenty to forty percent of the energy is found in cell 7 while the rest is evenly distributed among all other cells.

The distribution of energy at the surface is considerably different from those at higher levels with regard to the effect of the change in packing. Both during the day and at night a large fraction is found in cell 7. No such large contrast between the surface and upper levels is found in the effect of the modification.

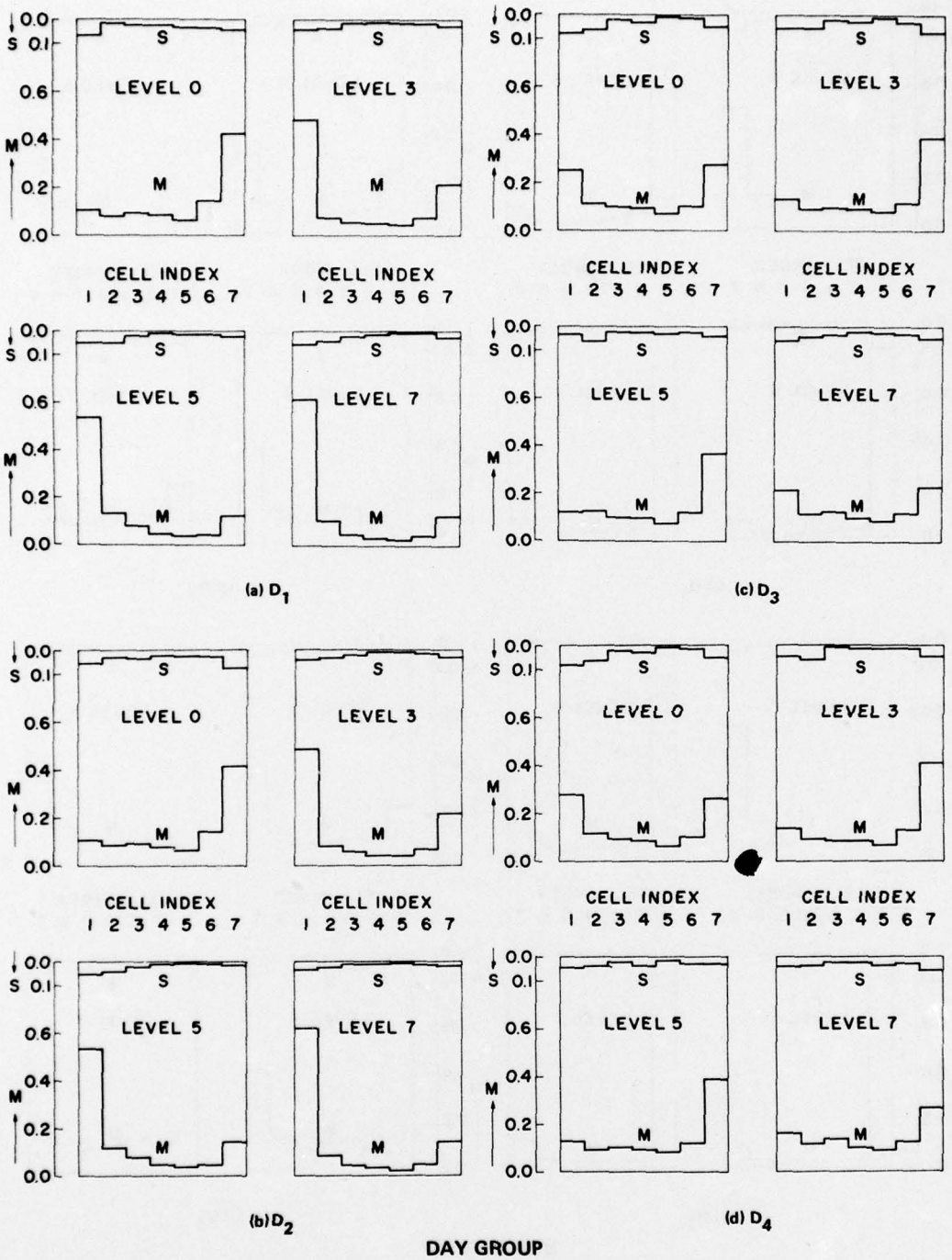


Figure 8. Distributions of Difference Energies into Spectral Cells

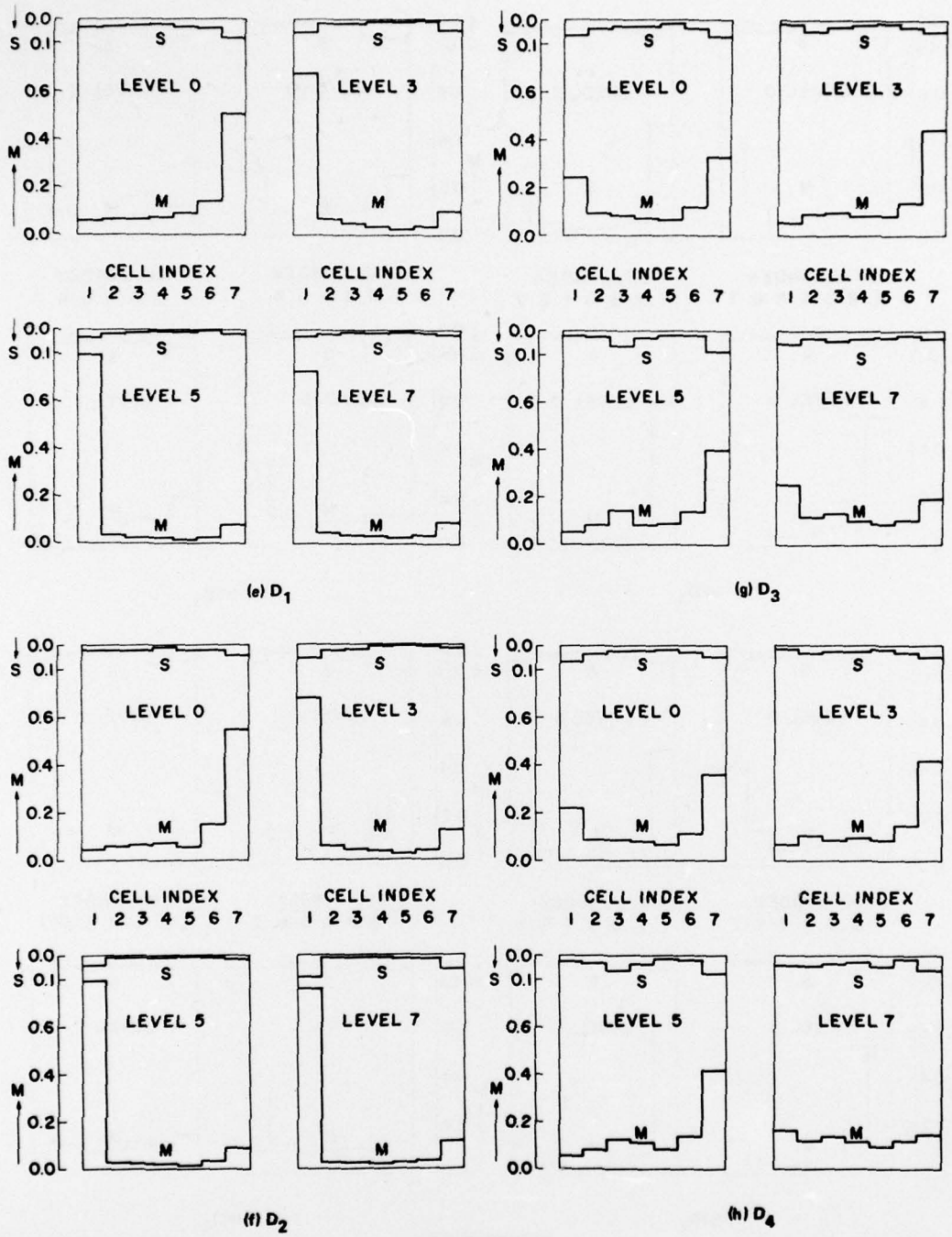


Figure 8. Distributions of Difference Energies into Spectral Cells (Cont)

Knowledge gained through the analyses shown in Figures 7 and 8 and in Tables 2 and 3 helps us to understand some of the relationships observed among the forecast errors of these models and elucidate the characteristics of the model-response to the changes introduced.

The assessment of the accuracies of the four models is summarized in Tables 4 and 5 and also in Table 6. Table 4 for the day group and Table 5 for the night group present the group means (M) and standard deviations (S) of the bias (m) and the standard deviation (s) of the forecast errors on four levels 0, 3, 5 and 7. Table 6, on the other hand, lists the group means (M) and standard deviations (S) of the root-mean-square (rms) forecast errors for the day and night groups of the four models on the four levels.

Table 4. Means and Standard Deviations of Forecast Errors of the 12-hr Temperature Forecasts, 12Z

			LEVEL							
			0		3		5		7	
			m	s	m	s	m	s	m	s
MODEL	3	M	-.66	3.59	-.70	3.62	-1.38	3.47	-.37	3.34
		S	.47	.36	.55	.38	.63	.31	.41	.23
	6	M	-.51	3.53	.25	3.56	-.02	3.24	1.14	3.30
		S	.46	.35	.56	.35	.61	.29	.43	.21
	17	M	-.25	3.75	-.49	3.75	-1.24	3.60	-.97	3.48
		S	.48	.36	.56	.39	.65	.31	.39	.30
	16	M	-.09	3.68	.52	3.66	.18	3.35	.76	3.37
		S	.48	.35	.57	.35	.65	.29	.41	.30

Table 5. Means and Standard Deviations of Forecast Errors of the 12-hr Temperature Forecasts, 0Z

			LEVEL							
			0		3		5		7	
			m	s	m	s	m	s	m	s
MODEL	3	M	.82	3.02	.63	2.96	.41	2.82	.76	3.02
		S	.23	.36	.60	.35	.38	.30	.25	.36
	6	M	1.04	3.08	3.56	3.27	2.75	2.97	2.64	3.23
		S	.27	.38	.60	.42	.44	.32	.23	.34
	17	M	.07	3.17	.65	2.93	.30	2.85	-.33	2.83
		S	.25	.40	.58	.33	.38	.30	.17	.23
	16	M	.14	3.11	3.62	3.16	2.69	2.95	1.98	2.97
		S	.33	.38	.63	.40	.40	.31	.20	.23

Table 6. Root-Mean-Square Forecast Errors of the 12-hr Temperature Forecasts (°K)

			LEVEL							
			0		3		5		7	
			12Z	0Z	12Z	0Z	12Z	0Z	12Z	0Z
MODEL	3	M	3.67	3.14	3.73	3.07	3.77	2.86	3.38	3.12
		S	.38	.34	.41	.45	.39	.33	.21	.36
	6	M	3.61	3.26	3.62	4.85	3.30	4.05	3.51	4.17
		S	.42	.35	.33	.65	.28	.45	.30	.35
	17	M	3.78	3.18	3.81	3.06	3.85	2.89	3.64	2.86
		S	.36	.40	.40	.42	.37	.33	.22	.22
	16	M	3.70	3.14	3.74	4.77	3.41	3.99	3.45	3.61
		S	.35	.37	.33	.63	.28	.42	.35	.33

Let us look first at the effects of the difference in packing on the day group. We find that in Table 2 the GL packing in Models 6 and 16 shifts the bias toward the positive direction in relation to the GWC packing of Models 3 and 17. We also observe that the shift increases with height from 0.2°K at the surface to the value of about 1.5 ~ 1.7°K at the top of the boundary layer. On the other hand, the standard deviation remains fairly constant about 0.7 ~ 0.8°K. The effect of the positive shift on the bias in changing from the GWC- to GL-packing is quite obvious upon comparing Model 3 with 6 or, Model 17 with 16 on the entries (m, M) of Table 4. At the same time, comparisons of the entries (s, M) between the same pairs of models show that the GL packing consistently yields smaller values than the GWC packing.

We also note from these two tables that the contribution from the difference in packing to the total variance of the forecast errors amounts to no more than 8 percent in any case. The overall result of all these features as they appear in the rms forecast errors is that the GL packing produces better accuracy chiefly by reducing the magnitude of the bias.

When we examine the same effect in the night group by inspecting Tables 2 and 5 and by associating them with Table 6, we find that the same agent, namely, the tendency of the GL packing to shift the bias toward the positive direction, is in this case causing Models 6 and 16 to perform less accurately than Models 3 and 17.

When these phenomena are considered in reference to the details of the model, they seem to indicate that the cause lies in a defect in the model mechanism that is responsible for distribution and dissipation of heat. We have, therefore, concluded that the physical assumptions and the mathematical procedures employed in estimating the eddy diffusivity for heat and moisture should be investigated for possible improvement.

Let us consider next the effect of the modification introduced in Models 16 and 17, namely, doubling the size of the time step and changing the temperature estimate at 2 km above ground. Table 3 shows that during the day the positive bias decreases with height in the lower three levels from 0.4 to 0.2°K and then changes into a negative bias at about -0.5°K at the top (level 7). The spatial standard deviation, on the other hand, remains at about 0.6°K in the lower three levels and jumps to 1.2°K at level 7. We believe that the out-of-line characteristics in both bias and standard deviation at level 7 is the result of the effect of the change in the temperature estimate at 2 km above ground. The individual samples, not presented here, reveal rather emphatically that the effect of this change hardly reaches below level 6 in the first 12 hours of prediction and is barely discernible at level 3 even after 24 hours.

These characteristiceffects, when embedded into the total model, led to smaller negative biases at the three lower levels and larger negative biases at level 7

(entries (m, M). They also led to slightly larger standard deviations of about 3.5°K with a nearly constant difference of 0.1°K (entries (s, M)) in both Models 17 and 16 in comparison with Models 3 and 6, as is shown in Table 4. When these biases and standard deviations are combined into rms forecast errors, we find in Table 6 that Models 16 and 17 consistently perform less accurately than Models 3 and 6 at all levels with the difference of 0.1°K in rms forecast error. With the estimated standard deviation of 0.13°K associated with these estimates of the rms forecast errors, we cannot attach too much statistical significance to a difference of this magnitude.

When we turn to the night group in Table 3, we find both the bias and the standard deviation to be quite different from those of the day group. The bias is now definitely negative at both level 0 and level 7 at about -0.8 or -0.9°K while approximately zero in the middle of the layer at both level 3 and level 5. At the same time the standard deviation becomes larger than that of the day group everywhere - above 1°K at levels 0 and 7 and about 1°K at levels 3 and 5. Relatively large values of the group variances of both parameters at night compared with those during day seem to imply a weaker influence of the effect at night.

Superimposed on the total model, these features produce reduction of the magnitude of bias at levels 0 and 7, but keep those at levels 3 and 5 unchanged between Models 3 and 17 and between Models 6 and 16. The standard deviation at level 0 increases slightly, but that at level 7 decreases with the modification, while those in the middle remain about the same. The total result on the rms forecast error is a virtual stand-off between the two pairs everywhere except level 7, where the modification leads to a smaller error.

When we put all these observations together under the light of the model structure, we must conclude that the model contains other sources of error which are of such a nature as to obscure and obliterate the expected larger truncation error of the larger time step in Models 16 and 17.

The study shows, furthermore, that neither the individual differences nor their combination can account for the major fraction of the observed forecast errors. As a matter of fact, they seem to produce a small change in the rms forecast error of the total model. While this does not exclude the possibility of the presence of an identifiable single defect which can account for a large fraction of the forecast error of the model, it may be a warning of an arduous task ahead for our search for the defects, even with the use of the proposed probing experiments.

6. CONCLUSIONS

How should one tackle problems in the nonlinear world with the tools designed for use in the linear world? Convinced that no universal method has yet been discovered, we attempted to develop a procedure by which the only alternative to a cure-all, the trial-and-error method, may be executed on a rational basis. In so doing, we found it necessary to clarify a few concepts and to define appropriate measures for some qualities that are pertinent to the problem.

With the state of meteorological observations as they exist today we feel that the method of twin experiment is the best technique for probing the nature of dynamical prediction models and developing the knowledge which we believe is essential to an understanding and correction of their defects.

As to representation and characterization of the forecast error, which is the ultimate measure of the model imperfection, we have suggested measures which are believed to provide valuable insight into the most readily discernible features. Constant vigilance is, however, believed to be the key to detection of any significant symptom; the mode of representation must be selected to insure that there is no distortion or obscuration of the symptom.

The results of the first series of twin experiments have been presented and discussed largely for purposes of illustrating the approach we have chosen to follow. However, in spite of the preliminary and tentative nature of these results they indicate the following:

(1) The physical assumptions and mathematical procedures of the model used in estimating the eddy diffusivity of heat and moisture should be a target of investigation,

(2) An intensive search should be made for a large source of error which could be masking the strong effect which would otherwise be expected from doubling the time-step.