NOAA Technical Memorandum ERL PMEL-41

DATA INTERCOMPARISON THEORY

IV. TERCILE TESTS FOR LOCATION, SPREAD AND PATTERN DIFFERENCES

Rudolph W. Preisendorfer
Curtis D. Mobley

Pacific Marine Environmental Laboratory
Seattle, Washington
December 1982

UNITED STATES
DEPARTMENT OF COMMERCE
Malcolm Baldrige,
Secretary

NATIONAL OCEANIC AND
ATMOSPHERIC ADMINISTRATION
John V. Byrne,
Administrator

Environmental Research
Laboratories
George H. Ludwig
Director

NOTICE

Mention of a commercial company or product does not constitute an endorsement by NOAA Environmental Research Laboratories. Use for publicity or advertising purposes of information from this publication concerning proprietary products or the tests of such products is not authorized.

Contribution No. 603 from NOAA's Pacific Marine Environmental Laboratory

TABLE OF CONTENTS

Abstract

Two data sets can be examined for closeness by considering them as field maps and pretending that one is trying to forecast the other. This is done by applying some recently devised forecast verification techniques to the pair of data sets. The technique we apply for the purpose is the tercile (or Trinomial Stochaster) technique wherein, over a given set of points in space, the two fields have their 0-class, 1-class and 2-class errors tallied, and examined for statistical significance. These class errors can be used to gauge the closeness of the three main attributes of the data sets: their *locations* (akin to averages), *spreads* (akin to variances), and (spatial or temporal) *patterns*. In illustration, the tercile technique is applied three times: to show how to gauge the effects of different objective analysis methods on the same raw data set; to examine the self-predictability (and hence noise or information content) of a data set; and to devise a new principal-component selection rule using the concept of self-predictability.

Data Intercomparison Theory

IV. Tercile Tests for Location, Spread and Pattern Differences

Rudolph W. Preisendorfer

Curtis D. Mobley

1. Introduction

The approaches to the problem of data intercomparison are manifold, as we have begun to see in the first three reports of the present series on Data Intercomparison Theory. In the present note we introduce still another approach to the data intercomparison problem, namely that based on the tercile classification of the range of values of a physical field. The present method draws out the thread of ideas that began to form in the closing section of DIT(III)*, wherein we developed the method of r-tile classification, and indicated its applications to intercomparing principal vectors (EOF's), principal components, and other objects derivable from the principal decomposition of data sets. In the present note we specifically set r = 3 and therefore work with one of the simplest r-tile classifications, namely *terciles*. Moreover, we shall work *directly* with observed field values rather than their principal parts. These two choices result in an easily interpreted set of tests for the differences in spatial or temporal aspects of location, spread, and pattern of two data sets. The reference distributions, by which we can decide on the significances of the differences of these various attributes between two data sets, can be generated under suitable conditions

---

* A complete list of titles in the present series of reports is given in §7, below.

by any one of the five general procedures (IOP,EOP,APP,PPP,CIP) defined and illustrated in DIT(II).

We shall illustrate by means of three examples the use of the tercile method of data intercomparison, using a combination of IOP and CIP, i.e., a combination of the Ideal Observation Procedure and Classical Intercomparison Procedure. The first example shows how the method can be used to detect whether *significant differences exist* in the locations, spreads, and patterns of three equatorial temperature fields arising from three different objective analysis schemes applied to a common set of raw data. The second example applies the Tercile Method to the *self-prediction property* of one of the three equatorial temperature data sets in the first example. The self-prediction property is the one that gives a measure of the non randomness (or information content) of a data set. The self-prediction property is one facet in the solution of the general climate predictability problem. Finally, the Tercile Method will be used to suggest a new form of *principal-component selection rule* based on the self-prediction idea.

Acknowledgments

We are indebted to Dr. Cho-Teng Liu for making his data set available prior to publication, and also for doing some special runs needed in the applications of §§4,5. The Liu data set computations and the work of Dr. Mobley were supported through grants from NOAA's Equatorial Pacific Ocean Climate Studies (EPOCS) program. Lai Lu typed the manuscript and Gini May drew the figures.

2. S-Synoptic Tests

We shall consider two main developments of the Tercile Method. The present development takes place in the spatial-synoptic (or S-Synoptic) setting wherein we shall test whether two given data sets $\underline{D}' = \{d'(t,x): t = 1,\cdots,n;\ x = 1,\cdots,p\}$ and $\underline{M}' = \{m'(t,x): t = 1,\cdots,n;\ x = 1,\cdots,p\}$ have significant differences in spatial location, spatial spread and spatial pattern. In §3 we shall summarize the *dual* development for temporal differences, resulting in the T-Synoptic Tests.

A. Finding Tercile Boundaries

The first step in the S-Synoptic Tests is to compute the time averages of the $d'(t,x)$ and $m'(t,x)$. Thus we write, with $x = 1,\cdots,p$,

$$'d_o(x)' \text{ for } n^{-1} \sum_{t=1}^{n} d'(t,x) \tag{2.1}$$

$$'m_o(x)' \text{ for } n^{-1} \sum_{t=1}^{n} m'(t,x) \tag{2.2}$$

These averages, specifically for the $\underline{D}'$ matrix, are shown in Fig. 2.1, which gives an overview of both the S-Synoptic and T-Synoptic Analyses.

Next, form the *anomalies*, for $t = 1,\cdots,n;\ x = 1,\cdots,p$,

$$d(t,x) \equiv d'(t,x) - d_o(x) \tag{2.3}$$

$$m(t,x) \equiv m'(t,x) - m_o(x) \tag{2.4}$$

From these, at each x, find the tercile class boundaries $a_d(x)$, $b_d(x)$ and $a_m(x)$, $b_m(x)$. These are shown schematically for the $\underline{D}'$ set in Fig. 2.1. Thus $a_d(x)$, $b_d(x)$ have the property that they split the n values $d(t,x)$, $t = 1,\cdots,n$

Determining tercile class boundaries for
S-Synoptic and T-Synoptic analyses

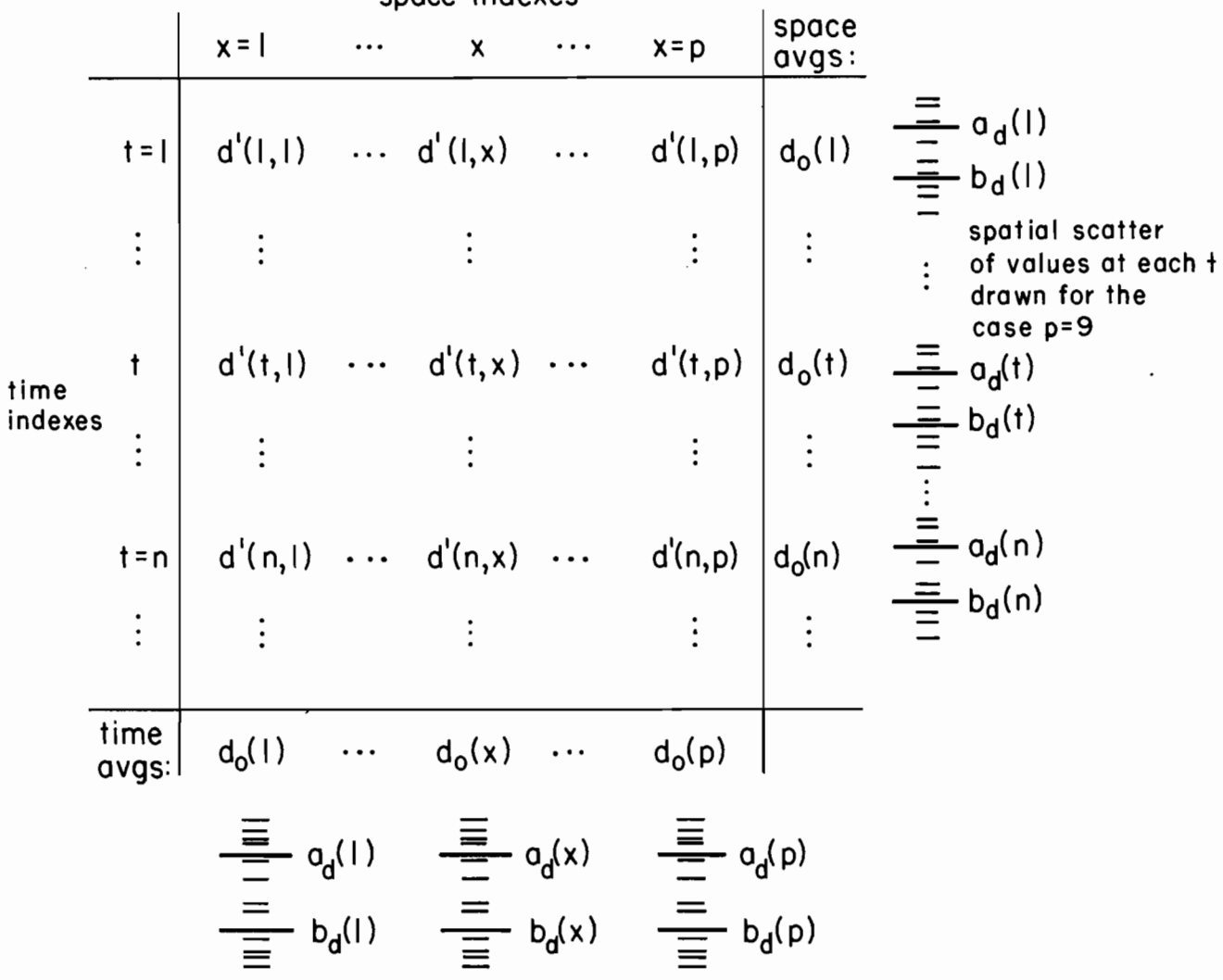


Fig. 2.1

at each x into three equally populous groups: the above-group (A), the normal-group (N), and the below-group (B). This may be done on the computer by ordering the n d(t,x) anomaly values at x in increasing size (and relabeling the t-values):

$$d(t_1,x) \leqq d(t_2,x) \leqq \cdots \leqq d(t_n,x) \tag{2.5}$$

and then partitioning these n values into three subsets:

$$d(t_1,x) \leqq \cdots \leqq d(t_b,x),\ d(t_{b+1},x) \leqq \cdots \leqq d(t_a,x),$$
$$d(t_{a+1},x) \leqq \cdots \leqq d(t_n,x) \tag{2.6}$$

At the subset boundaries we define:

$$a_d(x) \equiv \tfrac{1}{2}[d(t_a,x) + d(t_{a+1},x)] \tag{2.7}$$
$$b_d(x) \equiv \tfrac{1}{2}[d(t_b,x) + d(t_{b+1},x)] \tag{2.8}$$

The tercile boundaries $a_m(x), b_m(x)$ for the model set $\underline{M}'$ are found in the same way*, and are given at each x by formulas analogous to (2.7), (2.8) (simply replace "d" by "m" in (2.5)-(2.8)).

B. Pattern Test

The *Pattern Test* compares the spatial patterns of the maps $\underline{d}(t)$ and $\underline{m}(t)$ at each time t, where, for t = 1,···,n,

---

* If n is not divisible by three, then decide how to place the odd data points. For example, place the first odd point in N and the second (if it exists) in A.

$$\underline{d}(t) = [d(t,1),d(t,2),\cdots,d(t,p)]^T \tag{2.9}$$

$$\underline{m}(t) = [m(t,1),m(t,2),\cdots,m(t,p)]^T \tag{2.10}$$

Thus, at each time t find what tercile cell (A,N, or B) d(t,x) is in, and similarly for m(t,x). See top panel, Fig. 2.2. In that figure, d(t,x) is in B and m(t,x) is in N. Therefore these values subtend at x a 1-class error. (If m(t,x) happened to be in B or A, then m(t,x) and d(t,x) would subtend at x a 0-class or 2-class error, respectively.) We now go systematically through all points x and find the j-class errors (j = 0,1,2) subtended by $\underline{m}(t)$ and $\underline{d}(t)$ at those points. Let u(t), v(t), and w(t) be the total number of 0-,1-,2-class errors, respectively, subtended by $\underline{d}(t)$ and $\underline{m}(t)$. As a check we should have:

$$u(t) + v(t) + w(t) = p \tag{2.11}$$

for each $t = 1,\cdots,n$.

The two maps, as spatial patterns, are close when u(t) is large and the moment

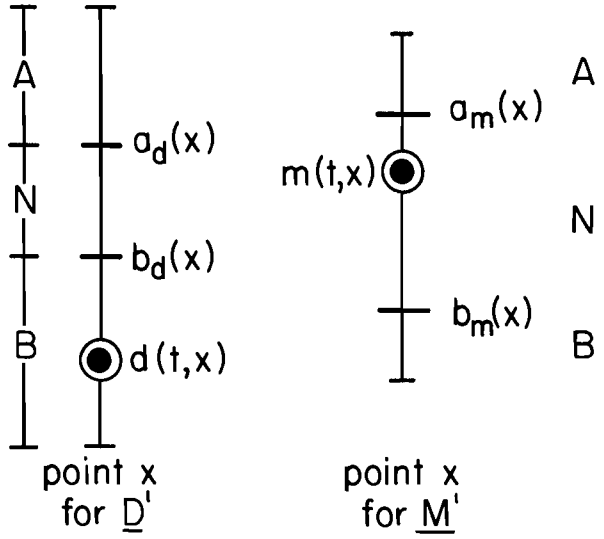
$$m(t) = v(t) + 2w(t) \tag{2.12}$$

is small. The two maps are identical at time t (as far as tercile classifications go) when

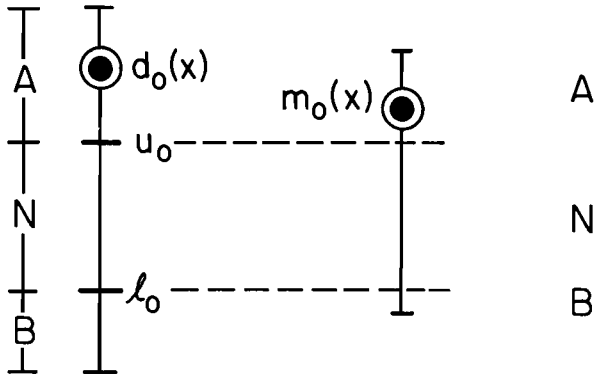
$$u(t) = p, \quad m(t) = 0. \tag{2.13}$$

An impression of the statistical significance of a score u(t),v(t),w(t) at time t is obtainable by comparing this score with that obtained by a

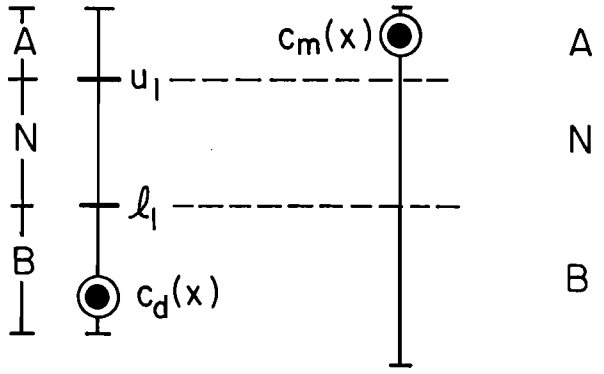
S-SYNOPTIC TESTS FOR LOCATION, SPREAD AND PATTERN



PATTERN TEST
Showing tercile cells A,N,B and a 1-class error between D', M' at point x



(D'-Base) LOCATION TEST
Showing tercile cells A,N,B and a 0-class error between D', M' at point x



(D'-Base) SPREAD TEST
Showing tercile cells A,N,B and a 2-class error between D', M' at point x

Fig. 2.2

process that selects maps like $\underline{d}(t)$, $\underline{m}(t)$ in a random fashion as follows. At each x, and for given fixed t, let d(t,x) be randomly placed in cell A, or N, or B with equal probability, namely 1/3. Similarly at each x, for the given fixed t let m(t,x) be placed with probability 1/3 into one of the terciles. These placements of d(t,x) and m(t,x) at x are independent of the placements at every x′ other than x. From these two randomly selected maps, we can find the u,v,w values for them. Under these conditions, the joint probability p(u,v,w) of the three class-errors u,v,w is readily found (Appendix A, using the Trinomial Stochaster). Moreover, the probability distributions for the individual u,v,w error counts are also readily obtained. From these, for a given significance level, 100α%, $0<\alpha<1$, we can judge if u(t), v(t), w(t) are significantly large or small, as the case may be. The use of these randomly found u,v,w scores will be illustrated in §4.

The net result of the Pattern Test is a decision at each t whether or not the triple of scores (u(t),v(t),w(t)) subtended by $\underline{d}(t)$ and $\underline{m}(t)$ is significant on some 100α% level, $0<\alpha<1$. If, for example, u(t) is significantly *large*, then the spatial patterns of $\underline{d}(t)$, $\underline{m}(t)$ may be said to be significantly close as regards 0-class errors. Other significance measures, in addition to u(t), are supplied by significantly *small* values of v(t), w(t), or m(t).

C. Location Test

To carry out a (spatial) *Location Test* on the n×p data sets $\underline{D}'$,$\underline{M}'$, first find the time averages $d_o(x)$, $m_o(x)$, $x = 1,\cdots,p$, as in (2.1), (2.2). Then order the $d_o(x)$ (after relabeling) in increasing size:

$$d_o(x_1) \leqq \cdots \leqq d_o(x_p) \tag{2.14}$$

and find the tercile boundaries ℓ_o, u_o, for these averages, analogously to (2.6):

$$d_o(x_1) \leqq \cdots \leqq d_o(x_\ell) \leqq d_o(x_{\ell+1}) \leqq \cdots \leqq d_o(x_u) \leqq d_o(x_{u+1}) \leqq \cdots \leqq d_o(x_p) \quad (2.15)$$

The class boundary values at x_ℓ, x_u, are defined as:

$$\ell_o \equiv \tfrac{1}{2}[d_o(x_\ell) + d_o(x_{\ell+1})] \quad (2.16)$$

$$u_o \equiv \tfrac{1}{2}[d_o(x_u) + d_o(x_{u+1})] \quad (2.17)$$

Thus the integers ℓ and u stand at or near the 1/3 marks of p. If p were of the form $p = 3q$, then $\ell = q$ and $u = 2q$. Otherwise, proceed as in the footnote below (2.7), (2.8).

The Location Test now follows along the lines summarized in the middle panel of Fig. 2.2. At each x see what tercile $d_o(x)$ is in, and what tercile $m_o(x)$ is in. For each of these decisions use ℓ_o and u_o as defined in (2.16), (2.17). In Fig. 2.2, $d_o(x)$ is in A and $m_o(x)$ is in A, and so they subtend a 0-class error. Repeat this $(d_o(x), m_o(x))$ class-error determination at each of the p places in physical space. Let u,v, and w be the resultant total member of 0-class, 1-class, and 2-class errors, so that, as in (2.11), these must sum to p. The statistical significances of u,v,w, individually or in common may now be determined exactly analogously as in the Pattern Test above, using the Trinomial Stochaster.

Observe that the present test has used the $\underline{D}'$ set as a basis of comparison (ℓ_o, u_o were derived from the $d_o(x)$'s). The dual test can be based on $\underline{M}'$. Generally the triple of scores: u,v,w resulting from this alternate test based on $\underline{M}'$ will not be the same as that based on $\underline{D}'$. Physical reasoning

should decide which base is to be preferred; or otherwise, simply average the two u's, two v's, and two w's from each type of Location Test. Note that the Pattern Test was symmetric from the outset.

D. Spread Test

To perform the (spatial) *Spread Test* on the n×p data sets $\underline{D}', \underline{M}'$ use the class boundaries $a_d(x), b_d(x)$ and $a_m(x), b_m(x)$ of the Pattern Test in par A to define, for $x = 1, \cdots, p$,

$$c_d(x) \equiv a_d(x) - b_d(x) \tag{2.18}$$

$$c_m(x) \equiv a_m(x) - b_m(x) \tag{2.19}$$

Then order the p values of $c_d(x)$, $x = 1, \cdots, p$, in increasing size after the manner of (2.15):

$$c_d(x_1) \leqq \cdots \leqq c_d(x_\ell) \leqq c_d(x_{\ell+1}) \leqq \cdots \leqq c_d(x_u) \leqq c_d(x_{u+1}) \leqq \cdots \leqq c_d(x_p) \tag{2.20}$$

and define:

$$\ell_1 \equiv \tfrac{1}{2}[c_d(x_\ell) + c_d(x_{\ell+1})] \tag{2.21}$$

$$u_1 \equiv \tfrac{1}{2}[c_d(x_u) + c_d(x_{u+1})] \tag{2.22}$$

The Spread Test now follows along the lines summarized in the lower panel of Fig. 2.2. At each x, determine in which tercile $c_d(x)$, $c_m(x)$ lie. In the Figure $c_d(x)$ lies in B and $c_m(x)$ lies in A, and so they subtend a 2-class error. Repeat this $(c_d(x), c_m(x))$ class-error determination at each of the p places in physical space. Let u,v, and w be the resultant total number of 0-, 1-, and 2-class errors. The statistical significances of u,v,w are deduced as

in the Pattern and Location tests. Observe that the Spread Test is based on $\underline{D}'$; an alternate test may be based on $\underline{M}'$, and score-averages found, as in the Location Test.

E. Notes on Alternate Procedures for Generating Tercile Tests

(*i*) Implicit in the test descriptions above was the assumption that there were enough maps $\underline{d}(t)$ and $\underline{m}(t)$, $t = 1,\cdots,n$ to generate a workable tercile decomposition of the values $d(t,x)$, $m(t,x)$, $t = 1,\cdots,n$ at each x. Thus we have implicitly used on $\underline{D}'$ and $\underline{M}'$ themselves the Ideal Observation Procedure (IOP) defined in DIT(II). When the number of samples $\underline{d}(t)$, $t = 1,\cdots,n$ is not adequate for this purpose, resort may be made to the Classical Intercomparison Procedure (CIP) using the suggestions made in DIT(II). Indeed, the u,v,w statistics defined above (and in Appendix A) are developed and applied in the CIP setting.

(*ii*) It may turn out that we have several $n\times p$ data sets $\underline{D}_i'$ $i = 1,\cdots,w$ available but individually the n maps $\underline{d}(t)$, $t = 1,\cdots,n$ in each are not enough to form a workable tercile decomposition of the map values. (For example n in such a case would be 5.) It may be meaningful, however, to pool the nw maps from the given set of $\underline{D}_i'$ matrices, and therefore have a sufficient number to work with.

(*iii*) When there are enough maps $\underline{d}(t)$, $\underline{m}(t)$, $t = 1,\cdots,n$ for IOP *within* $\underline{D}'$ and *within* $\underline{M}'$, then it may be of interest to apply the APP (of DIT(II)), suitably modified, to $\underline{D}'$ and $\underline{M}'$ and the tercile-forming method. Thus, any partitioned $\underline{D}'$ could be used to generate tercile decompositions and intercomparisons. Thereby, we could form DD cdfs of v, or w. These DD cdfs would be compared, in the manner described in APP, to the DM cdfs of the respective statistics v, or w. (For u, one would use the *left* tail of the DD cdf to reject H_o.)

(*iv*) The PPP may also be applied to the task of generating reference distributions of the u,v, and w statistics. The idea here would be to see where the u,v, or w, subtended by the given maps of $\underline{D}'$ and $\underline{M}'$, would fall in the reference distributions u,v, and w, respectively, generated by the pool and partition procedure. Significantly small u values, and significantly large v,w values would be the bases for rejecting H_o.

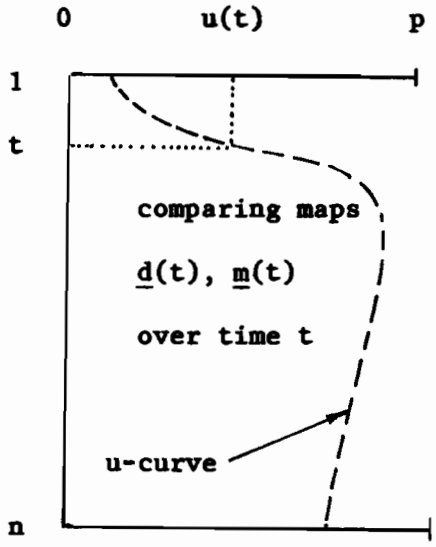
The above alternate procedures for generating terciles and reference distributions for u,v, or w have not been studied for their practicality. They are suggestions for further research.

3. T-Synoptic Tests

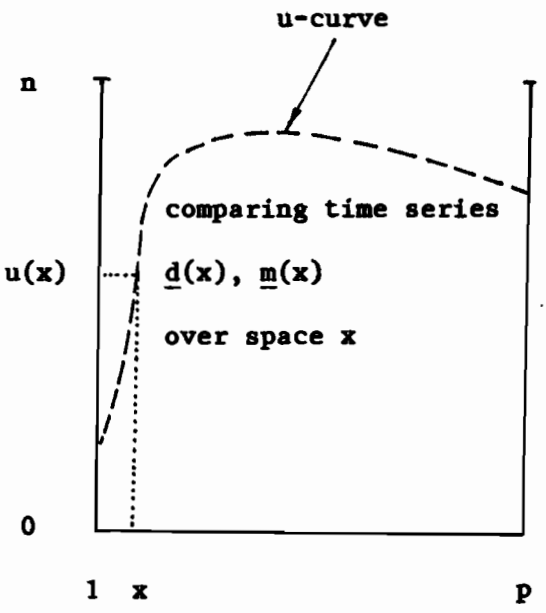
By means of Fig. 2.1 we see that there is a completely dual set of tests to the S-Synoptic set of tests described in §2. Thus, instead of averaging d(t,x), m(t,x) over t at a fixed x, we can average these quantities over x at a fixed t. To save notation we use the same symbols for this dual arithmetic activity. Thus $d_o(x)$ is now $d_o(t)$, and $a_d(x)$ becomes $a_d(t)$, and so on, where $t = 1,\cdots,n$. It is desirable (but not necessary) that p be divisible by 9; then all expected values of u,v,w and m for the Trinomial Stochaster are integers. The critical values for the individual statistics u,v,w, and m are obtainable from the Binomial distribution with the appropriate elementary probabilities (see Appendix A).

The interpretation of the results of a T-Synoptic Analysis is dual to their S-synoptic counterparts. Thus, in a *pattern* analysis in the present T-Synoptic case, the quantity u(x) will count the number of times, out of a sample of n, that the $\underline{M}'$ and $\underline{D}'$ sets have the same anomaly. The diagrams below will help visualize this case. A large x-set of high u(x) values will

indicate that the *temporal evolutions* of the two sets are alike at most of the points x of the region of interest. This is dual to the S-Synoptic Analyses wherein a long t-string of high u(t) values indicates that the *spatial patterns* of the two sets are alike at most of the times t of the period of interest.



S-Synoptic Tests



T-Synoptic Tests

Of course similar dual interpretations can be made for the remaining statistics v, w, and m.

Just as we can concentrate attention on a space subregion when doing S-Synoptic Analysis, so too can we concentrate attention on a time subperiod when doing T-Synoptic Analysis. The hypergeometric distribution can be used with equal facility in both types of analysis when we are seeking significant behavior in the appropriate subsets.

4. Application: Effects of Different Objective Analysis Schemes

A. In a recent study, Liu (1982) has produced an objectively analyzed sea surface temperature set consisting of semi-monthly averages of the SST over the tropical Pacific during the years 1975 to 1980, inclusive. Therefore there are 144 "snapshots" of the SST pattern over the region, (20°S, 20°N) and (80°W, 160°E) shown in Fig. 4.1. The objective analysis scheme was basically of the Cressman (1959) type, as modified by Levitus and Oort (1977), with additional smoothing operations, outlier cutoff, data weighting, and gradient-dependent features added by Liu. The four new features of Liu's scheme are summarized in the right column of Table 4.1. We were mainly interested in seeing how important the effect on the resulting data set was, as contributed by the gradient-dependent and data-weighting features of Liu's scheme. Accordingly, we asked Liu to remove these features from his scheme to obtain what we call (for historical reasons) the "modified Levitus-Oort" scheme. The resultant scheme is summarized in the middle column of Table 4.1. This scheme yielded a second data set to be compared with the Liu set. We also asked Liu to produce a third data set by completely removing the diffuser-smoother and curvature-corrector features of his program.* The scheme producing this third set is shown in the left column of Table 4.1. We designate this scheme as the "Original Levitus-Oort" scheme, since it incorporates only the essential features of Levitus and Oort (1977), namely grid-to-rack interpolator, an ordinary corrector-on-the-grid, an updater, and a Laplacian smoother.

---

* These various features of Liu's scheme will be developed in detail in DIT(V). They may also be discerned in Liu (1982). For our present purposes, the descriptions in Table 4.1 are adequate.

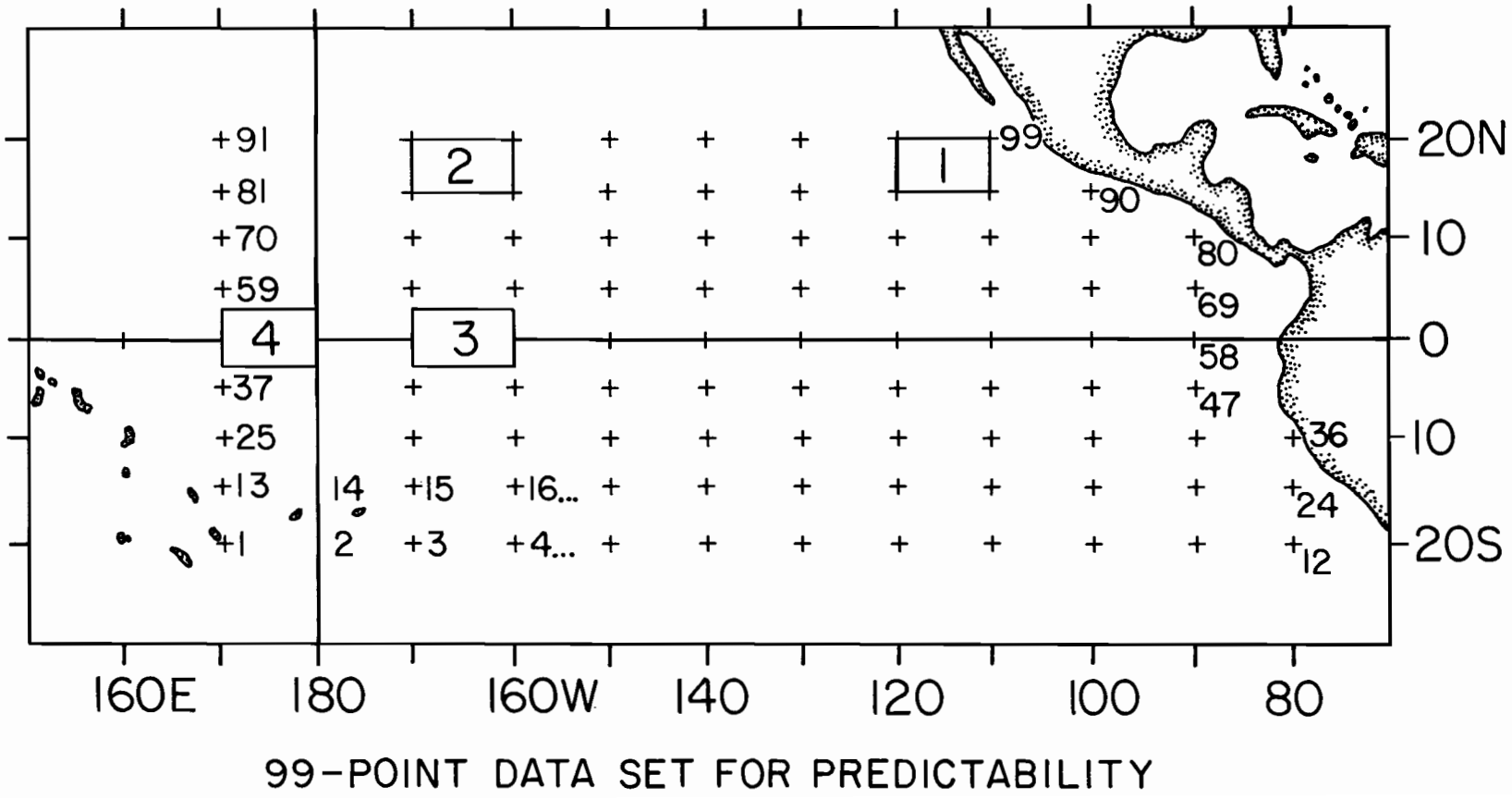

FOUR REGIONS (5° x 10°) FOR DATA INTERCOMPARISON (24 POINTS EACH)
1 = HIGH DATA DENSITY, LARGE SST GRADIENT
2 = HIGH DATA DENSITY, SMALL SST GRADIENT
3 = MODERATE DATA DENSITY, SMALL SST GRADIENT
4 = LOW DATA DENSITY, SMALL SST GRADIENT
1
2
3
4
99
90
80
69
58
47
36
24
12
+91
+81
+70
+59
+37
+25
+13
+1
14
2
+15
+3
+16...
+4...
20N
10
0
10
20S
160E
180
160W
140
120
100
80
99-POINT DATA SET FOR PREDICTABILITY

Fig. 4.1

Table 4.1

OBJECTIVE ANALYSIS SCHEMES

BASIC OPERATIONS ON ORIG. DATA	ORIG. LEV-ORT	MODF. LEV-ORT	LIU
GRID TO RACK INTER-POLATOR	X	X	X
CORRECTOR ON GRID	SST GRADIENT INDEPEND NO OUTLIER CUTOFF NO DATA WGT	SST GRADIENT INDEPEND OUTLIER CUTOFF NO DATA WGT	SST GRADIENT DEPEND OUTLIER CUTOFF DATA WGT
DIFFUSER-SMOOTHER	NONE	SST GRADIENT INDEPEND	SST GRADIENT DEPEND
UPDATER	X	X	X
CURVATURE-CORRECTOR	NONE	X	X
LAPLACIAN-SMOOTHER	X	X	X

B. We selected four regions in the Pacific area over which the three different data sets were to be examined for the 18 month period from January 1975 to June 1976. The regions were selected on the basis of spatial data density and SST gradient, and are shown in Fig. 4.1. Each region is 5°×10° in extent and has p = 24 points. Over the 18 month period, for a given region, we have two "snapshots" or maps per month; hence in the theory above, n = 36. Region 1 lies off the Coast of Mexico, and has high data density with large SST gradients. We selected the second region around Hawaii for its high data density but low SST gradient. Region 3 on the equator below Hawaii has moderate data density and small SST gradient. Finally, region 4 has low data density and small SST gradient. We anticipated considerably different resultant data sets as regards location, spread, and pattern, because of the different ways the sets were produced through modification of Liu's objective analysis scheme. The results of applying the S-Synoptic tests for pattern, location, and spread to these data sets are shown in Figures 4.2, 4.3, 4.4. In these figures the *averages* of u,v over the 36 semi-monthly periods are shown. (A primer in reading these diagrams is given in Appendix A. The horizontal and vertical lines in the figures give 5% significance levels for v and u, respectively.)

In Fig. 4.2, we intercompare the 36×24 set of Liu (=$\underline{D}'$) with the original 36×24 set of Lev-Ort (=$\underline{M}'$). See the open squares. Also Liu (=$\underline{D}'$) was compared with modified Lev-Ort (=$\underline{M}'$). See the solid circles. We can see by these dots, being crowded near the vertex, that data density and SST gradients play a small role in the objective analysis effects on spatial patterns. In other words, the gradient-dependent features of Liu's scheme are not important additions to the Cressman and Levitus-Oort techniques, as regards pattern changes. That is, by dropping these features from Liu's scheme we produced a

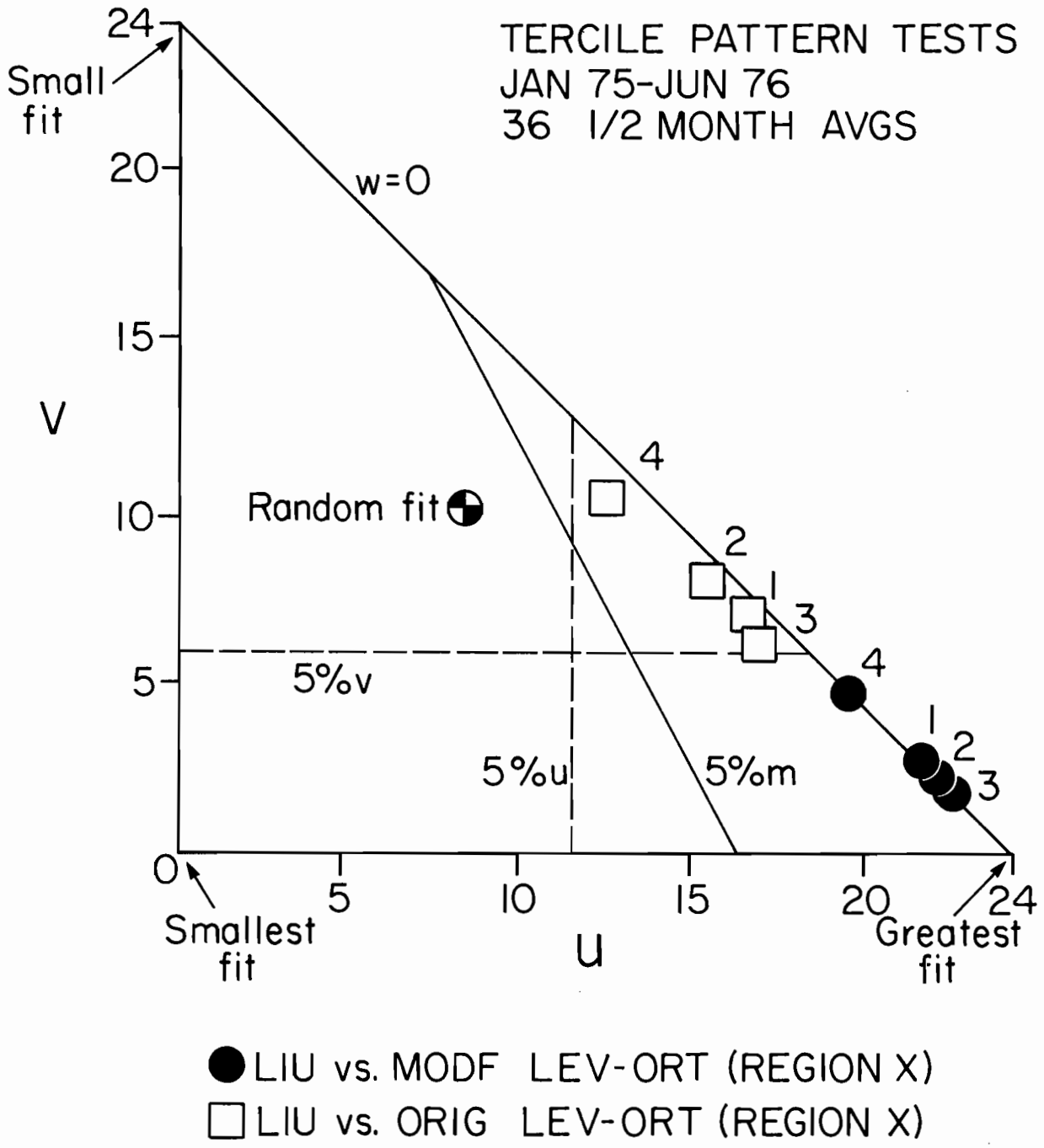

TERCILE PATTERN TESTS
JAN 75-JUN 76
36 1/2 MONTH AVGS
Small fit
w=0
Random fit
5%v
5%u
5%m
V
u
0
5
10
15
20
24
Smallest fit
Greatest fit
4
2
1
3
4
1
2
3
LIU vs. MODF LEV-ORT (REGION X)
LIU vs. ORIG LEV-ORT (REGION X)

Fig. 4.2

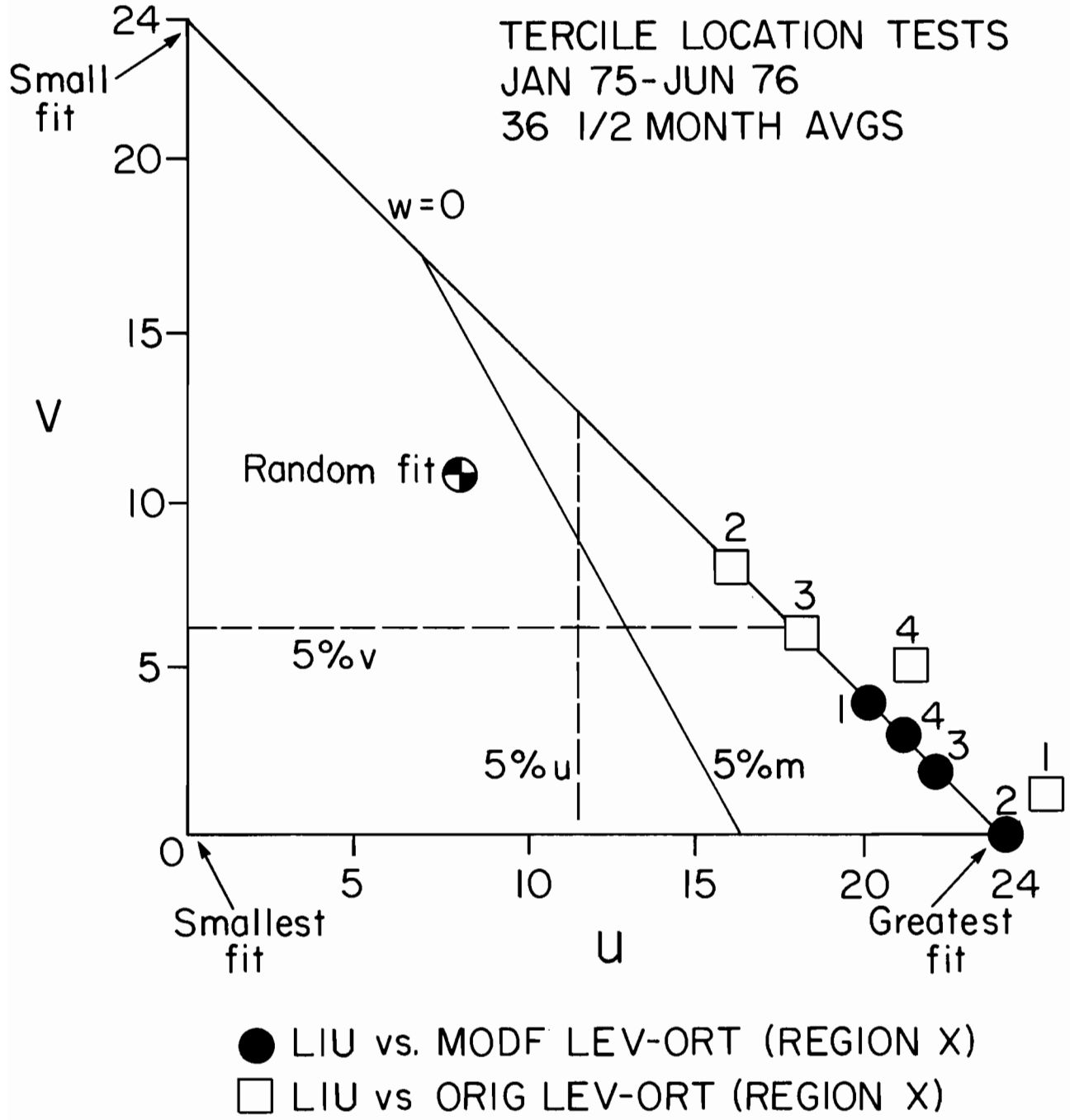

TERCILE LOCATION TESTS
JAN 75-JUN 76
36 1/2 MONTH AVGS
Small fit
w=0
Random fit
5%v
5%u
5%m
V
U
0
5
10
15
20
24
Smallest fit
Greatest fit
LIU vs. MODF LEV-ORT (REGION X)
LIU vs ORIG LEV-ORT (REGION X)

Fig. 4.3

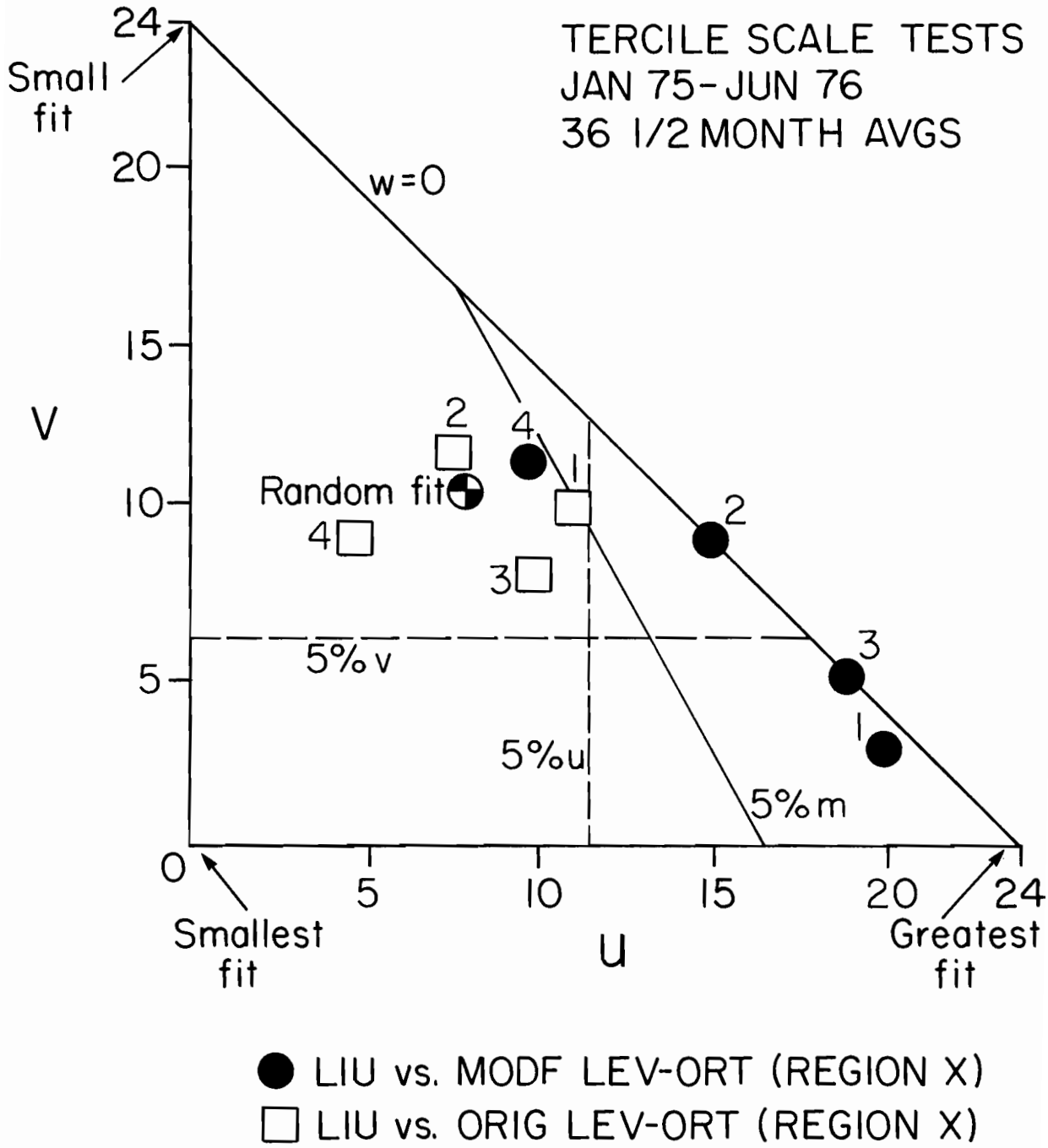

TERCILE SCALE TESTS
JAN 75-JUN 76
36 1/2 MONTH AVGS
Small fit
w=0
Random fit
5%v
5%u
5%m
V
U
0
5
10
15
20
24
Smallest fit
Greatest fit
LIU vs. MODF LEV-ORT (REGION X)
LIU vs. ORIG LEV-ORT (REGION X)

Fig. 4.4

set that is still significantly close to Liu's original set on the 5% level, in the sense of u, v, and m. Looking at Fig. 4.3 we see that the gradient-dependent features of Liu's scheme are not important to the location (overall average) of the data sets because the solid circular points once again crowd around the "perfect match" vertex of the diagram. In Fig. 4.4, where we examine the resultant spread or scale (variance) of the processed data sets, we see that there is a strong density and gradient effect on the spread of the sets. In region 1 where density and SST gradient are high, the Liu and modified Lev-Ort sets are close in the sense of spread (variance) (meaning the removal of SST-gradient features of Liu's scheme did not matter). However, in region 4 they differ in spread to the extent that any two randomly chosen 36×24 data sets would differ.

Notice that in each of the three diagrams of Figs. 4.2, 4.3, 4.4, the set of points for intercomparing Liu and modified Lev-Ort is closer to the "greatest fit" or "perfect match" vertex than the set of Liu vs. original Lev-Ort points. Also, the differences in effect between Liu's scheme and the original Levitus-Ort scheme on *pattern* and *spread* (or variance) of a data set are more pronounced than the differences in effect, between Liu vs. original Lev-Ort on *location*.

In sum, our conclusions concerning the gradient-dependent smoothing operation of Liu's objective analysis scheme, reached via the Trinity statistics SITES and SPRED in §5 of DIT(II), are corroborated by the present Tercile Method of intercomparing data sets. As the reader may verify, these two intercomparison techniques (Trinity and Tercile) are quite different in character. Therefore, the conclusions they agree on in this example may have some basis in reality. A final, detailed look at Liu's objective analysis scheme, and its effects on a raw data set, is made in DIT(V).

C. Recall that Figures 4.2, 4.3, 4.4 plot *time averages* of the u and v statistics. It is sometimes instructive to look at the individual time behaviors of v(t), v(t), and w(t). As an example, consider region 1, over which we intercompare the spatial patterns of Liu vs. original Lev-Ort. The results are displayed in Table 4.2. The average values of the u and v are respectively 16.4 and 7.2. These are the coordinates of open square 1 in Fig. 4.2. A "period" in Table 4.2 is a ½ month period and is denoted e.g. by symbols of the form "1/2/75", meaning the *second half* of January 1975. Note that the standard deviations of the u(t) and v(t) are quite large, indicating considerable variation in *pattern differences* over the 18 month period. In the present case, however, we see that only twice (3/1/75, 7/1/75) did u fall below its 5% significance level 11.5, indicating that, as far as u (0-class error count) goes, the SST patterns of Liu's set and the SST patterns of the original Levitus-Oort set are significantly close 34 out of 36 times over the 18 month period in region 1 of the Pacific. Once again this attests to the non essentiality of the SST-gradient-sensitive and other features in Liu's scheme, regarding *spatial patterns*.

Table 4.2

Time dependence of u,v,w scores over Region 1 while intercomparing Liu and original Lev-Ort SST Patterns

	Period	u	v	w
1	1/1/75	13	11	0
2	1/2/75	13	11	0
3	2/1/75	13	11	0
4	2/2/75	17	7	0
5	3/1/75	11	12	1
6	3/2/75	19	5	0
7	4/1/75	19	5	0
8	4/2/75	22	1	1
9	5/1/75	14	10	0
10	5/2/75	18	6	0
11	6/1/75	13	11	0
12	6/2/75	16	7	1
13	7/1/75	11	11	2
14	7/2/75	21	2	1
15	8/1/75	24	0	0
16	8/2/75	19	5	0
17	9/1/75	21	3	0
18	9/2/75	20	4	0
19	10/1/75	16	6	2
20	10/2/75	19	4	1
21	11/1/75	15	7	2
22	11/2/75	16	8	0
23	12/1/75	15	8	1
24	12/2/75	16	8	0
25	1/1/76	18	5	1
26	1/2/76	16	8	0
27	2/1/76	17	7	0
28	2/2/76	20	4	0
29	3/1/76	12	12	0
30	3/2/76	15	9	0
31	4/1/76	13	11	0
32	4/2/76	20	4	0
33	5/1/76	14	10	0
34	5/2/76	18	6	0
35	6/1/76	13	10	1
36	6/2/76	15	9	0
AVERAGE		16.4	7.2	.4
STD DEV		3.3	3.2	.6

5. Application: Self-predictability of a Data Set

The purpose of this discussion is two-fold. First, we wish to show how the concept of data pattern intercomparison, by the Tercile Method, may be used to generate *and* score a wide variety of possible long-term forecasts of oceanographic variables. Secondly, we introduce the notion of ***self-predictability*** as a possible measure of the information content of a data set. For we anticipate that the ***self-predictability*** of a data set varies ***inversely*** with the ***noise content*** of the set, as seen through the "eyes" of a given forecaster. By "self-predictability" of a data set we mean here the tercile score u (via the Pattern Test, above) between a data set's actual pattern at time $t+\ell$ and its predicted pattern at $t+\ell$ based on a pattern at an earlier time t (using some means of forecasting), over some fixed spatial domain, where ℓ is the ***lead*** between the later and earlier times. Before proceeding, the reader may wish to look over the definitions, given in Appendix A, of the various benchmark and empirical forecasters used in the self-prediction exercises described below.

A. The setting for the first example is the 99 point set in the tropical Pacific as depicted in Fig. 4.1. The points are numbered starting with the first in the southwest corner, and ending with the 99^{th} off the coast of Mexico in the northeast corner. Fig. 5.1 depicts the self-predictability of Liu's data set produced using his full objective analysis scheme as summarized in Table 4.1. We apply to this set the forecasting techniques of three ***empirical forecasters***: the (Most) Probable Markover, the Pure Analoger and the Persistence Analoger. The remaining two forecasters are the so-called ***benchmark forecasters***: the Stochaster and the Persister. The average number of 0-class errors of the Stochaster on the 99 point set is 33, and is shown

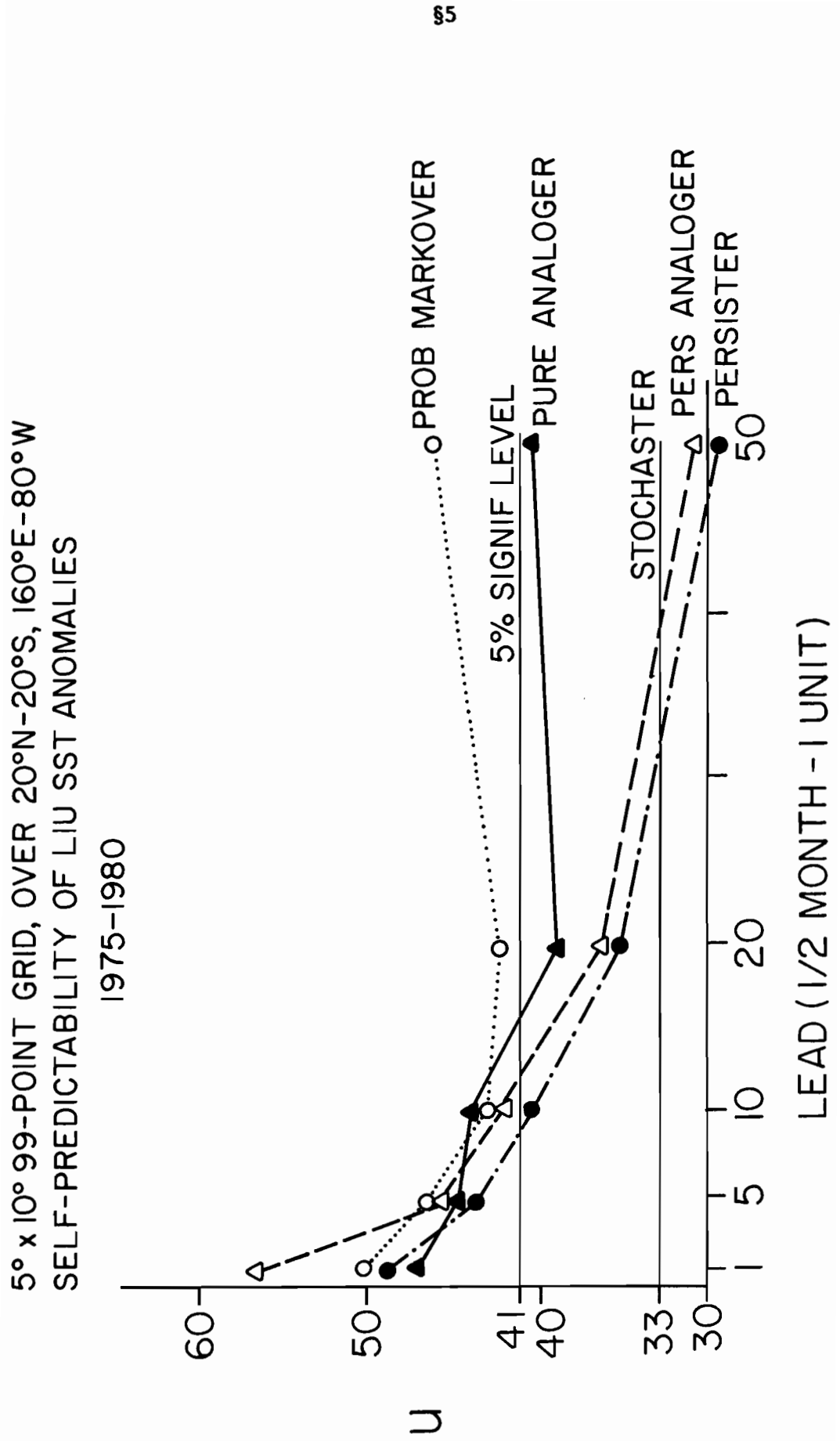

5° x 10° 99-POINT GRID, OVER 20°N-20°S, 160°E-80°W
SELF-PREDICTABILITY OF LIU SST ANOMALIES
1975-1980
PROB MARKOVER
5% SIGNIF LEVEL
PURE ANALOGER
STOCHASTER
PERS ANALOGER
PERSISTER
U
60
50
41
40
33
30
1
5
10
20
50
LEAD (1/2 MONTH - 1 UNIT)

Fig. 5.1

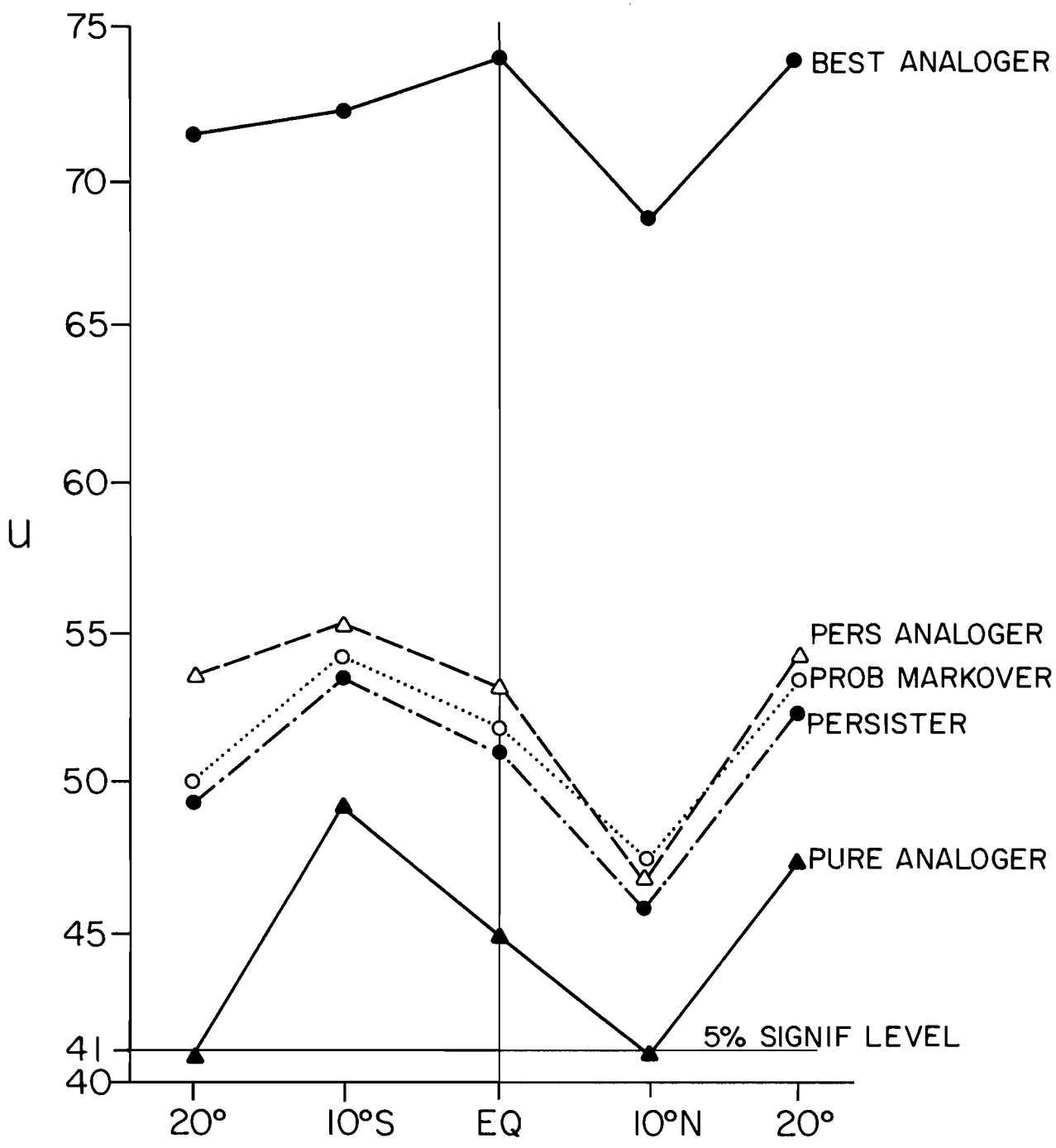
as a horizontal line in the figure. The Stochaster's 5% significance level is shown as a horizontal line at u = 41. Hence pure-chance forecasts will average around u = 33, and only 5% of the time will they produce u values of 41 or more.

We shall describe how a typical forecast went for the Persister. We started with a full sequence of 144 maps of SST over the 99 point set from January 1975, to December 1980. Each map was a ½ month average of SST over the 99 point set. We then applied the Pattern Test of the S-Synoptic method to the first and second maps of the sequence, found and then recorded the associated 0-class error count u. We then went to the second and third maps of the sequence, found and then recorded the associated u; and so on for all 143 possible ½-month-separated maps. We averaged the u scores, and found u = 49, and this became the ordinate of the Persister curve at ½ month lead time. This process was repeated with the Persister for various other lead times out to 50 ½-months, and the Pattern Test u scores averaged and plotted. For the other forecasters we applied their definitions in turn to each of the 144 maps, according as their definitions allowed, and plotted the average u scores for each lead out to 50 ½-months. We shall now describe some of the results.

The Persister begins with a very skillful forecast at a lead of ½ month, earning 49 0-class errors out of a possible 99. At the end of 10 ½-month leads, i.e., 5 months, the Persister's forecast is just below the 5% significance level relative to the Stochaster. The Persistence Analoger starts at ½ month lead with an excellent u = 57 out of a possible 99, and drops to the 5% level in about 6 months. Standing back and considering these two performances, we are somewhat surprised at the high skill of these self-forecasts, especially in view of our experience with season-ahead forecasts over 99-point sets of the U.S. Mainland. (This will be illustrated

later in Fig. 5.5.) For the moment we continue to look at these unexpectedly high self-predictions of the tropical sea surface temperature set, particularly through the "eyes" of the Pure Analoger and the (Most) Probable Markover. We see from Fig. 5.1 that the Pure Analoger has significant forecasts out to about a year, while the skill of the Probable Markover never falls below the 5% level for the duration of this particular set of forecasts (25 months).

B. The settings of the second example, shown in Fig. 5.2, are five parallels of latitude: 20°S, 10°S, equator, 10°N, 20°N. For each of these parallels, from 160°E to 80°W, we chose 99 points on or near each of these parallels and made a series of forecasts for ½ month leads. For example we applied the Pure Analoger to the task of forecasting ½ month ahead the SST along the 99 points of 20°S. The total record consisted of 144 such ½ month averaged SST snapshots from January 1975 to December 1980. Therefore the Pure Analoger made 143 forecasts of ½ month lead time at 20°S. The average u-score was 40.8--just below significance at the 5% level. However, at 10°S the average forecast score of the Pure Analoger soared to u = 49 and dropped to 45 at the equator. It reached a comparable low at 10°N to that at 20°S, and then rose again to about u = 47 at 20°N. The performances of the remaining forecasters are as shown in Fig. 5.2. They all (except for the Best Analoger, who is in a class by himself) exhibit the maximum self-predictability at 10°S and a local minimum at 10°N. While we will not conjecture here about the physical mechanisms behind these similarities in latitudinal behavior of self-predictability (and hence noise or information content of the SST field in the tropical Pacific), it is clear that the N-shaped curves in Fig. 5.2 are strongly influenced by the equatorial current and its counter currents.

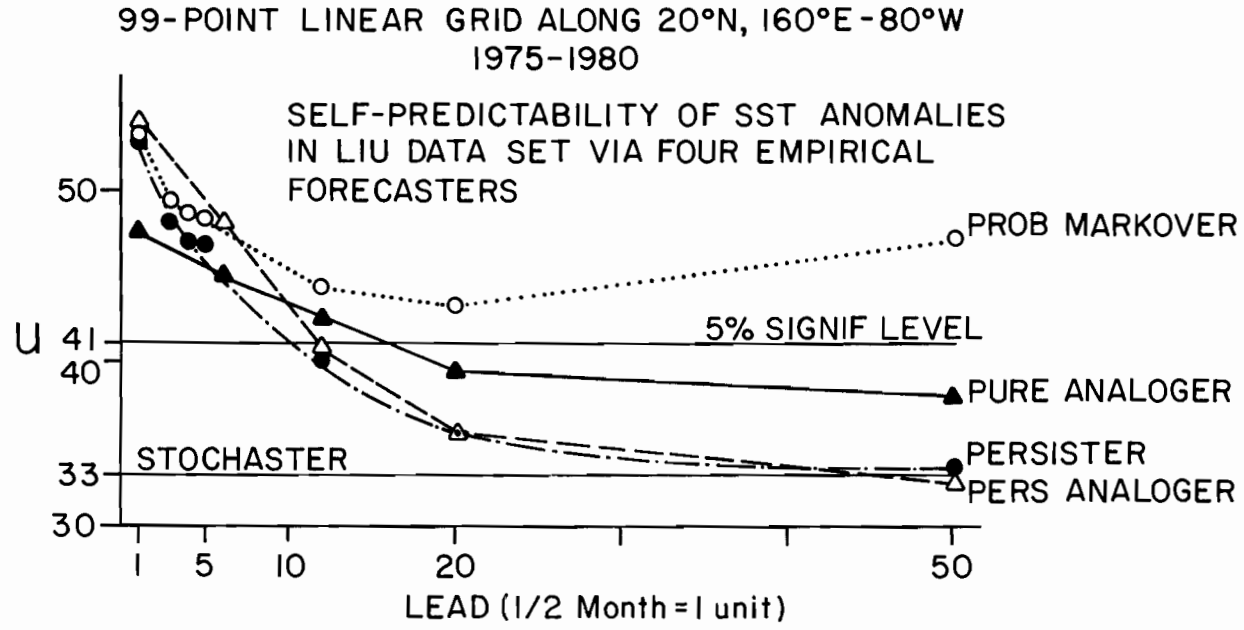


SELF-PREDICTABILITY OF LIU SST ANOMALIES,
JAN 75-DEC 80
144 1/2 MONTH AVGS
1/2 MONTH LEAD, ON 99-POINT GRIDS ALONG LATITUDES
SHOWN, FROM 160°E TO 80°W

Fig. 5.2

C. In the third example, shown in Fig. 5.3, we return to the setting of Fig. 5.2, in particular the 99 point grids along latitude parallels 10°N and 20°N. We were interested in the self-predictability of the SST set along these parallels as a function of lead time. Recall that the curves of Fig. 5.2 were based on a fixed ½ month lead time. Hence the curves in Figs. 5.2, 5.3 at lead time ½ month should match. The u scores were found analogously to those in Fig. 5.1. Then, as we step off along the lead-time axis, the u scores begin to drop. The main thing we were interested in was the relative occurrence of forecast lead times at which the curves fall below the Stochaster's 5% u-significance level. On the basis of Fig. 5.2, we anticipated that these lead times would be greater at 20°N than at 10°N. This is generally the case. Thus the Persister's forecasts along 20°N drop below the 5% level at 5 months lead time, while along 10°N they drop below at 4 months. The Pure Analoger's forecast curve becomes non significant along 20°N at 8 months, while along 10°N it is never quite significant. The Probable Markover, apparently extracting maximal information from the data set, manages to stay significant out to 25 months. The slight rise in his curve is quite interesting and, at this writing, we have not tested this behavior for plausibility.

D. In the fourth example, shown in Fig. 5.4, we return to the 99 point region of Figs. 5.1 and 4.1 and examine the effect on forecasts of averaging the data sets over time periods larger than ½ month. We selected for study averaging-time-periods of 1, 2, and 3 months. Other than this variable averaging time, the forecasts were made generally as those in Fig. 5.1. We therefore expect the curves to match at the ½ month lead time. Two extra forecasters were included in the experiment, namely the Climater and the Best


99-POINT LINEAR GRID ALONG 20°N, 160°E-80°W
1975-1980
SELF-PREDICTABILITY OF SST ANOMALIES
IN LIU DATA SET VIA FOUR EMPIRICAL
FORECASTERS
PROB MARKOVER
5% SIGNIF LEVEL
PURE ANALOGER
PERSISTER
PERS ANALOGER
STOCHASTER
U
50
41
40
33
30
1
5
10
20
50
LEAD (1/2 Month = 1 unit)

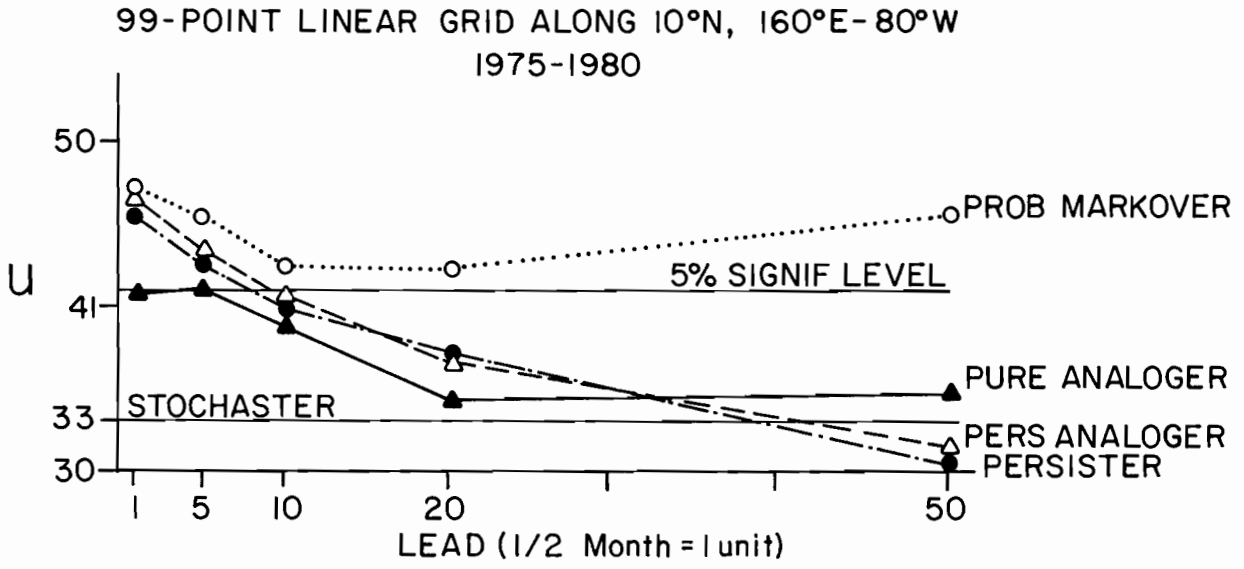

99-POINT LINEAR GRID ALONG 10°N, 160°E-80°W
1975-1980
PROB MARKOVER
5% SIGNIF LEVEL
PURE ANALOGER
PERS ANALOGER
PERSISTER
STOCHASTER
U
50
41
33
30
1
5
10
20
50
LEAD (1/2 Month = 1 unit)

Fig. 5.3

Analoger. We were particularly interested in the forecast skills at the three month (1 season) lead time.

It is noteworthy that in no case did the forecast skill diminish as we went from ½ month to 3 months lead time. Indeed, the Persistence Analoger, Probable Markover, Persister, and Pure Analoger end up with somewhat higher forecast skill scores at the season-ahead lead time than at the ½ month lead time. There are definite questions for further research that these results and those of the preceding three examples raise, questions that have implications for the season-ahead climate forecasts over the U.S. Mainland and other regions of the planet.

D. We conclude these examples of the use of the Tercile Method of data intercomparison by plotting in Fig. 5.5 the uv coordinates of the the five principal forecasters of Fig. 5.4 (those above the 5% significance level). We also added the performance of the Empirical Markover. These plotted points are shown as solid circles in Fig. 5.5. It so happened that in a time interval (1974-1982) containing this same period, 1975-1980, we had verified the average season-ahead forecasts, over the U.S. Mainland, of these same empirical forecasters.* In this earlier study, the U.S. Mainland was also laid-over with a 99-point grid; hence both sets of forecasts are comparable. In Fig. 5.5, the horizontal lines are of constant v, and the lines† with slope -2 are lines of constant moment m = v+2w. The results are remarkable: we see that, on the whole, the forecast skills of the empirical forecasters, (in

---

* See Preisendorfer and Mobley (1982) §15.

† The equations of the constant-m lines are of the form v = -2u + (2p-m). In Fig. 5.5 we have p = 99.

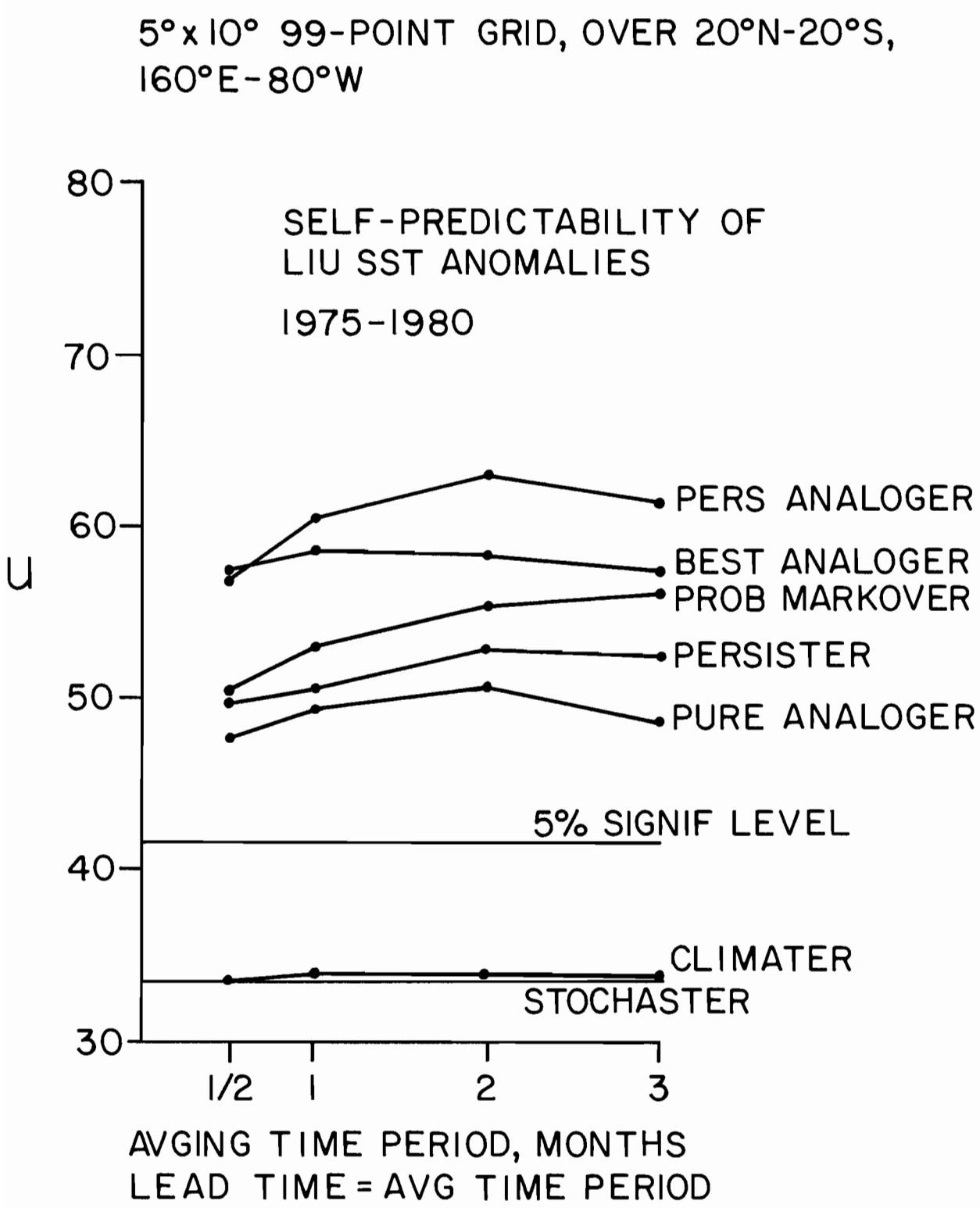

5°x 10° 99-POINT GRID, OVER 20°N-20°S,
160°E-80°W
SELF-PREDICTABILITY OF
LIU SST ANOMALIES
1975-1980
80
70
60
50
40
30
U
PERS ANALOGER
BEST ANALOGER
PROB MARKOVER
PERSISTER
PURE ANALOGER
5% SIGNIF LEVEL
CLIMATER
STOCHASTER
1/2
1
2
3
AVGING TIME PERIOD, MONTHS
LEAD TIME = AVG TIME PERIOD

Fig. 5.4

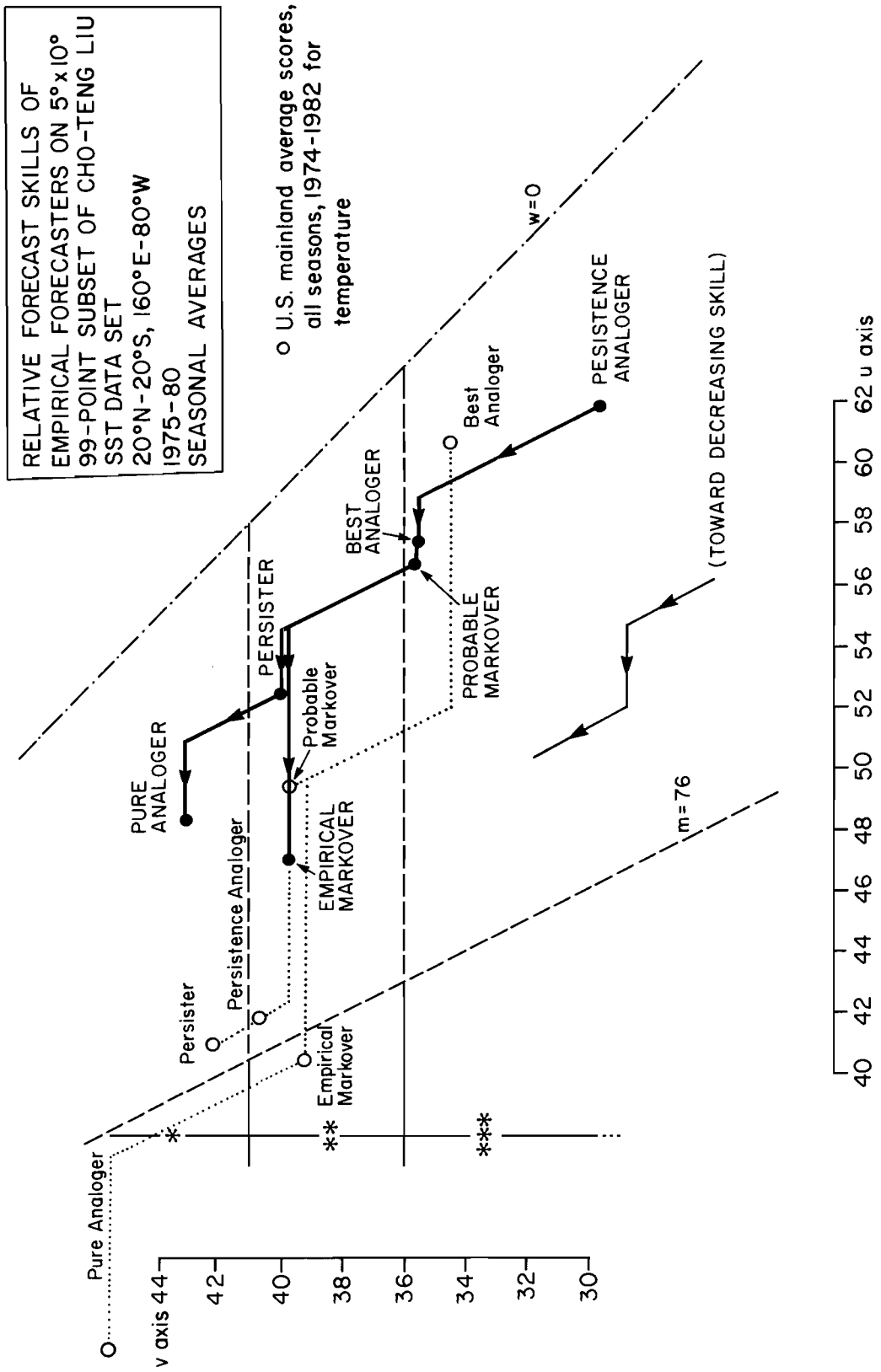

RELATIVE FORECAST SKILLS OF
EMPIRICAL FORECASTERS ON 5°x10°
99-POINT SUBSET OF CHO-TENG LIU
SST DATA SET
20°N-20°S, 160°E-80°W
1975-80
SEASONAL AVERAGES
O U.S. mainland average scores,
all seasons, 1974-1982 for
temperature
w=0
PESISTENCE
ANALOGER
Best
Analoger
BEST
ANALOGER
PERSISTER
Probable
Markover
PROBABLE
MARKOVER
PURE
ANALOGER
Persistence Analoger
EMPIRICAL
MARKOVER
Persister
Empirical
Markover
Pure Analoger
m=76
(TOWARD DECREASING SKILL)
v axis 44 42 40 38 36 34 32 30
40 42 44 46 48 50 52 54 56 58 60 62 u axis

Fig. 5.5

these two separate sets of forecasts), are almost co-extensive over the uv diagram. *The empirical forecasters on the whole are nearly as skillful at season-ahead forecasts over the U.S. Mainland as they are at season-ahead forecasts over the present Tropical Pacific domain, a region about five times the area of the U.S. Mainland.* On closer examination of the diagram, forecaster by forecaster, we may be led to expect the skills of the empirical forecasters over the tropical Pacific to be slightly superior to their temperature forecast skills on the U.S. Mainland. Future studies of this matter may be of interest to season-ahead forecast efforts.

6. Application: A Principal-Component Selection Rule

Adopting the assumption that a relatively high-u self-predictability index of an n×p data set $\underline{D}'$ is associated in some way to a relatively low noise content of the data set, we can explore the possibility of a principal component selection rule inherent in this assumption. As shown in Appendix B of DIT(III), we can represent the space-centered values d(t,x) of $\underline{D}'$ in principal component form:

$$d(t,x) = \sum_{j=1}^{p} a_j(t)e_j(x)$$

$$= \sum_{j=1}^{p} b_j(t,x) \qquad (6.1)$$

where we have written, for j = 1,···,p,

$$'b_j(t,x)' \quad \text{for} \quad a_j(t)e_j(x) \tag{6.2}$$

If we explicitly display the associated eigenvalues, we can write

$$\begin{aligned} b_j(t,x) &= \ell_j^{\frac{1}{2}}\alpha_j(t)e_j(x) \\ &\equiv \ell_j^{\frac{1}{2}}\beta_j(t,x) \end{aligned} \tag{6.3}$$

For each fixed j, we now may consider $\underline{B}_j \equiv \{b_j(t,x)\colon t = 1,\cdots,n;$ $x = 1,\cdots,p\}$, $j = 1,\cdots,p$, as a data set in its own right and subject it to an S-Synoptic Tercile Pattern Test to determine (in the manner of §5) its self-predictability. Thus, we would use $\underline{B}_j$ to determine the class boundaries $a_d(x)$, $b_d(x)$ as in (2.7), (2.8). Then we would select a fixed lead time ℓ of interest (say one season) and find for $\underline{b}_j(t) \equiv [b_j(t,1),\cdots,b_j(t,p)]^T$ and $\underline{b}_j(t+1)$ the associated values $u_j(t)$, $v_j(t)$, $w_j(t)$ and $m_j(t)$, $t = 1,\cdots,n$. To keep matters simple we would at first work only with the time averages of $u_j(t)$, $v_j(t)$ and plot the results (u_j,v_j), $j = 1,\cdots,p$ in a uv diagram, after the manner of Fig. 5.5. The result would be a set of p points in the uv diagram. These points can be partially ordered, as were the points on Fig. 5.5 (see (A3.4) or Preisendorfer and Mobley, 1982, §7). The simplest form of selection rule would be: *choose those points* (u_j,v_j) *of maximal forecast skill and that lie in the 5% u-significance region of the uv diagram.* If the j indexes of (u_j,v_j), so selected, comprise the integer set J, then the reduced form of (6.1) becomes

$$\begin{aligned} \hat{d}(t,x) &\equiv \sum_{j\varepsilon J} b_j(t,x) \\ &= \sum_{j\varepsilon J} a_j(t)e_j(x) \end{aligned} \tag{6.4}$$

Presumably the $\hat{d}(t,x)$ would make up a new data set of enhanced self-predictability and therefore of reduced noise content.

Now that the idea has been broached, several alternative selection rules come to mind. Suppose (for some fixed lead time ℓ) that instead of evaluating the self-predictive skill of the individual maps, $\underline{b}_j(t)$, $t = 1,\cdots,n$, we would evaluate the skill of various additive combinations

$$b(K|t,x) \equiv \sum_{j\varepsilon K} b_j(t,x), \quad t = 1,\cdots,n;\ x = 1,\cdots,p \qquad (6.5)$$

of the $\underline{b}_j(t)$, where K is any non empty subset of $\{1,\cdots,p\}$. Thus $\underline{B}(K) = \{b(K|t,x)\colon t = 1,\cdots,n;\ x = 1,\cdots,p\}$ is an n×p data set in its own right and, just as $\underline{B}_j$ above, it can be examined for self-predictive skill. The number of non empty subsets K of $\{1,\cdots p\}$ is 2^p-1 and therefore, for the usual range of p(10 to 100), very large. Efficient ways of examining these collections of the $\underline{B}(K)$ are called for. A random search for the K's may be the basis for their selection. It may simplify the search if it is limited to those subsets of indexes j in (6.5) for which the sum of the ℓ_j is (say) at least 80% of the total sum $\sum_{j=1}^{p} \ell_j$.

The class of selection rules that is forming, especially with the latter considerations of including the ℓ_j's in the decision process, is quite different from the dominant-variance and time-history rules developed in Preisendorfer, Zwiers and Barnett (1981). The present rules attempt to link the dimensional reduction of the given n×p data set $\underline{D}'$ to a *self-predictability* property of the set rather than to the purely *variance* or *correlational* properties of the principal components $a_j(t)$, $j = 1,\cdots,p$; $t = 1,\cdots,n$. We shall leave this matter here for possible future study.

7. References

Works in this series on Data Intercomparison Theory (NOAA Technical Memorandums, ERL-PMEL):

DIT(I): Minimal Spanning Tree Tests for Location and Scale Differences.

DIT(II): Trinity Statistics for Location, Spread and Pattern Differences.

DIT(III): S-Phase and T-Phase Tests for Spatial Pattern and Temporal Evolution Differences.

DIT(IV): Tercile Tests for Location, Spread and Pattern Differences.

DIT(V): Case Study: Effects of Objective Analysis on a Tropical Pacific Sea Surface Temperature Set.

Cressman, G. (1959) "An Operational Objective Analysis System," Mon. Wea. Rev. 87, 367.

Levitus, S., A. Oort (1977) "Global Analysis of Oceanographic Data," Bull. Am. Met. Soc. 58, 1270.

Liu, Cho-Teng (1982) Analysis of Tropical Pacific Sea Surface Temperatures for 1975 to 1980. Pacific Marine Environmental Laboratory, Seattle, Wash., NOAA Tech. Memo. ERL PMEL-34.

Preisendorfer, R. W. (1977) Climate Forecast Verification via Multinomial Stochasters. SIO Ref. Series 77-33, Scripps Institution of Oceanography, La Jolla, Cal.

Preisendorfer, R. W., F. W. Zwiers, T. P. Barnett (1981) Foundations of Principal Component Selection Rules. SIO Ref. Series 81-4. Scripps Institution of Oceanography, University of California, La Jolla, CA 92093.

Preisendorfer, R. W., C. D. Mobley (1982) Climate Forecast Verification, U.S. Mainland, 1974-1982. Pacific Marine Environmental Laboratory, Seattle, WA. NOAA Tech. Memo. ERL PMEL-36.

APPENDIX A

Trinomial Stochaster

1. Introduction

The applications of the Tercile Method, in §4 of the main text, draw on the theory of the Trinomial Stochaster, which we will briefly review here. In §2 below we shall outline the Stochaster's theory and in §3 we will list and define the empirical forecasters. The reader wishing more detail on the Stochaster and empirical forecasters may consult Preisendorfer (1977) and Preisendorfer and Mobley (1982).

2. Trinomial Stochaster

Consider p points in space and three terciles of values of some field at each point x. Thus, at each point x a range of n possible values of a given field has been divided into three equally populous cells A, N, and B, as in §2A of the main text. It is immaterial to the Stochaster defined below how these terciles at each point were constructed; just that they exist.

The *Trinomial Stochaster* generates a random field over the p points, as follows. At point x = 1, he randomly flips an unbiased cube with opposite faces marked with "A", "N", and "B". Thus with equal probability, namely 1/3, an A, N, or B shows up. He notes the outcome and assigns it to point x = 1. In general, at the x^{th} point, he assigns an A, N, or B with probability 1/3, and in this way produces a random field in tercile form. He then passes over the p points once again and generates a second random field, statistically independent of the first. Next, a tally is made of the number u of 0-class errors (matches) between the values of the two fields. At the same time the v

1-class and w 2-class errors are also tallied over the p values (recall Fig. 2.2). Clearly, for these u,v,w scores, we must have

$$u + v + w = p \qquad \text{(A2.1)}$$

Moreover, since the fields were independently produced, the probability of a match (i.e., a 0-class error) occurring in any A or N or B cell at a given x is $(1/3)^2 = 1/9$. The schematic diagram below shows more generally that 1/9 is the probability of any one of the nine types of occurrence of the two fields at each x:

	A	N	B
A	1/9	1/9	1/9
N	1/9	1/9	1/9
B	1/9	1/9	1/9

tercile value of first field at x

tercile value of second field at x

Thus the three possible types of match for the first and second fields are AA, NN, BB - down the main diagonal, and each is of probability 1/9. Therefore the probability a_0 of a 0-class error at any x is the sum of these probabilities, namely 1/3. Moreover, the probability a_1 of a 1-class error is obtained by summing all four of the 1/9 fractions occurring just off the main diagonal of the above diagram. Thus $a_1 = 4/9$. Finally, the probability a_2 of a 2-class error between two randomly produced fields at x is obtained by summing up the two $1/9^{ths}$ at the AB, BA boxes: $a_2 = 2/9$.

APPENDIX A

Having determined the elementary probabilities a_0, a_1, a_2 of occurrences of 0-, 1-, and 2-class errors at each x, we can now compute the probabilities of u, v, w 0-, 1-, 1-class errors individually or jointly as subtended by two randomly produced fields with terciled values on a set of p points. The probability $p_o(u)$ of u matches of two fields out of a possible p matches is

$$p_o(u) = \binom{p}{u} a_o^u (1-a_o)^{p-u} , \qquad a_o = 1/3 \tag{A2.2}$$

where

$$\binom{p}{u} = \frac{p!}{u!(p-u)!}$$

In a similar manner we have the probability of v 1-class errors between the two random fields as

$$p_1(v) = \binom{p}{v} a_1^v (1-a_1)^{p-v} , \qquad a_1 = 4/9 \tag{A2.3}$$

and for 2-class errors:

$$p_2(w) = \binom{p}{w} a_2^w (1-a_2)^{p-w} , \qquad a_2 = 2/9 \tag{A2.4}$$

The joint probability of u,v,w 0-, 1-, 2-class errors between the two randomly produced fields is:

$$p(u,v,w) = \frac{p!}{u!v!w!} a_0^u a_1^v a_2^w \tag{A2.5}$$

where

$$a_0 + a_1 + a_2 = 1, \quad u + v + w = p$$

By summing p(u,v,w) over all possible values of v and w, for example, we return to $p_o(u)$ of (A2.2).

3. Display of Scores: The uv Diagram

The tercile intercomparison of two maps in p-space results in a triple of scores u,v,w. Since any two of these are independent (cf A2.1) we may pick, say, u and v and plot a point with these coordinates in the uv plane. Since u,v are each bounded by 0 and p, and since (A2.1) holds, the region of possible (u,v) points is a triangle as shown in Figs. 4.2, 4.3, 4.4. The diagrams are drawn for the case p = 24. The average score of the Stochaster is at (u,v) = (8,10.7) = (a_0p, a_1p), and is denoted by "random fit" in the figures. From (A2.2), (A2.3), (A2.4), we can find the 5% significance levels for u,v,w and these are, respectively, 11.5, 6.1, 1.6. The 11.5 was found, e.g.*, by summing $p_o(u)$ from u = 0 up to the u which gave a sum of 0.95, or just above. The u,v, 5% significance levels are included in Figs. 4.2, 4.3, 4.4, and help in discerning significantly close fits between fields. The first indication of a significantly close fit is given by a u score exceeding 11.5; the second is given by a v score being less than 6.1; and finally by a w score being less than 1.6. Having all three scores significant produces a joint significance level which can be calculated if desired,† from (A2.5).

Another useful measure of significantly close fit is given by the *moment*

$$m = v + 2w \qquad \text{(A3.1)}$$

---

* The 5% significance level 41 for u in the case p = 99 (as used e.g. in §5 of the main text) is found in exactly the same way. The Stochaster's average u for p = 99, is $a_op = 99/3 = 33$.

† See, e.g., Preisendorfer (1977) for an example of Extensive tables (cf Table A there) that can be compiled, and which can be used to find the probability content of any subset of the uv diagram.

APPENDIX A

Under moderately large p values the m of the Stochaster is distributed nearly normally. The average value $\bar{m}$ and variance σ_m^2 of m are given rigorously by

$$\bar{m} = p(a_1 + 2a_2) = 24 \cdot (2/9) = 21.33 \qquad \text{(A3.1)}$$

$$\sigma_m^2 = p\,[(a_1+4a_2) - (a_1+2a_2)^2] = (24) \cdot (0.5432) = 13.03 \qquad \text{(A3.2)}$$

The 5% critical level for m is therefore approximately* (with $\sigma_m = 3.610$):

$$m_c = \bar{m} - 1.64\ \sigma_m = 15.41 \qquad \text{(A3.31)}$$

The smaller m is, the better the fit between the two fields. For m values below 15.41, we have significance on the 5% level. The line m = 15.41 is drawn in the Figures and labeled "5%m" for the case p = 24.

One may see the systematic use of m, in scoring and ranking intercomparisons, by consulting the references in the Introduction to this Appendix. In fact, in ranking fits, we tend to favor the use of v and m. Thus if $f_1 = (u_1, v_1, w_1)$ and $f_2 = (u_2, v_2, w_2)$ are the tercile scores of two data fits and $m_1 = v_1+2w_1$, $m_2 = v_2+2w_2$, we would write

$$f_2 \leqq f_1 \qquad \text{(A3.4)}$$

if and only if $m_1 \leqq m_2$ and $v_1 \leqq v_2$. We would read (A3.4) as saying: "data fit f_1 is greater than or equal to data fit f_2". This ordering definition satisfies the postulates of a partial order (reflexivity, symmetry, transitivity).

---

* when p = 99, $m_c = 76$, see Fig. 5.5.

4. The Benchmark and Empirical Forecasters

We introduce the present forecasters with the idea in mind of using them for *data intercomparison* purposes. Thus, as shown in the main text above, these forecasters can have their roles extended to become also *intercomparators*. We shall, however, retain their "forecaster" name for simplicity.

The *benchmark forecasters* are defined as follows:

Climater: always predicts normal

Persister: always predicts present

Stochaster: always predicts randomly

The Stochaster was defined in §2 of this Appendix. In the context of tercile representations of data fields, the other two benchmark forecasters may be further described this way:

The *Climater* never makes a calculated try at forecasting, and therefore never tries to predict above or below normal. Verification uv scoring diagrams for the Climater show the characteristic tell-tale pattern of his forecasts: all his verification points have w = 0.

The *Persister* is the lazy cousin of the Climater. Whatever pattern now exists over the map, the Persister says *that* will be the pattern of the next period, e.g. the next season.

The *empirical forecasters* are defined here in the context of tercile representations of data fields.

In the definitions below, it will be helpful to think of the data maps laid out on a table, in a straight line, from earliest on the left to the latest on the right. In each definition, one of these maps will be singled out as the *present* or *current season map*, and the *forecast* will be for the successor just to its right, i.e., the *next season map*. Here, of course

"season" is used simply to fix ideas. In Table 4.2 of the main text, e.g., we work with periods of time ½ month long. Therefore, the following definitions are quite generally applicable.

Pure Analoger: Go through the entire record and find that map of the record which has the smallest m-value (m = v + 2w) with respect to the current season's map. This is the *analog* of the current season's map. The forecast for the next season is the seasonal map succeeding the analog's map.

Persistence Analoger: Proceed, as in the case of the Pure Analoger, to find the current season's analog. The forecast for the next season is the map of the analog.

Best Analoger (non operable): Single out an arbitrary map of the record, and designate it as the *current season map*. Then go to the next season map. (This is where the procedure *in practice* becomes non-operable.) Find the analog of the next season's map, i.e., that map of the record which is closest to the next season map in the sense of the m-metric. The forecast for the next season map is this analog of the next season map. This produces the best possible pure or persistence analog forecast that can be made for the next season map, as contained in the record, and using the m-metric.

Empirical Markover: Fix attention on any one of the p points on some arbitrary current map. Go through the record and build the 3×3 table, at the point, consisting of relative frequencies of transition from any

given one of the 3 states A, N, B, in the current season, to any other given one of these states in the next season. (The rows of the matrices are thus probabilities, adding up to 1.) Repeat this at each of the p points, resulting in p, 3×3 transition probability matrices. To forecast the next season's map, use the present season's A, N, or B state and the 3×3 matrix at each point. Draw a random number in a suitable manner to determine which final state the initial A, N, or B state will go to, in accordance with the transition probability. Repeat this at each of the p points. The result is a map which is a realization of a random variable.

(Most) Probable Markover: Establish the 3×3 matrices as above. At a given point, currently in state A, N, or B, find the maximum entry in the associated row of the transition matrix. This maximum entry gives the most probable final state. Use this most probable final state to forecast next season's anomaly state from the present season's observed anomaly state at the given point. Repeat this at each of the p points. The result is a deterministic map. (In the event of a tie, i.e., two or more maximal row elements, at a point, make a prediction using the Climater.

In the experiments summarized in §5 of the main text the empirical forecasters were allowed to range over the entire record of 144 maps in order to produce their analogs or their 3×3 transition probability tables, as the case may be.

NOAA ERL technical reports, technical memoranda, and data reports published by authors at Pacific Marine Environmental Laboratory in Seattle, Washington, are listed below. Microfiche copies are available from the USDOC, National Technical Information Service (NTIS), 5285 Port Royal Road, Springfield, Virginia 22161 (703-487-4650). Hard copies of some of these publications are available from the ERL Library in Boulder, Colorado (303-497-3271). Hard copies of some of the technical reports are sold by the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402 (202-275-9251).

NOAA Technical Report Series

ERL82-POL1 Naugler, Frederic P. (1968)
Bathymetry of a region (PORL-421-2) North of the Hawaiian Ridge, pre-NTIS.

ERL93-POL2 Grim, Paul J. (1968)
Seamap deep-sea channel, Jan. 1969, 2 824 50 060, pre-NTIS.

ERL118-POL3 Le Mehaute, Bernard (1969)
An introduction to hydrodynamics and water waves, 2 vols. 725 pp.
NTIS: PB192 065, PB192 066.

ERL146-POL4 Rea, David K. (1970)
Bathymetry and magnetics of a region (POL-421-3) 29° to 35°N, 155° to 165°W.
NTIS: COM-71-00173.

ERL191-POL5 Reed, R.K. (1970)
Results from some parachute drogue measurements in the central North Pacific Ocean, 1961-1962, 9 pp.
NTIS: COM-71-50020.

ERL214-POL6 Lucas, William H. (1971)
Gravity anomalies and their relation to major tectonic features in the North Central Pacific, 19 pp.
NTIS: COM-71-50409.

ERL229-POL7 Halpern, David (1972)
Current meter observations in Massachusetts Bay, 36 pp.
NTIS: AD-745 465.

ERL230-POL8 Lucas, William H. (1972)
South Pacific RP-7-SU-71 Pago Pago to Callao to Seattle.
NTIS: COM-72-50454.

ERL231-POL9 Halpern, David (1972)
Description of an experimental investigation on the the response of the upper ocean to variable winds, 51 pp.
NTIS: COM-72-50452.

ERL232-POL10 Stevens, H. R., Jr. (1972)
RP-1-OC-71 Northeast Pacific geophysical survey, 91 pp.
NTIS: COM-72-50677.

ERL234-POL11 Lucas, William H. (1972)
Juan de Fuca Ridge and Sovanco fracture zone.
RP-5-OC-71, 39 pp.
NTIS: COM-72-50854.

ERL240-POL12 Halpern, David (1972)
Wind recorder, current meter and thermistor chain measurements in the northeast Pacific-August/September 1971, 37 pp.
NTIS: COM-73-50107.

ERL247-POL13 Cannon, G. A. and Norman P. Laird (1972)
Observations of currents and water properties in Puget Sound, 1972, 42 pp.
NTIS: COM-73-50402.

ERL252-POL14 Cannon, G. A., N. P. Laird, T. V. Ryan (1973)
Currents observed in Juan de Fuca submarine canyon and vicinity, 1971. 57 pp.
NTIS: COM-73-50401.

ERL258-POL15 Lucas, William H., and Richard R. Uhlhorn (1973)
Bathymetric and magnetic data from the northeast Pacific 40° to 58°N, 125° to 160°W. 9 pp.
NTIS: COM-73-50577.

ERL259-POL16 Ryan, T. V., N. P. Laird, G. A. Cannon (1973)
RP-6-OC-71 Data Report: Oceanographic conditions off the Washington coast, October-November 1971, 43 pp.
NTIS: COM-73-50922

ERL260-POL17 Cannon, GIenn A. (1973)
Observations of currents in Puget Sound, 1970, 77 pp.
NTIS: COM-73-50666/9.

ERL261-POL18 Stevens, H. R. Jr., (1973)
RP-1-OC-70 Southeast Pacific geophysical survey, 60 pp.
NTIS: not available.

ERL271-POL19 Reed, Ronald K., and David Halpern (1973)
STD observations in the northeast Pacific, September-October 1972, 58 pp.
NTIS: COM-73-50923/4.

ERL292-PMEL20 Reed, R. K. (1973)
Distribution and variation of physical properties along the SEAMAP standard section, 16 pp.
NTIS: COM-74-50334/3.

ERL323-PMEL21 Erickson, B. H. (1975)
Nazca plate program of the international decade of ocean exploration--OCEANOGRAPHER Cruise-RP 2-OC-73, 78 pp.
NTIS: COM-7540911/6.

ERL325-PMEL22 Halpern, D., J. M. Helseth, J. R. Holbrook, and R. M. Reynolds (1975)
Surface wave height measurements made near the Oregon coast during August 1972, and July and August 1973, 168 pp.
NTIS: COM-75-10900/9.

ERL327-PMEL23 Laird, N. P., and Jerry A. Galt (1975)
Observations of currents and water properties in Puget Sound, 1973, 141 pp.
NTIS: COM-73-50666/9.

ERL333-PMEL24 Schumacher, J. D., and R. M. Reynolds (1975)
STD, current meter, and drogue observations in Rosario Strait, January-March 1974, 212 pp.
NTIS: COM-75-11391/0.

ERL339-PMEL25 Galt, J. A. (1975)
Development of a simplified diagnostic model for interpretation of oceanographic data.
NTIS: PB-247 357/7.

ERL352-PMEL26 Reed, R. K., (1975)
An evaluation of formulas for estimating clear-sky insolation over the ocean, 25 pp.
NTIS: PB-253 055/8.

ERL384-PMEL27 Garwood, Roland (1977)
A general model of the ocean mixed layer using a two-component turbulent kinetic energy budget with mean turbulent field closure, 81 pp.
NTIS: PB-265 434/1.

ERL390-PMEL28 Hayes, S. P., and W. Zenk (1977)
Observations of the Antarctic Polar Front by a moored array during FDRAKE-76, 47 pp.
NTIS: PB-281 460/6.

ERL390-PMEL29 Hayes, S. P., and W. Zenk (1977)
Observations of the Antarctic Polar Front by a moored array during FDRAKE-76, 49 pp.
NTIS: PB-281 460/6.

ERL403-PMEL30 Chester, Alexander J. (1978)
Microzooplankton in the surface waters of the Strait of Juan de Fuca, 26 pp.
NTIS: PB 297233/AS.

ERL404-PMEL31 Schumacher, J. D., R. Sillcox, D. Dreves, and R. D. Muench (1978)
Winter circulation and hydrography over the continental shelf of the northwest Gulf of Alaska, 16 pp.
NTIS: PB 296 914/AS.

ERL407-PMEL32 Overland, J. E., M. H. Hitchman, and Y. J. Han (1979)
A regional surface wind model for mountainous coastal areas, 34 pp.
NTIS: PB 80 146 152.

ERL412-PMEL33 Holbrook, J. R., R. D. Muench, D. G. Kachel, and C. Wright (1980)
Circulation in the Strait of Juan de Fuca: Recent oceanographic observations in the Eastern Basin, 42 pp.
NTIS: PB 81-135352.

ERL415-PMEL34 Feely, R. A., and G. J. Massoth (1982)
Sources, composition, and transport of suspended particulste matter in lower Cook Inlet and northern Shelikof Strait, Alaska, 28 pp.
NTIS: PB 82-193263

ERL417-PMEL35 Baker, E. T. (1982)
Suspended particulate matter in Elliott Bay, 44 pp.
NTIS: PB 82-246943.

ERL419-PMEL36 Pease, C. J., S. A. Schoenberg, J. E. Overlsnd (1982)
A climatology of the Bering Sea and its relation to sea ice extent, 29 pp.
NTIS: not yet available.

ERL422-PMEL37 Reed, R. K. (1982)
Energy fluxes over the eastern tropical Pacific Ocean, 1979-1982, 15 pp.
NTIS: PB 83 138305

NOAA Data Report Series

ERL PMEL-1 Mangum, L., N. N. Soreide, B. D. Davies, B. D. Spell, and S. P. Hsyes (1980)
CTD/O_2 measurements during the equatorial Pacific Ocean climate study (EPOCS) in 1979, 643 pp.
NTIS: PB 81 211203.

ERL PMEL-2 Katz, C. N., and J. D. Cline (1980)
Low molecular weight hydrocarbon concentrations (C_1-C_4), Alsskan continental shelf, 1975-1979, 328 pp.
NTIS: PB 82 154211.

ERL PMEL-3 Taft, B. A., and P. Kovala (1981)
Vertical sections of temperature, salinity, thermosteric anomaly, and zonal geostrophic velocity from NORPAX shuttle experiment, part 1, 98 pp.
NTIS: PB 82 163106.

ERL PMEL-4 Pullen, P. E., and H. Michael Byrne (1982)
Hydrogrsphic measurements during the 1978 cooperative Soviet-American tsunami expedition, 168 pp.
NTIS: not yet available.

ERL PMEL-5 Taft, B.A., P. Kovala, and A. Cantos-Figuerola (1982)
Vertical sections of temperature, salinity, thermosteric anomaly and zonal geostrophic velocity from NORPAX Shuttle Experiment--Part 2, 94 pp.
NTIS:

ERL PMEL-6 Katz, C.N., J.D. Cline, and K. Kelly-Hansen (1982)
Dissolved methane concentrations in the southeastern Bering Sea, 1980 and 1981, 194 pp.
NTIS:

NOAA Technical Memorandum Series

ERL PMEL-1 Sokolowski, T. J. and G. R. Miller (1968)
Deep sea release mechanism, Joint Tsunami Research Effort, pre-NTIS.

ERL PMEL-2 Halpern, David (1972)
STD observations in the northeast Pacific near 47°N, 128°W (August/September 1971), 28 pp.
NTIS: COM-72-10839.

ERL PMEL-3 Reynolds, R. Michael and Bernard Walter, Jr. (1975)
Current meter measurements in the Gulf of Alaska--Part I: Results from NEGOA moorings 60, 61, 62A, 28 pp.
NTIS: PB-247 922/8.

ERL PMEL-4 Tracy, Dan E. (1975)
STD and current meter observations in the north San Juan Islands, October 1973.
NTIS: PB-248 825/2.

ERL PMEL-5 Holbrook, James R. (1975)
STD measurements off Washington and Vancouver Island during September 1973.
NTIS: PB-249 918/4.

ERL PMEL-6 Charnell, R. L. and G. A. Krancus (1976)
A processing system for Aanderaa current meter data, 53 pp.
NTIS: PB-259 589/0.

ERL PMEL-7 Mofjeld, Harold O. and Dennis Mayer (1976)
Formulas used to analyze wind-driven currents as first-order autoregressive processes, 22 pp.
NTIS: PB-262 463/3.

ERL PMEL-8 Reed, R. K. (1976)
An evaluation of cloud factors for estimating insolation over the ocean, 23 pp.
NTIS: PB-264 174/4.

ERL PMEL-9 Nakamura, A. I. and R. R. Harvey (1977)
Versatile release timer for free vehicle instrumentation over the ocean, 21 pp.
NTIS: PB 270321/AS.

ERL PMEL-10 Holbrook, James R. and David Halpern (1977)
A compilation of wind, current, bottom pressure, and STD/CTD measurements in the northeast Gulf of Alaska, February-May 1975.
NTIS: PB 270285.

ERL PMEL-11 Nakamura, A. I. and R. R. Harvey (1978)
Conversion from film to magnetic cassette recording for the Geodyne 102 current meter, 17 pp.
NTIS: PB-283 349/9.

ERL PMEL-12 Hayes, S. P., J. Glenn, N. Soreide (1978)
A shallow water pressure-temperature gage (PTG): Design, calibration, and operation, 35 pp.
NTIS: PB 286 754/7.

ERL PMEL-13 Schumacher, J. D., R. K. Reed, M. Grigsby, D. Dreves (1979)
Circulation and hydrography near Kodiak Island, September to November 1977, 52 pp.
NTIS: PB 297421/AS.

ERL PMEL-14 Pashinski, D. J., and R. L. Charnell (1979)
Recovery record for surface drift cards released in the Puget Sound-Strait of Juan de Fuca system during calendar years 1976-1977, 32 pp.
NTIS: PB 299047/AS.

ERL PMEL-15 Han, Y.-J. and J. A. Galt (1979)
A numerical investigation of the Bering Sea circulation using a linear homogeneous model, 40 pp.
NTIS: PB 299884/AS.

ERL PMEL-16 Loomis, Harold G. (1979)
A primer on tsunamis written for boaters in Hawaii, 10 pp.
NTIS: PB80-161003.

ERL PMEL-17 Muench, R. D. and J. D. Schumacher (1980); (Hayes, Charnell, Lagerloef, and Pearson, contributors)
Some observations of physical oceanographic conditions on the northeast Gulf of Alaska continental shelf, 90 pp.
NTIS: PB81-102584.

ERL PMEL-18 Gordon, Howard R., ed. (1980)
Ocean remote sensing using lasers, 205 pp.
NTIS: PB80-223282.

ERL PMEL-19 Cardone, V. J. (1980)
Case studies of four severe Gulf of Alaska storms, 58 pp.
NTIS: PB81-102519.

ERL PMEL-20 Overland, J. E., R. A. Brown, and C. D. Mobley (1980)
METLIB--A program library for calculating and plotting marine boundary layer wind fields, 82 pp.
NTIS: PB81-141038.

ERL PMEL-21 Salo, S. A., C. H. Pease, and R. W. Lindsay (1980)
Physical environment of the eastern Bering Sea, March 1979, 127 pp.
NTIS: PB81-148496.

ERL PMEL-22 Muench, R. D., and J. D. Schumacher (1980)
Physical oceanographic and meteorological conditions in the northwest Gulf of Alaska, 147 pp.
NTIS: PB81-199473.

ERL PMEL-23 Wright, Cathleen (1980)
Observations in the Alaskan Stream during 1980, 34 pp.
NTIS: PB81-207441.

ERL PMEL-24 McNutt, L. (1980)
Ice conditions in the eastern Bering Sea from NOAA and LANDSAT imagery: Winter conditions 1974, 1976, 1977, 1979, 179 pp.
NTIS: PB81-220188.

ERL PMEL-25 Wright, C., and R. K. Reed (1980)
Comparison of ocean and island rainfall in the tropical South Pacific, Atlantic, and Indian Oceans, 17 pp.
NTIS: PB81-225401.

ERL PMEL-26 Katz, C. N. and J. D. Cline (1980)
Processes affecting distribution of low-molecular-weight aliphatic hydrocarbons in Cook Inlet, Alaska, 84 pp.
NTIS: not yet available.

ERL PMEL-27 Feely, R. A., G. J. Massoth, A. J. Paulson (1981)
Distribution and elemental composition of suspended matter in Alaskan coastal waters, 119 pp.
NTIS: PB82-124538.

ERL PMEL-28 Muench, R. D., J. D. Schumacher, and C. A. Pearson (1980)
Circulation in the lower Cook Inlet, Alaska, 26 pp.
NTIS: PB82-126418.

ERL PMEL-29 Pearson, C. A. (1981)
Guide to R2D2--Rapid retrieval data display, 148 pp.
NTIS: PB82-150384.

ERL PMEL-30 Hamilton, S. E., and J. D. Cline (1981)
Hydrocarbons associated with suspended matter in the Green River, Washington, 116 pp.
NTIS: PB82-148677.

ERL PMEL-31 Reynolds, R. M., S. A. Macklin, and T. R. Heister (1981)
Observations of South Alaskan coastal winds, 49 pp.
NTIS: PB82-164823.

ERL PMEL-32 Pease, C. H., and S. A. Salo (1981)
Drift characteristics of northeastern Bering Sea ice during 1980, 79 pp.
NTIS: PB 83 112466

ERL PMEL-33 Ikeda, Motoyoshi (1982)
Eddies detached from a jet crossing over a submarine ridge: A study using a simple numerical model, 38 pp.
NTIS: PB82-217563.

ERL PMEL-34 Liu, Cho-Teng (1982)
Tropical Pacific sea surface temperature measured by SEASAT microwave radiometer and by ships, 160 pp.
NTIS: not yet available.

ERL PMEL-35 Lindsay, R.W., and A.L. Comiskey (1982):
Surface and upper-air observations in the eastern Bering Sea, 90 pp.
NTIS: not yet available.

ERL PMEL-36 Preisendorfer, R., and C. E. Mobley (1982)
Climate forecast verifications off the U. S. mainland, 1974-1982, 225 pp.
NTIS: not yet available.