NOAA Technical Memorandum ERL PMEL-39

DATA INTERCOMPARISON THEORY

II. TRINITY STATISTICS 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. 601 from NOAA's Pacific Marine Environmental Laboratory

TABLE OF CONTENTS

Abstract

In this report, the second in a series of five on data intercomparison theory, we examine three basic measures of data-set separation and procedures by which these measures can be assigned statistical significance. The three measures are for the distances between *means* (SITES), *variances* (SPRED), and *patterns* (SHAPE) of space-time geophysical multivariate data sets. The patterns, in turn, are resolved into spatial and temporal patterns.

The problem of determining procedures to generate reference distributions, by which statistical significance is decided, is resolved into five parts, depending on the amount of data available for use. We classify availability of data into five categories: *adequate, semi-adequate, borderline, semi-inadequate,* and *inadequate*. For each of these settings we develop procedures to generate reference distributions for SITES and SPRED, and determine the power curves for these two statistics under selected procedures. These power curves are compared with those generated by some classical tests for the relative location and spread of multivariate data sets. The proposed statistics SITES and SPRED and some of their distribution-producing procedures appear to be relatively powerful and robust.

Data Intercomparison Theory

II. Trinity Statistics for Location, Spread and Pattern Differences

Rudolph W. Preisendorfer

Curtis D. Mobley

1. Introduction

The central problem of data intercomparison theory in climate studies is two-fold: given two (often spatially extensive) data sets $\underline{D}$ and $\underline{M}$, each in the form (say) of p time series of some physical field sampled n times, it is required first of all to gauge the "distance" (in some sense) between $\underline{D}$ and $\underline{M}$, and then to decide if this distance is "significantly large" or not. The outcome of such a decision could for example lead (if the "distance" is not large) to the acceptance of some dynamical hypothesis about the field; or it could in another instance be instrumental (if the "distance" is too large) in discarding or modifying a computer model attempting to simulate the physical processes of the field. It is therefore a matter of considerable importance in the study of climate processes to have some confidence in the adequacy of the statistical decision procedures by which a theory will stand or fall, or by which a model is declared faithful or not. More often than not, the classical statistical tools of gaussian reference distributions and their various derivates, and even the oft-invoked declarations of statistical independence, are not adequate to the tasks of modern climatology. In this note we examine the general problem of defining distance gauges between various attributes (location, spread, pattern) of pairs of geophysical data sets. This problem leads directly to that of constructing the reference distribution for the statistic, the means by which significance decisions are made.

There are no unique solutions to these problems, no solutions that are maximally powerful in all settings encountered in practice. We therefore cannot emphasize just one procedure by which the general intercomparison problem may be solved. Rather we will develop some general principles of data intercomparison activity which can serve to guide the construction of intercomparison procedures for a wide range of specific settings.

This wide range of settings can be split up into a set of five categories. These are defined loosely by the relative size of the data sets that are available within them for constructing the reference distributions. These categories are arranged in order of decreasing availability of *real* data sets, as follows: adequate, semi-adequate, borderline, semi-inadequate, and inadequate. We shall illustrate a reference-distribution building-procedure for each of these categories. The procedures begin with the ideal case of adequate *real* data sets. This is the real counterpart to the theoretical setting wherein a classical distribution (e.g. normal, chi square, etc.) serves as the reference distribution from which an arbitrary, unlimited number of samples can be drawn. In some real data sets now becoming available (e.g. the HSSTP assembled at CIRES by Fletcher and colleagues at Boulder, CO) we have adequate numbers of data values which fall into this first category, for some (but not all) statistical studies. In this setting all statistical questions, by definition, can be answered in adequate detail. Next come the semi-adequate data settings where there are "almost enough" samples to constitute a sample space of statistic values. By means of a simple and natural permutation procedure such a setting can be made to blossom into an adequate one, and from which statistical decisions can be drawn. Next come two categories, the borderline and semi-inadequate cases, where, despite the absence of further data or dynamic principles, relatively powerful measures can still be taken to generate the requisite

reference distributions. These two cases are perhaps the more important of the five categories studied here and accordingly we shall give them most of our attention, developing respectively the *auto-cross*, and the *pool*, *permutation procedures* (APP, PPP, resp.) for the borderline and semi-inadequate cases. The fifth (the "inadequate") category of data availability represents such a barren setting for reference distribution construction that we mention it only for completeness and consider in passing some of the (admittedly desperate) measures that one may invoke in order to generate a statistical background for a given statistic. Nevertheless, knowing that something can be done even in this setting may occasionally lead an investigator to a not unreasonable conclusion about his data set or model.

In the pursuit of all these matters it is occasionally helpful to have in mind a simple geometric visualization of the data sets being intercompared. The geometrization of a data set is illustrated in Fig. 1.1. There are four activities in the process of visualization: assembling raw data, producing digitized data (which is also filtered at this stage), the compression of a synoptic pattern into a point of E_p (p-dimensional euclidean space), and producing the space-time trajectory of the point. In this last stage, two data sets $\underline{D}$, $\underline{M}$, originally in the form of p time series (of sea surface temperature, or sea level air pressure, say) sampled n times, are now seen as two trajectories in E_p, discretely sampled at times t=1,...,n.

These geometric visualizations are heuristic, in the sense of suggesting to an investigator various possible definitions of measures of distance between the main attributes (location, spread, shape) of data sets. However, human ability to effectively visualize these attributes is limited to E_2 or E_3. When the number of time series, say, increases to four or more, resort must be made to objective reasoning by analogy and by symbol-manipulations. To illustrate

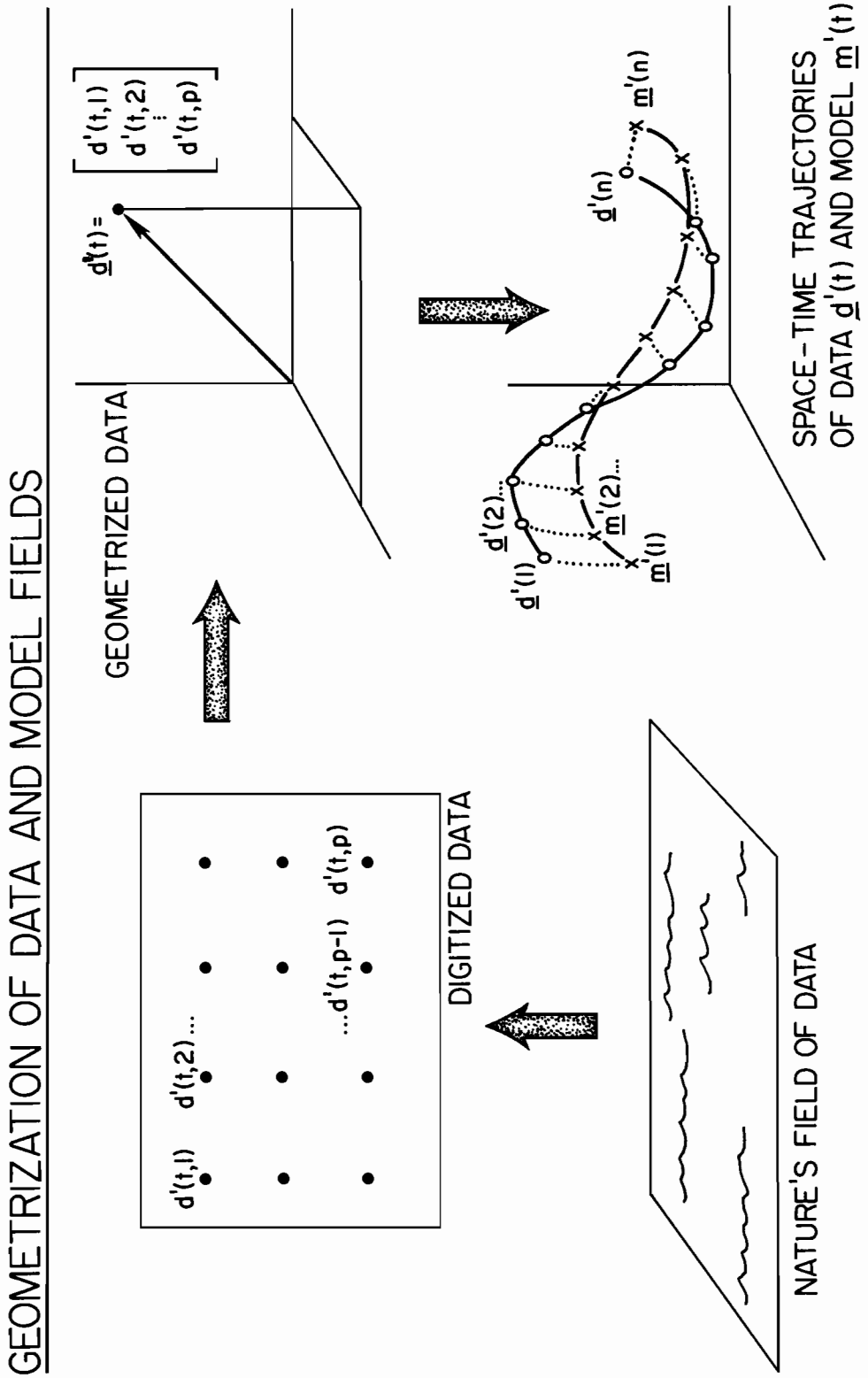

GEOMETRIZATION OF DATA AND MODEL FIELDS
GEOMETRIZED DATA
d'(t) =
d'(t,1)
d'(t,2)
d'(t,p)
DIGITIZED DATA
d'(t,1) d'(t,2)...
...d'(t,p-1) d'(t,p)
NATURE'S FIELD OF DATA
SPACE-TIME TRAJECTORIES
OF DATA d'(t) AND MODEL m'(t)
d'(1)
d'(2)...
d'(n)
m'(1)
m'(2)...
m'(n)

Fig. 1.1

the necessity of such objective measures, we have prepared three visual experiments and recorded them in Figs. 1.2, 1.3, 1.4. In each of the figures we started from the same data set shown by the circles. Each circle summarizes the simultaneous temperature (say) at two stations on the ocean surface. There were 48 such observations made at each station. In one of the figures the circles were all displaced a fixed amount parallel to one of the axes and their final locations marked by crosses. In another figure, the circles were rotated by 20° in a certain direction about their centroid. In still another figure the set of circles was subject to a transformation affecting only the overall variance of the set. Can the reader discern, visually, which diagram illustrates which of these transformations? This example may serve to indicate the occasional need for objective measures of distance between the various commonly used attributes of data sets. Visual examples could be made to illustrate the need for an even more subtle concept of data intercomparison theory, that of statistical significance. But perhaps this set of three diagrams will serve to make the point of this discussion.

Acknowledgments

Our interest in the data intercomparison problem is due largely to stimulating discussions over the past years by one of us (RWP) with Dr. Tim P. Barnett of the Climate Research Group at Scripps Institution of Oceanography. Moreover, the example in §6J is drawn from a joint study by Dr. Barnett and the authors. Ryan Whitney of PMEL typed the manuscript and Gini May of PMEL drew the figures.

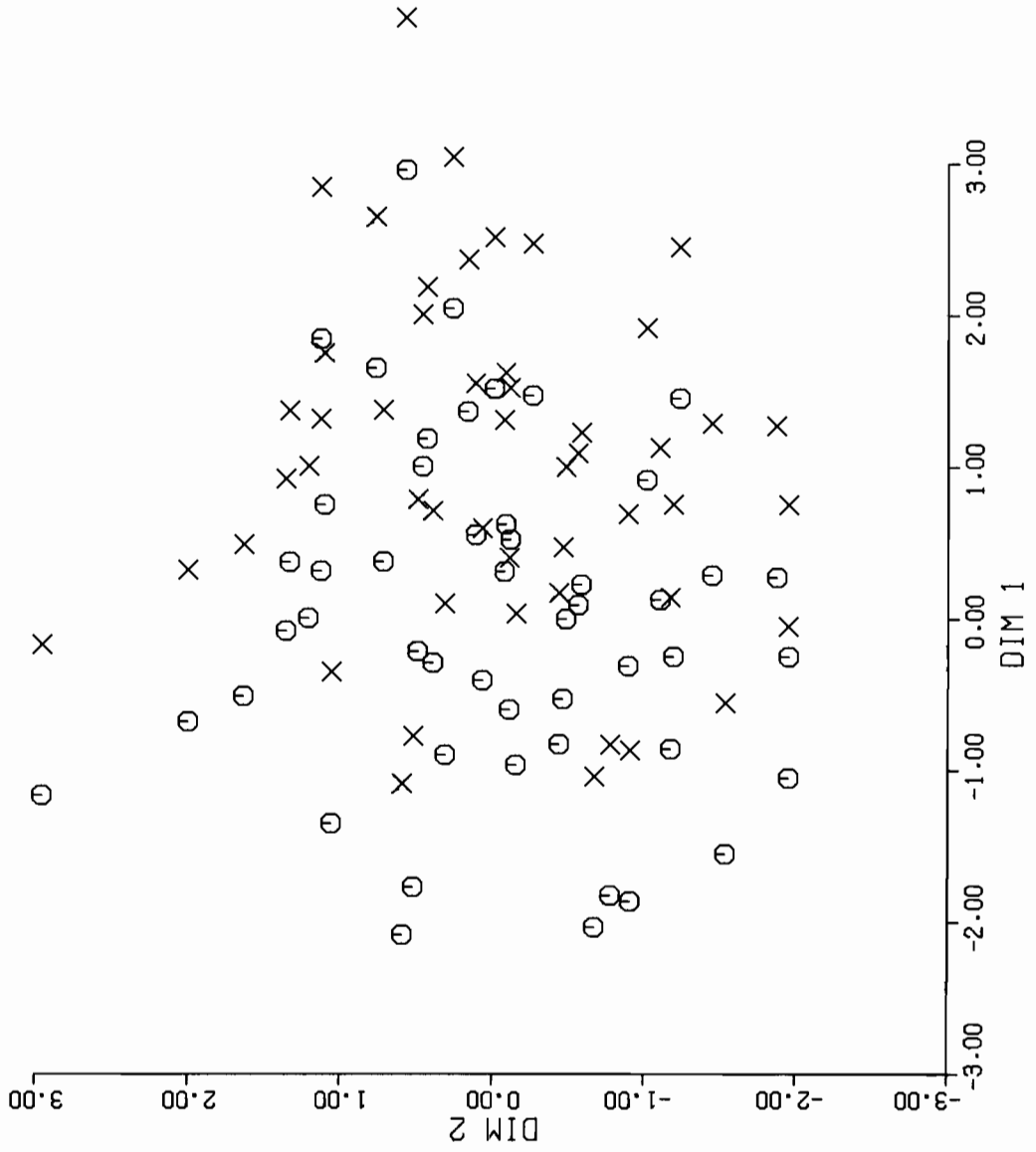

DIM 1
3.00
2.00
1.00
0.00
-1.00
-2.00
-3.00
DIM 2
-3.00
-2.00
-1.00
0.00
1.00
2.00
3.00

Fig. 1.2

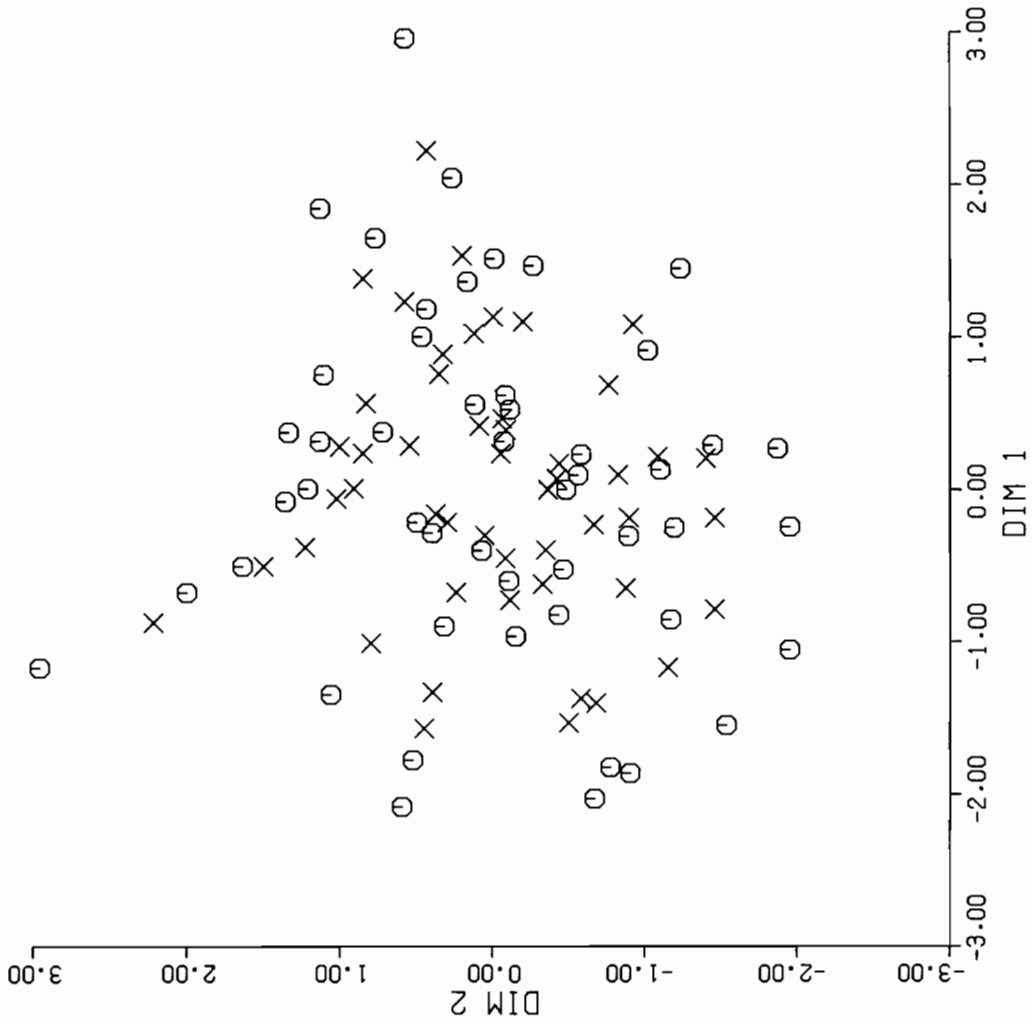

3.00
2.00
1.00
0.00
-1.00
-2.00
-3.00
DIM 1
-3.00
-2.00
-1.00
0.00
1.00
2.00
3.00
DIM 2

Fig. 1.3

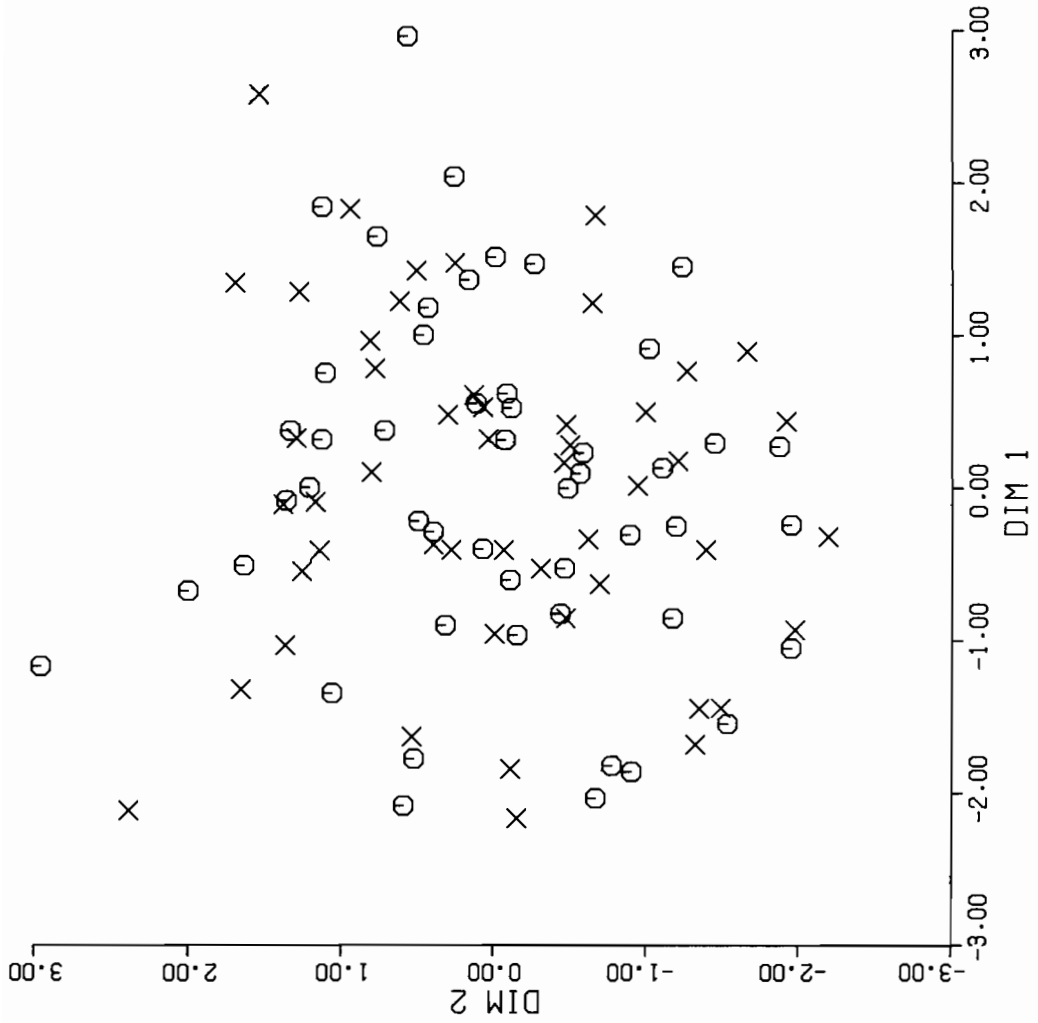

DIM 1
3.00
2.00
1.00
0.00
-1.00
-2.00
-3.00
DIM 2
3.00
2.00
1.00
0.00
-1.00
-2.00
-3.00

Fig. 1.4

2. Trinity Statistics

In the geometrization process of a data set described in the Introduction, we begin with two data sets in raw observation form. After objective analysis, each data set is in digitized form (the primes in Fig. 1.1 have been dropped for brevity):

$$\underline{D}: \quad \{d(t,x): \quad t=1,\ldots,n; \; x=1,\ldots,p\} \tag{2.1}$$

$$\underline{M}: \quad \{m(t,x): \quad t=1,\ldots,n; \; x=1,\ldots,p\} \tag{2.2}$$

Here t, x are time and space indexes. We may think of $\underline{D}$ and $\underline{M}$ either as nxp matrices, as points in euclidean space E_{np} of dimension np, or as two swarms of n points each in euclidean space E_p of dimension p. Each of these representations will be used in the developments below.

The basic measure of separation $L(\underline{D},\underline{M})$ of two points in E_{np}, such as $\underline{D},\underline{M}$ in (2.1), (2.2), is defined via

$$L^2(\underline{D},\underline{M}) \equiv \sum_{t=1}^{n} \sum_{x=1}^{p} [d(t,x) - m(t,x)]^2 \; . \tag{2.3}$$

On the other hand, we may visualize $\underline{D},\underline{M}$ as n-point swarms in E_p, i.e., as

$$\underline{D}: \quad \{\underline{d}(t): \quad t=1,\ldots,n\} \tag{2.4}$$

$$\underline{M}: \quad \{\underline{m}(t): \quad t=1,\ldots,n\} \tag{2.5}$$

where (with "T" denoting transpose):

$$\underline{d}(t) \equiv [d(t,1),\ldots,d(t,p)]^T \; . \tag{2.6}$$

$$\underline{m}(t) \equiv [m(t,1),\ldots,m(t,p)]^T \; . \tag{2.7}$$

The measure of distance $\|\underline{u}-\underline{v}\|$ between points $\underline{u},\underline{v}$, of E_p may be defined via

$$\|\underline{u}-\underline{v}\|^2 \equiv \sum_{x=1}^{p} [u(x) - v(x)]^2 \ . \tag{2.8}$$

for any two points $\underline{u} = [u(1),\ldots,u(p)]^T$, $\underline{v} = [v(1),\ldots,v(p)]^T$ in E_p. The length $\|\underline{u}\|$ of $\underline{u}$ is its distance from $\underline{v} = \underline{0}$, the origin of E_p, i.e.,

$$\|\underline{u}\|^2 = \underline{u}^T\underline{u} = \sum_{x=1}^{p} u^2(x) \ . \tag{2.9}$$

Therefore, viewing $\underline{D},\underline{M}$ as two swarms of n points each in E_p, we can write (2.3) as

$$L^2(\underline{D},\underline{M}) = \sum_{t=1}^{n} \|\underline{d}(t)-\underline{m}(t)\|^2 \ . \tag{2.10}$$

The *centroids* of the n-point swarms $\underline{D},\underline{M}$ are the points in E_p defined by

$$\underline{d} \equiv n^{-1} \sum_{t=1}^{n} \underline{d}(t) \ , \ \underline{m} \equiv n^{-1} \sum_{t=1}^{n} \underline{m}(t) \ . \tag{2.11}$$

These serve to define the locations of the swarms in E_p. The separation of swarm centers is then $\|\underline{d}-\underline{m}\|$. The radial extent or *scatters* of $\underline{D},\underline{M}$ about their centroids are given by

$$\sigma_D^2 \equiv \sum_{t=1}^{n} \|\underline{d}(t) - \underline{d}\|^2 \ , \ \sigma_M^2 \equiv \sum_{t=1}^{n} \|\underline{m}(t) - \underline{m}\|^2 \ . \tag{2.12}$$

With these notations established, we can now return to $L^2(\underline{D},\underline{M})$ in (2.10) and rearrange it, by means of a succession of algebraic transformations, into the form which will yield the measures of distance between the three main attributes of the data sets, namely the sites, the spreads, and the shapes of the sets. Thus, starting with (2.10), we have

$$L^2(\underline{D},\underline{M}) = \sum_{t=1}^{n} \|\underline{m}(t) - \underline{d}(t)\|^2$$

$$= \sum_{t=1}^{n} \{(\underline{m}(t)-\underline{m}) - (\underline{d}(t)-\underline{d}) + (\underline{m}-\underline{d})\}^T \{(\underline{m}(t)-\underline{m}) - (\underline{d}(t)-\underline{d}) + (\underline{m}-\underline{d})\}$$

$$= \sum_{t=1}^{n} \{(\underline{m}(t)-\underline{m})^T (\underline{m}(t)-\underline{m}) + \sum_{t=1}^{n} (\underline{d}(t)-\underline{d})^T (\underline{d}(t)-\underline{d}) + \sum_{t=1}^{n} (\underline{m}-\underline{d})^T (\underline{m}-\underline{d})$$

$$- \sum_{t=1}^{n} (\underline{d}(t)-\underline{d})^T (\underline{m}(t)-\underline{m}) - \sum_{t=1}^{n} (\underline{d}(t)-\underline{d})^T (\underline{m}-\underline{d})$$

$$- \sum_{t=1}^{n} (\underline{m}(t)-\underline{m})^T (\underline{d}(t)-\underline{d}) + \sum_{t=1}^{n} (\underline{m}(t)-\underline{m})^T (\underline{m}-\underline{d})$$

$$- \sum_{t=1}^{n} (\underline{m}-\underline{d})^T (\underline{d}(t)-\underline{d}) + \sum_{t=1}^{n} (\underline{m}-\underline{d})^T (\underline{m}(t)-\underline{m}) \qquad (2.13)$$

This reduces to

$$L^2(\underline{D},\underline{M}) = n \|\underline{d}-\underline{m}\|^2 \qquad \text{(site)}$$

$$+ \sum_{t=1}^{n} \|\underline{d}(t)-\underline{d}\|^2 + \sum_{t=1}^{n} \|\underline{m}(t)-\underline{m}\|^2 \qquad \text{(spread)}$$

$$- 2\sum_{t=1}^{n} [\underline{d}(t)-\underline{d}]^T [\underline{m}(t)-\underline{m}] \qquad \text{(shape)} \qquad (2.14)$$

This may be rearranged further into the following form with the help of (2.12):

$$L^2(\underline{D},\underline{M}) = n \|\underline{d}-\underline{m}\|^2 + (\sigma_D-\sigma_M)^2 + 2\{\sigma_D\sigma_M - \sum_{t=1}^{n} [\underline{d}(t)-\underline{d}]^T [\underline{m}(t)-\underline{m}]\} . \qquad (2.15)$$

In order to obtain dimensionless measures of separation of the site, spread and shape terms, (2.15) suggests division of each term by $\sigma_D\sigma_M$. When this is done we will have attained the trinity of statistics of the present study, and (2.15) becomes

$$\text{DIST2} = \text{SITES} + \text{SPRED} + \text{SHAPE} \tag{2.16}$$

where we have written

$$\text{"DIST2"} \quad \text{for} \quad L^2(\underline{D},\underline{M})/\sigma_D\sigma_M \tag{2.17}$$

$$\text{"SITES"} \quad \text{for} \quad n\,\|\underline{d}-\underline{m}\|^2/\sigma_D\sigma_M \tag{2.18}$$

$$\text{"SPRED"} \quad \text{for} \quad (\sigma_D-\sigma_M)^2/\sigma_D\sigma_M \tag{2.19}$$

$$\text{"SHAPE"} \quad \text{for} \quad 2\left\{1 - \sum_{t=1}^{n}\left[\frac{\underline{d}(t)-\underline{d}}{\sigma_D}\right]^T\left[\frac{\underline{m}(t)-\underline{m}}{\sigma_M}\right]\right\} \tag{2.20}$$

The *trinity* of statistics {SITES, SPRED, SHAPE}, which have been carved out of DIST2, will serve as our measures of separation of the three main attributes of $\underline{D},\underline{M}$, namely location, radial scale, and pattern. Fig. 2.1 illustrates, for the case p = 2, the various concepts defined in the present discussion. Here we have a schematic depiction of $\underline{D},\underline{M}$ as 8-point swarms in E_2. Their centroids are shown in part (a) of Fig. 2.1 as $\underline{d},\underline{m}$, being separated by distance $\|\underline{d}-\underline{m}\|$. The radii of $\underline{D},\underline{M}$ are schematically shown (and not to scale) by σ_D, σ_M. In part (b) of the figure the essential terms of the statistic SHAPE are interpreted graphically. Thus the vectors $\underline{u}(t) = [\underline{d}(t)-\underline{d}]/\sigma_D$ and $\underline{v}(t) = [\underline{m}(t)-\underline{m}]/\sigma_M$ are vectors in E_p for each time t = 1,...,n. They are placed end-to-end to simulate successive displacements of the data points in E_p. If $\underline{u}(t)$ and $\underline{v}(t)$ happen to point in the same direction for every t, and in particular $\underline{d}(t)-\underline{d} = (\sigma_D/\sigma_M)(\underline{m}(t)-\underline{m})$, we deduce from (2.20) that SHAPE = 0. Thus in this instance the spatial patterns spun out in time by the $\underline{d}(t)$ and $\underline{m}(t)$ points are identical, even though the swarms may have different locations and spreads about their centroids. Returning to Figs. 1.2, 1.3, 1.4, and recalling their manner of construction, the reader may verify that SHAPE = 0 for the three pairs of data sets shown in these figures.

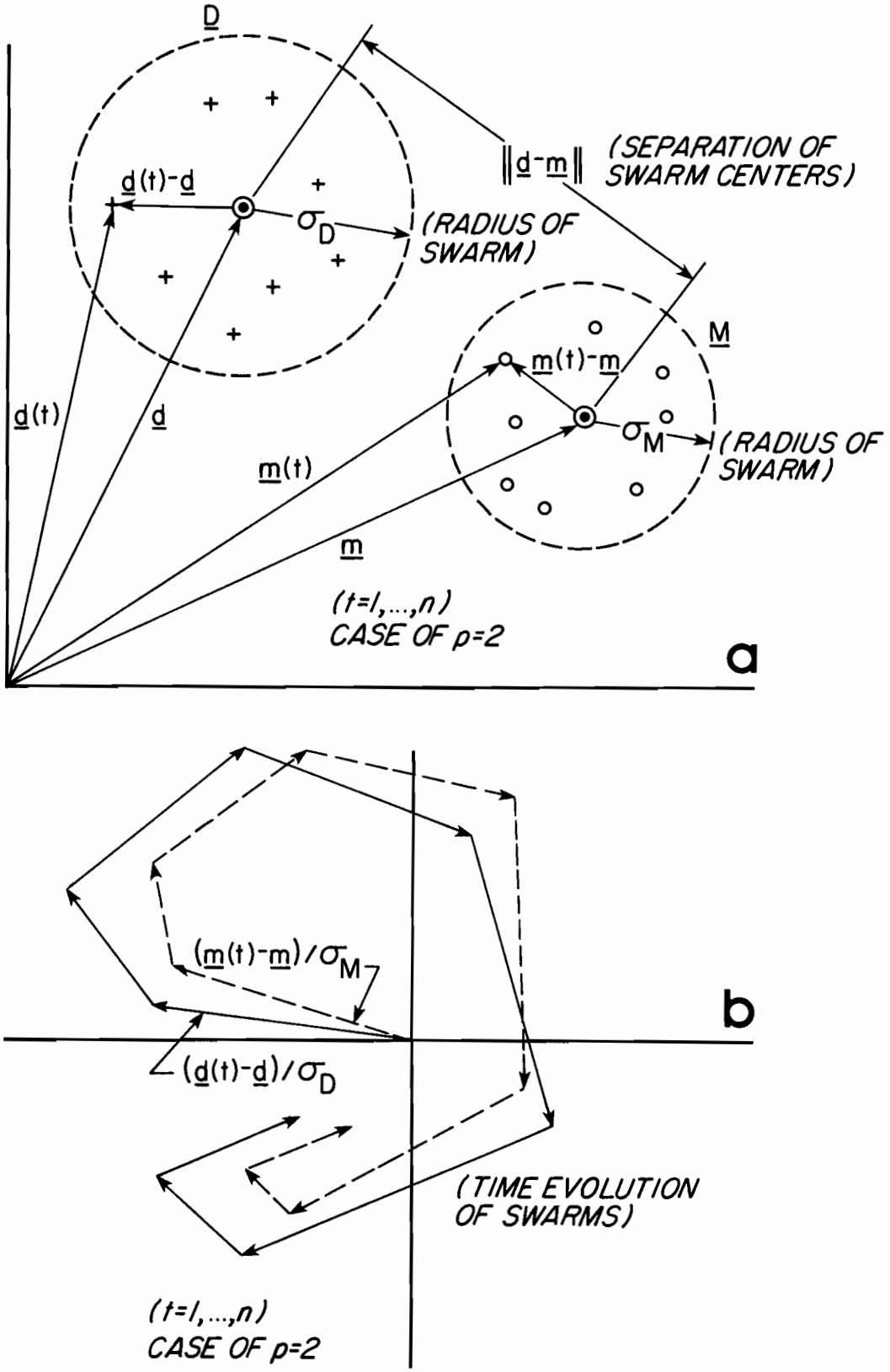

D
d(t)-d
σD
(RADIUS OF SWARM)
‖d-m‖ (SEPARATION OF SWARM CENTERS)
M
m(t)-m
σM
(RADIUS OF SWARM)
d(t)
d
m(t)
m
(t=1,...,n)
CASE OF p=2
a
(m(t)-m)/σM
(d(t)-d)/σD
(TIME EVOLUTION OF SWARMS)
(t=1,...,n)
CASE OF p=2
b

Fig. 2.1

In the present study we shall concentrate mostly on SITES and SPRED, being the simpler members of the trinity within DIST2. The statistic SHAPE, as it turns out, because of its more complex internal structure, is less powerful than SITES or SPRED using the presently adopted reference-distribution building-procedures. Its detailed examination and that of its various logical descendants will be made in part III of the present series of Data Intercomparison Theory Studies.

3. Reference Distributions in Adequate Settings (IOP)

We shall give two examples of adequate settings arising in climate work wherein the data are sufficient to the task of deciding on the statistical significance of some event. When such data sets are encountered then they can be mined for all the statistical information they possess. We shall refer to this process as the "Ideal Observation Procedure (IOP)." The most basic form of statistical information is of course in the cumulative distribution function of the statistic of interest, from which all the moments of the distribution and other derivates can be obtained. The longer the time series (or the data gathering activity) has been in existence, the more detailed can be the structures of the extracted distributions. In some statistical-gathering activities (such as actuarial, astronomical, and some but not all aspects of meteorology and oceanography) very detailed statistical tables are available for use in diagnostic and prediction research. Here are two representative examples that illustrate this basic mode of distribution-construction. One example is from a routine collection of observed records, the other is from model-building records.

A. Example of IOP: Significant Temperature Rise in Homer, Alaska

A meteorologist on a TV weather program is often heard to make a statement to the effect that today was a record breaker for hot (or cold) days, for this date in a period of, say, twenty years. If one has the appropriate record of temperatures, then it is a simple matter to deduce or verify the statement made. On another occasion a less simple and more thought-provoking statement may be that: *during the first three weeks of the past month there was a significant rise in temperature of early mornings, and that rise occurred over the first couple of days of the second week.* On first reading, the statement may well appear cryptic. Then, as the sense of it emerges, some questions arise. What could "*significant rise*" mean in this case?

In what follows we give an everyday type of example of how such a judgment of significance in temperature changes can be given a precise, verifiable meaning. We went to the monthly weather summaries that are routinely compiled for each major U.S. city by the NOAA National Climate Center at Asheville, NC, and randomly chose a city and a month. It turned out to be Homer, Alaska, for October, 1979 (whose temperature record served as the basis of the above italicized remark, and the analysis, below). We selected the first twenty-five days of the month and graphed the temperatures in the city as they were recorded at 0200 hrs local time each day. These temperatures are shown as solid dots in the upper graph of Fig. 3.1. We then took an average of these temperatures for each successive pairs of days. These averages are shown as open circles. This *averaging* was done because it was known that the speaker of the italicized observation above was referring (implicitly at least) to temperature averages over successive days. Next, since he had made a statement about temperature *changes* over pairs of days, we took the differences of these averages by subtracting the earlier average from the later. The results are graphed in

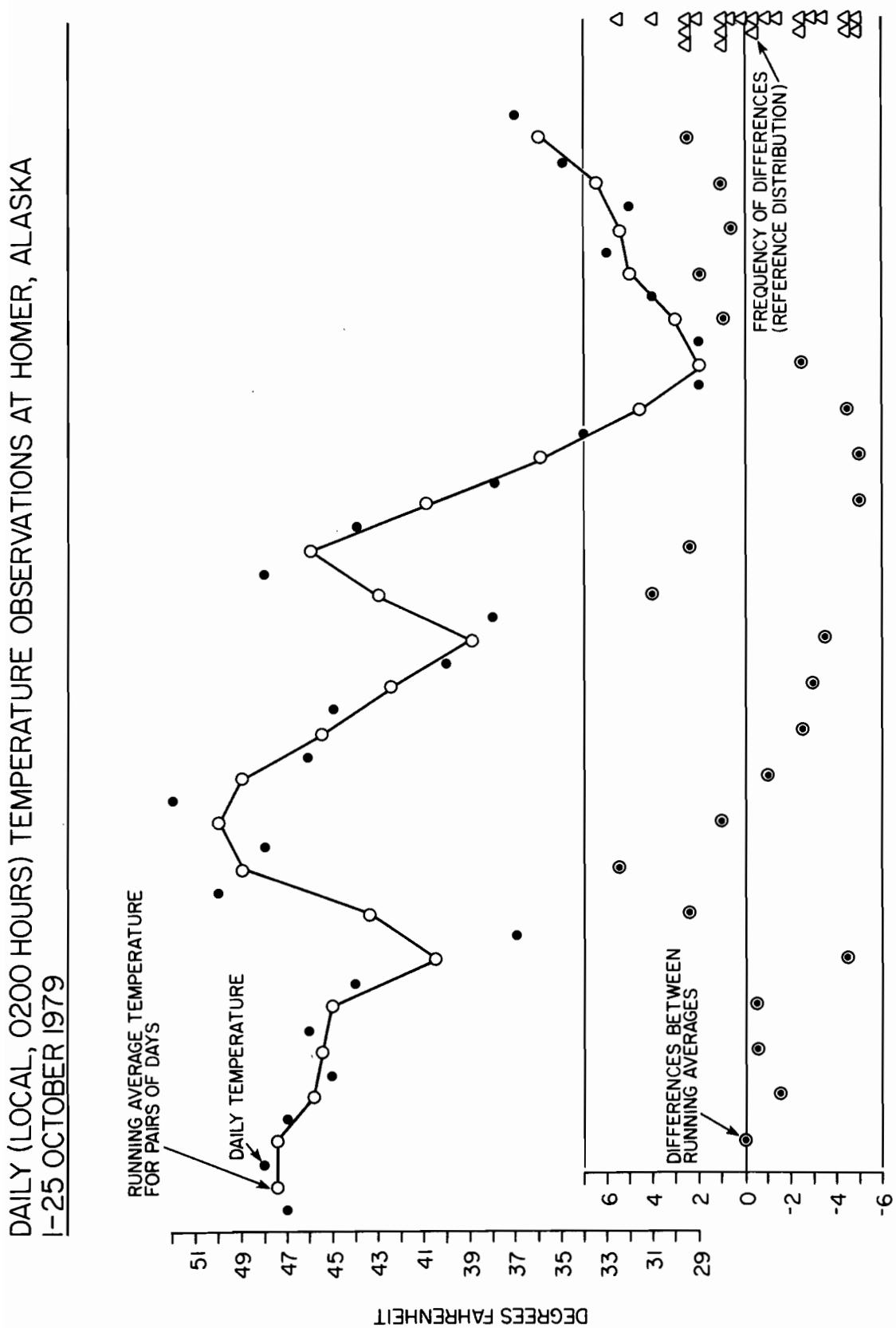

DAILY (LOCAL, 0200 HOURS) TEMPERATURE OBSERVATIONS AT HOMER, ALASKA
1-25 OCTOBER 1979
RUNNING AVERAGE TEMPERATURE FOR PAIRS OF DAYS
DAILY TEMPERATURE
DIFFERENCES BETWEEN RUNNING AVERAGES
FREQUENCY OF DIFFERENCES (REFERENCE DISTRIBUTION)
DEGREES FAHRENHEIT
DAY OF MONTH

Fig. 3.1

the lower of the two diagrams and are indicated as circled dots. The range of these differences is shown on the vertical axis to the left of the graph. The frequencies of occurrence of these differences (classified in bins on the ½°F marks) are shown on the axis to the right of the graph. The frequencies of occurrence of temperature *rises* occur on the upper half of the axis. Out of the 10 rises, that at 5½°F was the *greatest* and of the *smallest frequency of occurrence* (whence the significance of the event). This event occurred between the eighth and ninth days of the month. Thus we could, on this basis, and with confidence 90% (since it occurred once out of 10 trials) make the italicized statement about the event at the outset of this example. In fact this is what we agree is *meant* by the statement, and in particular what is meant by the words "*significant rise*" in that statement.

A quite different approach to that above, which would also be instructive, would be to obtain a great many October temperature records for Homer, pool the data, make a histogram just as above, but now for all pooled temperatures, and then make the same analysis. That would produce a larger data base on which to make the italicized claim above. In the example, as given, we are content to use only one month as a basis for the italicized remark, for we are at the moment interested in illustrating an idea, that of an adequate data setting.

This example, although of simple content, contains the essence of the activity involved in any verifiable statistical statement in climate studies: (i) an adequate record of a specific physical field; (ii) at a given location; (iii) over a particular time period (or an ensemble of time periods), of such extent that the appropriate statistics can be compiled into: (iv) a distribution from which the statement may then be read, its significance gauged, and from which the confidence of that assertion can be deduced.

B. Example of IOP: Significant Interannual Differences of a Computer Simulation of Regional Temperatures

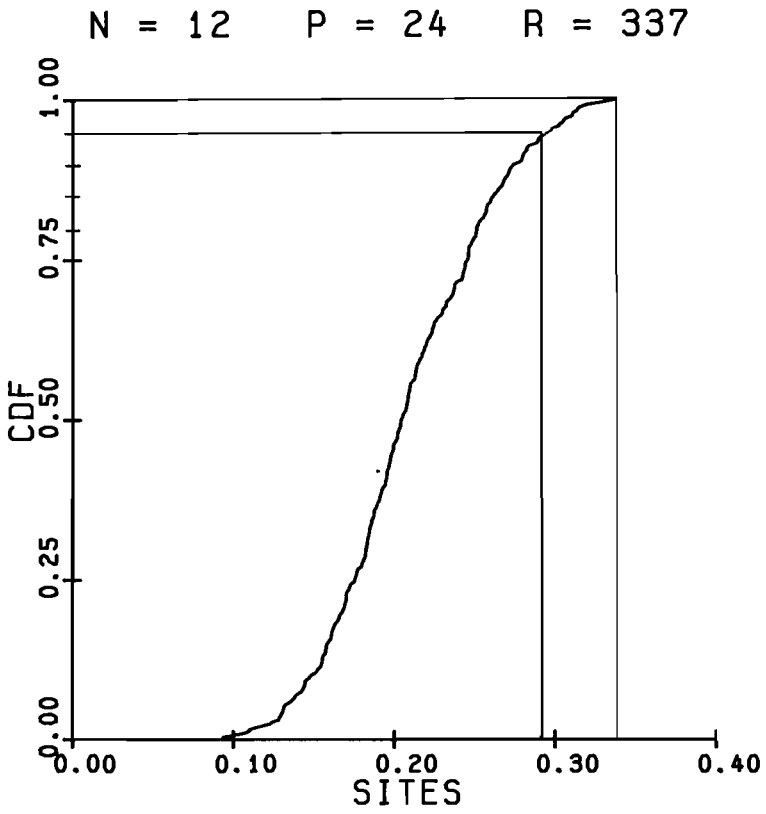
A computer model of a temperature field over a certain region yields a sample of 360 readings at each of 24 stations of that region. (The model is described in Appendix A, and will be used for other examples in this work.) The samples spaced in time are intended to represent monthly averages, and hence the period of time simulates 30 years of observations. The object of the study is to determine the range and distribution of simulated interannual differences in average yearly temperature over the 24 station network, and specifically to note the manner in which *significantly large differences (on the 5% level) in these multivariate interannual means are distributed throughout the 30-year period.*

In order to analyze the computer output of the model in accordance with the above goals we must define quantitatively the several key terms in the italicized statement. Towards this end, let $\underline{d}(t)$ be the 24-component vector of temperatures at time t in the output record, where t = 1,...,360. These are formed via (A1.9) in Appendix A. For this example, a *year of temperature records* is any set $\{\underline{d}(t+1),\ldots, \underline{d}(t+12)\}$ of temperature vectors with $0 \leq t < t + 12 \leq 360$. Hence two *adjacent years* of temperature records are given by

$$\underline{D}_j = \{\underline{d}(t): \quad t = j+1,\ldots,j+12\} \tag{3.1}$$
$$\underline{M}_j = \{\underline{d}(t): \quad t = j+13,\ldots,j+24\} \equiv \{\underline{m}(t): \quad t = j+1,\ldots,j+12\}$$

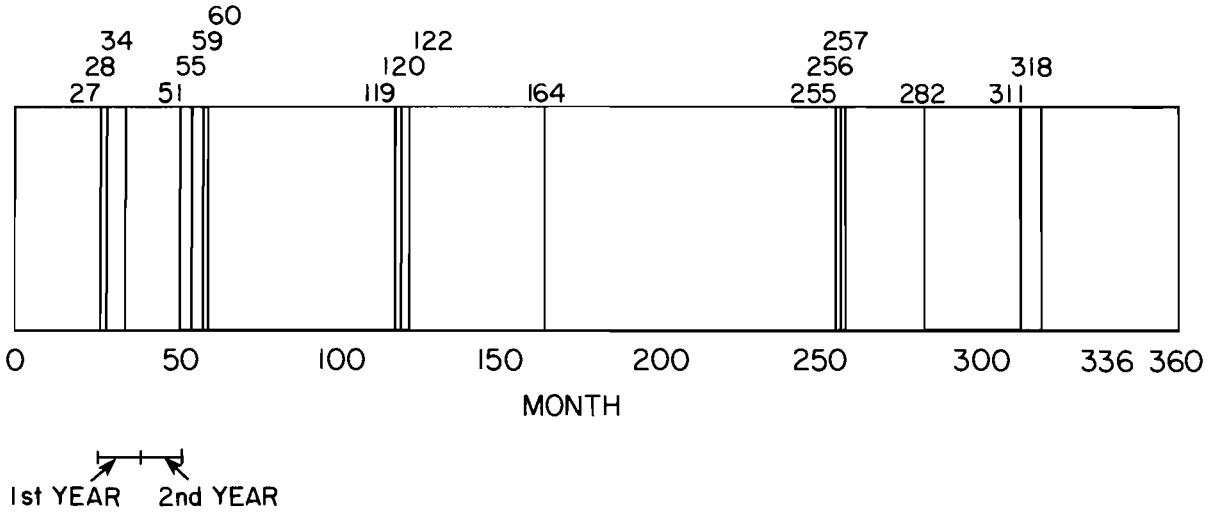
for $j = 0,\ldots,360 - 2\times12 = 336$.

The *jth interannual temperature difference* over the region is given by $S_j \equiv \text{SITES}(\underline{D}_j,\underline{M}_j)$, $j = 0,\ldots,336$.



a

OCCURRENCE IN TIME OF SIGNIFICANT DIFFERENCES OF ANNUAL MEANS OF TWO SUCCESSIVE YEARS OVER 30 YEARS



b

Fig. 3.2

The italicized statement above then directs us to order the S_j in increasing size, and select the largest 17 of them (since 5% of 337 is 17, to the nearest integer). The result is a cumulative distribution function (cdf) of the S_j. The graph in Fig. 3.2(a) gives the cdf of S_j values in this example, while Fig. 3.2(b) displays the seventeen years (i.e., the indexes j in (3.1)) which produced a significantly large multivariate mean difference over the 24-point region. The largest 17 SITES values can be seen from Fig. 3.2(a) to range from 0.296 to 0.338. The fields associated with these 5% significant values were distributed in time over the 30 years as shown in Fig. 3.2(b). Thus the first significant difference in annual means occurred over the 24 month period beginning with month 27. The last occurred over the 24 month period beginning with month 318. It is clear that the significantly large differences can occur in clusters of 3 to 4, and that these clusters and the singletons occur fairly uniformly over the years. It is not our intent here to analyze further this interesting clustering behavior (it is a function of the choice of the fundamental periods in the deterministic part of the model, the signal to noise ratio built into it, the significance level, and the sampling procedure). The main point of this example, however, has now been made: *that under ideal conditions, i.e., with an adequately large data set, basic questions of significant behavior of a statistic (SITES in this case) can be fully explored without resort to the usual classical statistical constructs (gaussianity, independence, etc.).*

4. Reference Distributions in Semi-Adequate Settings (EOP)

The present EOP (Empirical Observation Procedure) setting is one in which the available data set is not large enough to directly make the requisite

distributions in the fullness that is possible in the IOP settings of §3 above. Yet there is enough information in the typical data set of this setting to allow a simple permutation procedure to expand the data set into the requisite sample space. The outline of the procedure is given below, followed by explanatory comments and an example.

A. Given: n×p data sets $\underline{D},\underline{M}$ { over region R_o, time interval I_o where p = no. points in R_o, n = no. samples at each point.

B. Question: Is STSTC($\underline{D},\underline{M}$) significantly large? (STSTC = SITES, or SPRED)

C. Data Space: Available collection $\underline{D}_1,\ldots,\underline{D}_\Omega$, $\Omega \geq 15$ of n×p data sets over region R and time interval I, which are relevant to pars A, B above.

D. Sample Space: All ordered pairs $(\underline{D}_i,\underline{D}_j)$, $i, j = 1,\ldots,\Omega \geq 15$, $i \neq j$

E. Statistics: Form $S_k = \text{STSTC}(\underline{D}_i,\underline{D}_j)$, $k = 1,\ldots,r \leq \omega = \frac{1}{2}\Omega(\Omega-1)$

F. Reference Distribution: Order the S_k of par E according to increasing size and relabel:

$$S_{(1)} \leq S_{(2)} \leq \cdots \leq S_{(crit)} \leq \cdots \leq S_{(r)}, \quad crit \equiv (1-\alpha)r$$

where α is the significance level of the test, $0 < \alpha < 1$.

G. Null Hypothesis H_o: $(\underline{D},\underline{M})$ is in the sample space

H. Accept H_o: if $\text{STSTC}(\underline{D},\underline{M}) \leq S_{(crit)}$

Reject H_o: if $\text{STSTC}(\underline{D},\underline{M}) > S_{(crit)}$

In the latter case, STSTC($\underline{D},\underline{M}$) is significantly large; in the former case it is not.

I. Comments

The central feature of EOP, the one that distinguishes it from APP and PPP below, is step C wherein it is explicitly required that there are available at least 15 n×p data sets $\underline{D}_j$ which are *physically relevant* to the given data sets $\underline{M},\underline{D}$. This means, e.g., that if $\underline{D}$ and $\underline{M}$ are arrays of SLP measurements

over some region R_o of the Pacific and consist of monthly averages, then so too should the $\underline{D}_j$ be SLP monthly averages. But they need not always be over the same spatial or temporal domain as $\underline{M}$ and $\underline{D}$. There is accordingly, in the gathering of the data space $\underline{D}_1,\ldots,\underline{D}_\Omega$, considerable freedom as to choice of R and I, as long as the choice leads to a mathematically well-defined and physically reasonable sample space. Note that the sample space constructed in step D is open-ended in that more elements can be added beyond those generated from the present selection of the data space. We also observe that the sample space is built on the assumption that the temporal order of the $\underline{D}_j$ is immaterial to the question at hand. Thus when forming H_o we are not interested in the fact that $\underline{D}_i$ may precede $\underline{D}_j$ in time, but only in some measure of separation STSTC($\underline{D}_i$,$\underline{D}_j$) of $\underline{D}_i$ and $\underline{D}_j$. If in some investigation the time order of $\underline{D}_i$,$\underline{D}_j$ is important, this should be stated in the question of par B and the appropriate sample space in par D constructed. (For example, the illustrations of the IOP in §3 took congnizance of time ordering of the data sets.) In this extension of EOP to questions of temporal order, the condition on Ω (i.e. $\geqq 15$) may have to be changed. The present choice of $\Omega \geqq 15$ assures that $\omega \geqq 105$ so that a reasonably detailed cumulative distribution curve can be obtained and its (say) 5% significance level determined.

J. Example of EOP: model-model intercomparison

We use the model of Appendix A to generate 15 realizations of an n×p data set $\underline{D}$; where n = 36, p = 24, a = 1, μ = 0, and σ = 1.000. Let the jth of these sets be denoted by "$\underline{D}_j$", j = 1,...,15. From this we form the 105 points of the ***EOP reference distribution*** made up of the values SITES($\underline{D}_i$,$\underline{D}_j$), i, j = 1,...,15, $i \neq j$. This distribution is generated to answer the question of

whether the pair ($\underline{D}$,$\underline{M}$) (say) is in the sample space of 105 pairs ($\underline{D}_i$,$\underline{D}_j$) where $\underline{M}$ is generated by the same model but with $\mu = 0.300$. Figure 4.1 shows the reference distribution of SITES values formed, and Table 4.1 shows the values of SITES($\underline{D}_j$,$\underline{M}$), j = 1,...,15. From Fig. 4.1 we see that the critical value of SITES for the 5% significance level is at SITES = 0.063. From Table 4.1 we find SITES($\underline{D}_1$,$\underline{M}$) = 0.106. Hence in this case we would discard H_o, and conclude that the centroids of $\underline{M}$ and $\underline{D}_1$ are significantly far apart. In other words, increasing μ of the model from 0.000 to 0.300 produces a significant change in the mean value of the two resultant data sets. Indeed, if we go systematically through the entire given set of 15 $\underline{D}_j$ data matrices, and examine SITES($\underline{D}_j$,$\underline{M}$), j = 1,...,15, then, as seen from Table 4.1, except for indexes 7, 8, every one of these SITES values is significant on the 5% level. The important exception is SITES($\underline{D}_7$,$\underline{M}$) = 0.008.

In summary, we have shown that a small population of data sets (such as the $\underline{D}_j$) can be augmented by a simple pairing and permutation procedure so as to generate a collection of ordered pairs (such as ($\underline{D}_i$,$\underline{D}_j$), $i \neq j$) which then provide a reference distribution against which the statistic of some given pair ($\underline{D}_k$,$\underline{M}$) can be compared for significant size.

EOP DISTRIBUTION FOR SITES ($\underline{D}_j$, $\underline{M}$)
ESTABLISHING 5% SIGNIFICANCE
LEVEL FOR ENTRIES IN TABLE 4.1

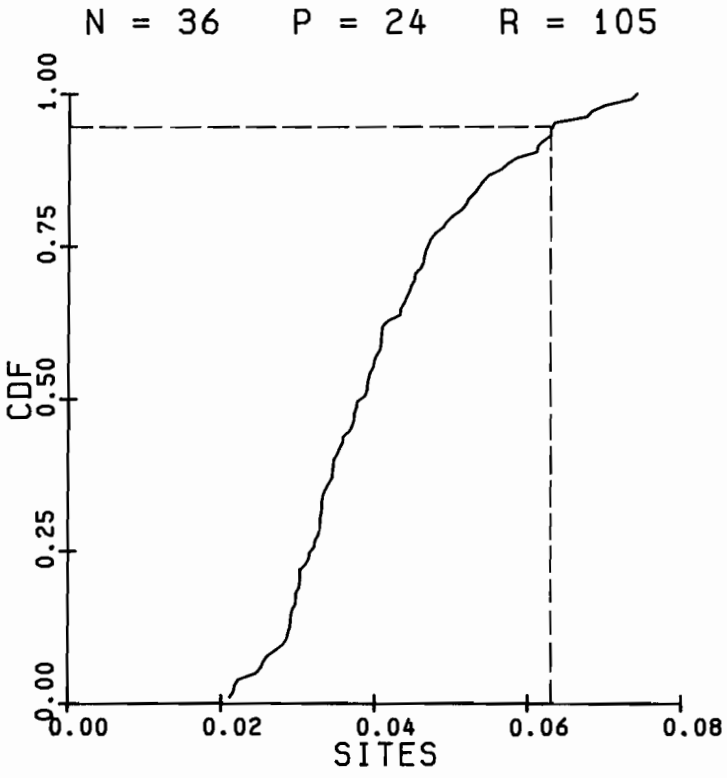


Fig. 4.1

TABLE 4.1

List of sites $(\underline{D}_j, \underline{M})$ values for j = 1,...,15, for the EOP model-model intercomparison example.

INDEX j	SITES$(\underline{D}_j, \underline{M})$
1	0.106
2	0.065
3	0.117
4	0.096
5	0.123
6	0.075
7	0.008
8	0.063
9	0.082
10	0.090
11	0.064
12	0.083
13	0.083
14	0.122
15	0.102

5. Reference Distributions in Borderline Settings (APP)

We consider now a borderline setting in the sense that we have only one n×p data set $\underline{D}$ to generate the reference distribution, in contrast to the 15 or more realizations of $\underline{D}$ available for the EOP of §4, and the 337 samples of §3. The present setting is, moreover, on the average, somewhat richer than that considered in §6 below, in that n for the present case should on average in practice be twice the n in the procedure* of §6. We call the present procedure the "Auto-cross Permutation Procedure (APP)." Some further discussion of this procedure is given in the comments below, and an example is appended to illustrate its use.

A. Given: n×p data sets $\underline{D},\underline{M}$ { over region R_o, time Interval I_o, where; p = no. points in R_o, n = no. of samples at each point

B. Question: is STSTC($\underline{D},\underline{M}$) significantly large? (STSTC = SITES, or SPRED)

C. Data Space: $\underline{D}$ only

D. Sample Space: The present elements of this are constructed from $\underline{D},\underline{M}$, as follows.

(i) Represent $\underline{D}$ as $\{\underline{d}(1),\ldots,\underline{d}(n)\}$ and

$\underline{M}$ as $\{\underline{m}(1),\ldots,\underline{m}(n)\}$

where $\underline{d}(t)$, $\underline{m}(t)$, $t = 1,\ldots,n$, are vectors in E_p.

(ii) Construct a random permutation ϕ of the set of integers $1,\ldots,n$. Let $n = 2m$.

(iii) Partition $\underline{D},\underline{M}$ via ϕ and form new n×p data sets. Thus form either the sets

$$\underline{D}_1,\underline{D}_2 \equiv \{\underline{d}(\phi(1)),\ldots,\underline{d}(\phi(m))\},\ \{\underline{d}(\phi(m+1)),\ldots,\underline{d}(\phi(n))\}$$

$$\underline{D}_1,\underline{M}_2 \equiv \{\underline{d}(\phi(1)),\ldots,\underline{d}(\phi(m))\},\ \{\underline{m}(\phi(m+1)),\ldots,\underline{m}(\phi(n))\}$$

} $\underline{D}$-based partitions

---

* This is only a crude intuitive rule of thumb. Subsequent comparative studies of APP and PPP under varying n-conditions should provide a better rule.

or, alternately, form the sets

$$\left.\begin{aligned} \underline{M}_1,\underline{M}_2 &\equiv \{\underline{m}(\phi(1)),\ldots,\underline{m}(\phi(m))\},\ \{\underline{m}(\phi(m+1)),\ldots,\underline{m}(\phi(n))\} \\ \underline{M}_1,\underline{D}_2 &\equiv \{\underline{m}(\phi(1)),\ldots,\underline{m}(\phi(m))\},\ \{\underline{d}(\phi(m+1)),\ldots,\underline{d}(\phi(n))\} \end{aligned}\right\} \underline{M}\text{-based partitions}$$

(For the case of n = 2m+1, partition according to $[\phi(1),\ldots,\phi(m+1)]$, $[\phi(m+2),\ldots,\phi(n)]$)

(iv) Repeat (ii), (iii) r times, $r \geq 100$, thereby forming r ordered pairs $(\underline{D}_1,\underline{D}_2)_1,\ldots,(\underline{D}_1,\underline{D}_2)_r$ along with $(\underline{D}_1,\underline{M}_2)_1,\ldots,(\underline{D}_1,\underline{M}_2)_r$ from the $\underline{D}$-based sets (say). These two collections of ordered pairs form the sample spaces of current interest and are respectively the *auto* and *cross* sample spaces.

E. Statistics: Form the statistics $A_k \equiv \text{STSTC}(\underline{D}_1,\underline{D}_2)_k$,

$C_\ell \equiv \text{STSTC}\ (\underline{D}_1,\underline{M}_2)_\ell$, $k,\ell = 1,\ldots,r$

F. Reference Distribution: Order the *auto statistics* A_k of par E in increasing size and relabel. Do likewise with the *cross statistics* C_ℓ. Thus form the sequence

$$A_{(1)} \leqq A_{(2)} \leqq \cdots \leqq A_{(\text{crit})} \leqq \cdots \leqq A_{(r)}\ ,\ \text{crit} = [(1-\alpha)r]$$

where α is the significance level of the test, and the sequence

$$C_{(1)} \leqq C_{(2)} \leqq \cdots \leqq C_{(r/2)} \leqq \cdots \leqq C_{(r)}\ ,\ C_{(r/2)} = \text{median}$$

Here r/2 is taken to the nearest larger integer. The $A_{(j)}$ sequence above produces the *auto-reference* (or DD) cumulative distribution function (cdf), while the $C_{(j)}$ produce the *cross-reference* (or DM) cumulative distribution function (cdf).

G. Null Hypothesis H_o: $(\underline{D}_1,\underline{M}_2)_\ell$, $\ell = (r/2)$, is in the auto sample space

H. Accept H_o: if $C_{(r/2)} \leqq A_{(\text{crit})}$

Reject H_o: if $C_{(r/2)} > A_{(\text{crit})}$

In the latter case, at least 50% of the time will a randomly selected element $(\underline{D}_1,\underline{M}_2)_k$ of the cross sample space lie beyond the critical level of the auto sample space. Thus we would conclude that, with at least probability 0.50, STSTC($\underline{D}$,$\underline{M}$) is significantly large, with significance level α.

I. Comments

The rationale for this procedure is as follows. First of all, the time order of the samples is presumed immaterial to the present procedure.

If both $\underline{D}$ and $\underline{M}$ are drawn from the same population then, as before in EOP, the switching of identity-tags on the elements $\underline{d}(t)$ and $\underline{m}(t)$ of $\underline{D}$ and $\underline{M}$ is permissible. One can readily see this by visualizing the $\underline{D}$ and $\underline{M}$ as n-point swarms in some common euclidean space E_p. The null hypothesis H_o in effect says that the labels "$\underline{d}(t)$," and "$\underline{m}(t)$" used to identify the points on each swarm are interchangeable in arbitrary ways. In principle there are up to n! points that can be in the present sample spaces. However we will usually only need $r \cong 100$ or so. The auto reference distribution provides a standard by which the location of the cross reference distribution is gauged. If the median of the cross distribution lies beyond (to the right of) the critical value of the auto distribution, then we would have reason to believe that $\underline{D},\underline{M}$ are drawn from different populations. Observe that we have used only the D-based partitions in this exposition. We know, from statistical reasoning and also numerical experiments, that for given parent distributions from which $\underline{D},\underline{M}$ are drawn, on the average, the cross distribution for $\underline{M}_1,\underline{D}_2$ equals that of $\underline{D}_1,\underline{M}_2$; and moreover that the $\underline{M}$-based partitions will, on the average, produce the same results as the D-based partitions. Thus either set of partitions may be used in practice. If the time order of the samples is relevant to a study, then the present procedure does not apply and must be modified accordingly.

J. Example of APP: Testing Effects of Different Objective Analysis Procedures on a Raw Data Set

In a recent study, Liu (1982) developed an improved Cressman type objective analysis procedure (Cressman, 1959) and applied it to raw data for sea surface temperature (SST) over a rectangle of the equatorial Pacific from 25°N to 30°S, and from the western coastline of the Americas to 150°E, and in the time

interval 1975 to 1980, inclusive. The analysis yielded 144 realizations of ½-month-averaged SST fields over the whole domain, and each realization was contoured for surface isotherms. The fields therefore included the weak El Niño of 1976 and its post period. This is a well-produced set of data fields of high quality that can be used, among other things, to examine the non-seasonal SST variation over the space-time extent of the data set.

In this example we will consider the effect of Liu's objective analysis procedure on the *mean-value* and the *variance* of the final processed field. We will compare the final Liu data set with two others arrived at in different ways and which we will designate as "Modified Levitus-Oort" and "Original Levitus-Oort." The second of these designations refers to the objective analysis scheme of Levitus and Oort (1977) which Liu extended to his present form. The first scheme was produced for this example by modifying Liu's scheme as explained below. All these schemes are variations of the Cressman procedure.

Table 5.1 below summarizes the salient features of each scheme. On the left column of Table 5.1 we list the basic operations that occur in Liu's procedure. It is not essential for our present example that the reader know the detailed mathematical forms of these operations. (They may be reviewed in Liu's (1982) study, or in more detail in DIT(V) of this series by Preisendorfer and Mobley.) The X's in Table 5.1 denote those operations of the left column that (as far as possible) are common to all three objective analysis schemes. It is seen then that the novelty in Liu's approach is fivefold: the "corrector on grid" operator (a basic feature of Cressman's original scheme) is (i) made SST gradient dependent, (ii) there is an outlier cutoff feature, (iii) the various data entering the corrector scheme are weighted in importance of quality, and (iv) the diffuser-smoother operator, absent in the original

Table 5.1 Defining the three objective analysis schemes to be intercompared

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

Levitus-Oort (and Cressman) scheme, is added in Liu's scheme and made dependent on the SST gradient. Finally, (v) a curvature-corrector term is added to Liu's scheme, relative to the original Levitus-Oort procedure.

The main motivation of this example is to see what happens to the *average values* and *variance* of the raw data set as it is passed through the three objective analysis schemes described in Table 5.1. In particular, we ask whether the SST gradient-dependence of the "corrector on the grid" and "diffuser-smoother" operators is important to the mean and variance attributes of the data set. Accordingly we shut off these features of the Liu scheme and also dropped the data weight feature. The result is the "modified Levitus-Oort" scheme. Subsequently we dropped from the latter scheme the outlier cutoff feature and dropped both the diffuser-smoother and curvature-corrector operators, thereby reverting back to the original Levitus-Oort scheme.

After defining the above three schemes, we selected four regions in the Pacific over which we would intercompare the results of the three schemes. Each of these four regions is a rectangle of extent $5° \times 10°$ in the north $\times$ east directions, and with the properties as summarized in Table 5.2. The data sets defined on these regions were of the $n \times p$ form $\underline{D},\underline{M}$, where $p = 24$ and $n = 36$. Thus in each of the four regions we looked at 24 time series of 36 ½-month averages starting in January 1975 and extending 18 months through June 1976. $\underline{D},\underline{M}$ then took on the forms of the three possible distinct combinations, as shown in the three left boxes of Table 5.3. The size of the significance level α in the intercomparison was taken as 0.10, and r was taken as 100. We applied the APP to each of these three pairs of data sets.

Table 5.2

Four 5° × 10° Regions of p = 24 points each for Data Intercomparison of Liu's Objective Analyses Scheme with Original Levitus-Oort and Modified Levitus-Oort Schemes

Location of SW corner of Rectangle	General Area where Located	Quality of Original Data in Rectangle
15°N, 120°W	Mexico coast	High data density Large SST gradient
15°N, 170°W	Hawaii	High data density Small SST gradient
2½°S, 170°W	Below Hawaii on Equator	Moderate data density Small SST gradient
2½°S, 170°W	On dateline on Equator	Low data density Small SST gradient

A study of the results in Table 5.3 suggests the following interesting features of the three objective analyses schemes (of Table 5.1) applied to data sets in regions 1-4 defined in Table 5.2.

First of all, *the locations (i.e. mean values) of the resultant data sets are not significantly affected by relaxing the gradient-dependent smoothing operations of Liu's scheme.* This is signified by the "A"s (for H_o "accepted") in the "SITES RESULT" column of Table 5.2. The actual sample means in region 1, e.g., are 24.97°C for Liu's set and 25.03°C for the original Lev-Oort set, as shown in the Table. This amounts to a 0.06°C average difference over the 24 points and 18 month period. The fact that the SITES test, using APP,

Table 5.3

Intercomparison of SITES and SPRED attributes of three pairs of data sets defined on the regions of Table 5.2 and produced using the schemes of Table 5.1. APP was used to produce this table. A = Accept H_o, R = Reject H_o.

DATA SETS	REGION	SAMPLE MEANS	SITES RESULT	SAMPLE STD DEVS	SPRED RESULT
LIU VS. ORIG LEV-OORT	1	24.97°C 25.03	A	1.154°C 1.663	R
	2	25.75 25.23	A	0.839 1.414	R
	3	27.17 27.05	A	0.7513 1.338	R
	4	28.25 28.25	A	0.566 1.096	R
LIU VS. MODIFIED LEV-OORT	1	24.97 24.80	A	1.154 1.145	A
	2	25.75 25.75	A	0.839 0.832	A
	3	27.17 27.16	A	0.7513 0.755	A
	4	28.25 28.21	A	0.566 0.522	A
MODIFIED LEV-OORT VS. ORIGINAL LEV-OORT	1	24.80 25.03	A	1.145 1.663	R
	2	25.75 25.83	A	0.832 0.414	R
	3	27.16 27.05	A	0.755 1.338	R
	4	28.21 28.25	A	0.522 1.096	R

accepted H_o in this case is therefore intuitively reasonable. We conclude in particular that data density and SST gradient properties do not play an important role in determining the mean structure of the present data sets via the defined objective analysis schemes.

On turning to the "SPRED RESULT" column, we encounter another story. In the Liu-vs-original Lev-Oort box we see (by the listed rejections R of H_o) *that by omitting the new features of Liu's procedure, we produce significantly different variance properties in the final data sets* (as gauged by SPRED via APP). The °C differences in sample standard deviations in this case are on the order of 0.5°C in region 1. When we return some of these features to the Original Lev-Oort procedure to obtain the modified Lev-Oort scheme, we see, by the Liu-vs-Modified Lev-Oort box, that the spreads of the fields are no longer significantly different (the "A"s are denoting acceptance of H_o) and the sample standard deviations are differing by 0.01°C in region 1. The conclusion we reach in this aspect of the study is that: *relaxing just the SST gradient dependent smoothing properties of the Liu objective analysis scheme does not significantly affect the variance properties of the Liu data set in any of the various regions*. However, by dropping the new smoothing operations altogether, and other features (see Table 5.1), so as to revert to the Original Levitus-Oort scheme, the variance properties are significantly affected in all regions and by amounts that are climatologically important (i.e., on the order of 0.5°C).

The preceding conclusion leads us to the next consideration (which, however, is beyond the scope of this study): which of the two schemes (Original Levitus-Oort, or Liu) is to be preferred in the objective analysis of the equatorial data? The tests have made their decisions as to the disparate SPRED properties of the resultant data sets. Table 5.3 indicates that the application of the new smoothing operations in the Liu scheme *reduces* the

average standard deviations of the resultant sets by the order of 0.5°C. The detailed pursuit of this consideration, however, will not be made here.

6. Reference Distributions in Semi-Inadequate Settings (PPP)

We descend now to the setting where the number n of samples may be so low that it may no longer be feasible to split up the $\underline{D}$ and $\underline{M}$ sets, as in the APP of §5, in order to generate the reference distribution. Instead we go the other way: we pool the n-point swarms in E_p first, and then repeatedly partition their union to generate the requisite reference distribution. We shall call this the "pool permutation procedure (PPP)." The details follow.

A. Given: $n \times p$ data sets $\underline{D},\underline{M}$ { over region R_o, time interval I_o where; p = no. points in R_o, n = no. of samples at each point

B. Question: is STSTC($\underline{D},\underline{M}$) significantly large? (STSTC = SITES or SPRED)

C. Data Space: $\underline{D}$ only

D. Sample Space: The elements of the sample space are constructed from $\underline{D},\underline{M}$ as follows.

(i) Represent $\underline{D}$ as $\{\underline{d}(1),\ldots,\underline{d}(n)\}$ and

$\underline{M}$ as $\{\underline{m}(1),\ldots,\underline{m}(n)\}$

where $\underline{d}(t)$, $\underline{m}(t)$, $t = 1,\ldots,n$, are vectors in E_p

(ii) Form the union of $\underline{D},\underline{M}$, i.e., pool the $\underline{m}(t)$ and $\underline{d}(t)$ vectors to find

$$\underline{U} \equiv \{\underline{u}(1),\ldots,\underline{u}(n), \underline{u}(n+1),\ldots,\underline{u}(2n)\}$$

$$\equiv \{\underline{d}(1),\ldots,\underline{d}(n), \underline{m}(1),\ldots,\underline{m}(n)\}$$

(iii) Construct a random permutation ϕ of the set of integers $1,\ldots,2n$. Arrange the permuted set in the order: $\phi(1),\ldots,\phi(n),\phi(n+1),\ldots,\phi(2n)$.

(iv) Partition $\underline{U}$ via ϕ:

$$\{\underline{U}_1,\underline{U}_2\} \equiv \{[u(\phi(1)),\ldots,\underline{u}(\phi(n))], [\underline{u}(\phi(n+1)),\ldots,\underline{u}(\phi(2n))]\}$$

(v) Repeat (iii), (iv) r times, $r \geq 100$, thereby forming r ordered pairs $(\underline{U}_1,\underline{U}_2)_1,\ldots,(\underline{U}_1,\underline{U}_2)_r$. These constitute elements of the *sample space*.

E. Statistics: Form the statistic values $S_k \equiv \text{STSTC}\ (\underline{U}_1,\underline{U}_2)_k$, $k = 1,\ldots,r$.

F. Reference Distribution: Order the S_k of par E according to increasing size; and relabel:

$$S_{(1)} \leqq S_{(2)} \leqq \cdots \leqq S_{(crit)} \leqq \cdots \leqq S_{(r)}\ ,\ crit = [(1-\alpha)r]$$

where α is the significance level of the test.

G. Null Hypothesis H_o: $(\underline{D},\underline{M})$ is in the sample space

H. Accept H_o: if $\text{STSTC}(\underline{D},\underline{M}) \leqq S_{(crit)}$

Reject H_o: if $\text{STSTC}(\underline{D},\underline{M}) > S_{(crit)}$

In the latter case, $\text{STSTC}(\underline{D},\underline{M})$ is *significantly large*; in the former case, it is not.

I. Comments

The present scheme has the following rationale. For the purpose of forming H_o, we envision the two data sets $\underline{D},\underline{M}$ as being produced by some common, statistically steady physical process. Thus under this assumption the sets $\underline{U}_1,\underline{U}_2$ in step D(iv) above are just as likely to have evolved under this process as $\underline{D},\underline{M}$. This assumption is tantamount to saying tht the labels "$\underline{d}(t)$," "$\underline{m}(t)$," used to identify the points of each swarm, are interchangeable in arbitrary ways within the union $\underline{U}$ of step D(ii). Granted this, the sample space may be constructed and, ultimately, one of the decisions concerning H_o is reached.

Looking at this decision process more closely and thinking specifically of the range of values $\text{SITES}(\underline{U}_1,\underline{U}_2)_k$, $k = 1,\ldots,r$ (of step E), we observe the following: if $\underline{D}$ and $\underline{M}$ have their centroids separated a distance that is large compared to σ_D,σ_M (recall Fig. 2.1a), then when we pool $\underline{D},\underline{M}$ and produce a partition $\{\underline{U}_1,\underline{U}_2\}$, the centroids of $\underline{U}_1$ and $\underline{U}_2$ will tend on the average to have $\text{SITES}(\underline{U}_1,\underline{U}_2)$ smaller than $\text{SITES}(\underline{D},\underline{M})$. If the latter value is significantly

large (in the sense that it exceeds $S_{(crit)}$) then in step H we would reject H_o.

A similar intuitive review of the decision process in step H can be made for the range of values $SPRED(\underline{U}_1,\underline{U}_2)_k$, k = 1,...,r. For this purpose it is helpful to write $SPRED(\underline{D},\underline{M})$ as $(\sigma_D/\sigma_M) + (\sigma_M/\sigma_D) - 2$. Then it is clear that if σ_D and σ_M are markedly different, SPRED will be relatively large. As these swarms $\underline{D},\underline{M}$ (with large SPRED) are pooled and partitioned in various ways, we produce $\underline{U}_1,\underline{U}_2$ with SPRED values that are, more often than not, less than $SPRED(\underline{D},\underline{M})$. If the latter value is significantly large (i.e. if it exceeds $S_{(crit)}$) then in Step H we would reject H_o. This may be visualized by experimenting with a figure like that in Fig. 2.1a. A preliminary numerical analysis of the power of SPRED under various conditions on centroid separation, sample size n, and number of positions p in space is made in §8, below. Similar numerical analyses for the power of SITES are also found in that section.

It should be noted that the number of elements of the sample space formed in step D can be quite large even for modest sample sizes n in $\underline{D}$ and $\underline{M}$. For example, if n = 10, then there are $(2n!)/(n!)^2 = 184,756$ distinct partitions that can be made and added to the sample space. Our experience shows that the number of elements in the sample space can be considerably smaller, say on the order of r = 100, from which the general shape and location of the distribution function is already discernable and, indeed, useable.

Observe, finally, that the temporal order of $(\underline{D},\underline{M})$ and hence of $(\underline{U}_1,\underline{U}_2)$ is presumed immaterial to the question at hand. Tests sensitive to the temporal order of $\underline{D}$ and $\underline{M}$ can be constructed in the above framework of PPP (analogously as in the case of the other procedures); but this matter will not be considered here.

J. Example* of PPP: Simulation of January sea level pressure by a GCM

A general circulation model may be used to simulate atmospheric and or oceanographic time series in a region and epoch of interest, and the results compared with reality. For example, we were interested in comparing with real observations a GCM simulation of a typical January sea level pressure in the relatively energetic region between 20°-80°N and 10°-180°W. A series of runs, in the space-time domain, of an early version of the NCAR GCM was provided us by R. Chervin. These were realizations obtained by five separate integrations starting from the same initial conditions. Each realization was the last 30 days of a 60-day integration. The model output was projected onto a 5° latitude by 10° longitude grid, of dimensions 36 × 33. The model matrix $\underline{M}$ was therefore of dimensions n = 5, p = 1188.

We selected from the observed data five Januarys that were closest to the long-term average January. The method was as follows: let $p(\underline{x},t)$ be the sea level pressure at location $\underline{x}$ and time t. Twenty-four years of January data were used. Define

$$\bar{p}(\underline{x}) \equiv \frac{1}{24} \cdot \sum_{t=1}^{24} p(\underline{x},t)$$

where $\underline{x}$ gives the coordinates of a typical point in the 36 × 33 grid defined above, and t indexes the January of interest. The standard deviation at $\underline{x}$ is

$$\sigma(\underline{x}) = \left\{ \frac{1}{23} \sum_{t=1}^{24} (p(\underline{x},t) - \bar{p}(\underline{x}))^2 \right\}^{\frac{1}{2}}$$

A separation index $\gamma(t)$ was defined by

$$\gamma(t) = \sum_{\underline{x}} \left[\frac{p(\underline{x},t) - \bar{p}(\underline{x})}{\sigma(\underline{x})} \right]^2$$

---

* This example is drawn from a joint study by T. P. Barnett and the authors.

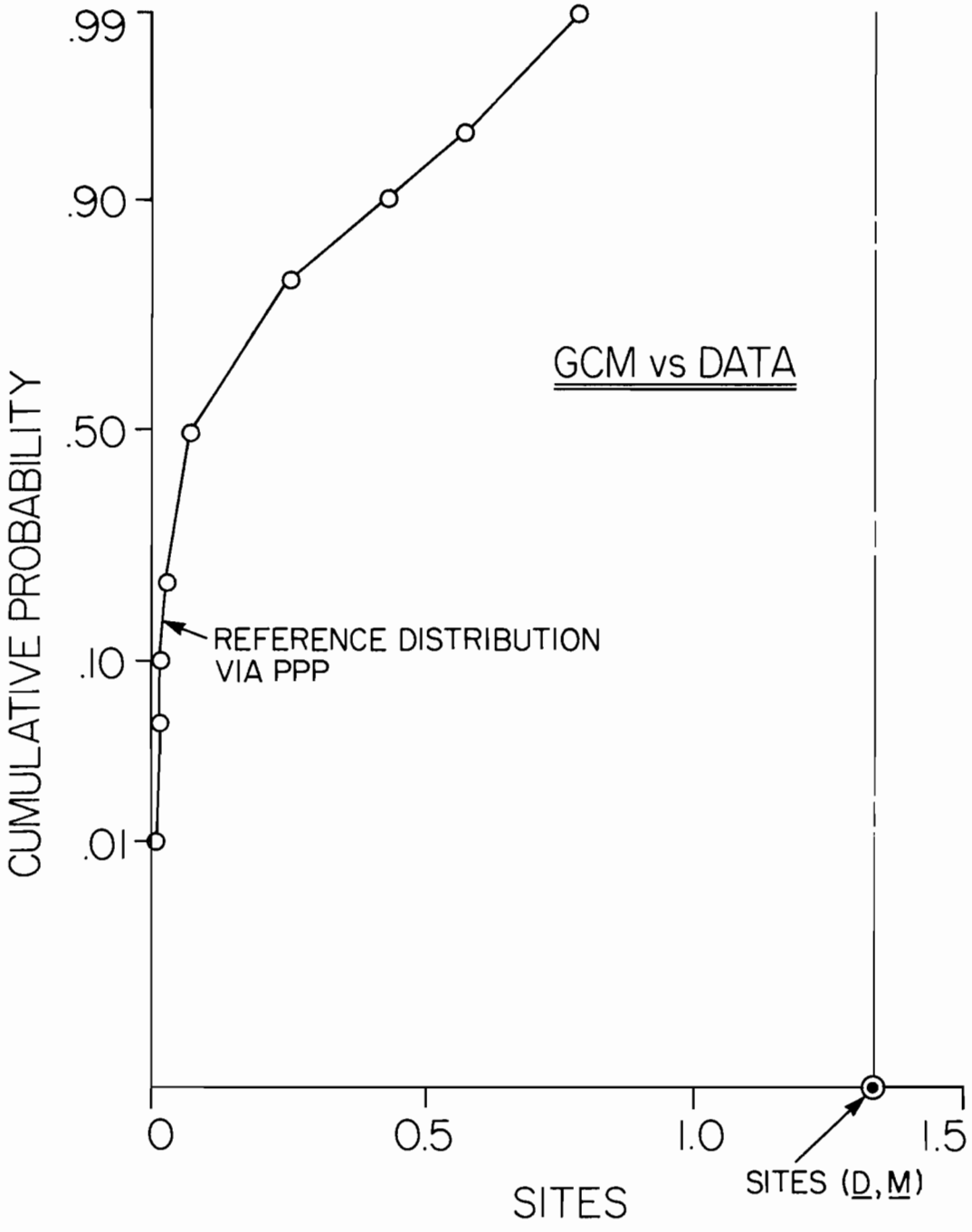

.99
.90
.50
.10
.01
CUMULATIVE PROBABILITY
GCM vs DATA
REFERENCE DISTRIBUTION
VIA PPP
0
0.5
1.0
1.5
SITES
SITES (D, M)

Fig. 6.1

The five Januarys with the smallest values of $\gamma(t)$ were selected for the test, and considered "typical." This produced the 5×1188 matrix $\underline{D}$.

The question posed was, "Did the GCM reproduce the observed mean January sea level pressure field in the domain of interest?" This question was answered by the PPP using SITES. Specifically, we asked (as in Step B of PPP) "is SITES($\underline{D}$,$\underline{M}$) significantly large?" The answer resides in Fig. 6.1. In that figure is the resultant cumulative distribution curve formed, as in Step F, using r = 100 realizations out of the possible 252 values formed by pooling the two 5-point swarms in E_{1188}. The SITES($\underline{D}$,$\underline{M}$) value is shown on the abscissa of the graph, and is seen to lie beyond the 99% critical level. We therefore reject H_o and *conclude with confidence exceeding 99% that, in the region of interest, the GCM's "typical" January sea level pressure (SLP) field is different, in the sense of mean value, from a "typical" observed January SLP.*

7. Reference Distributions in Inadequate Settings (CIP)

We come now to the rock-bottom of the hierarchy of data settings. This is typified by data-model matrix pairs of the following two kinds

(a) $2 \times p$ $\underline{D}$, $1 \times p$ $\underline{M}$

(b) $1 \times p$ $\underline{D}$, $1 \times p$ $\underline{M}$

Thus, in (a) we have $\underline{D}$ consisting of two maps of some physical field at a common set of p points, while the "model" set $\underline{M}$ consists of one map over the same set of points. The object of the intercomparison test is to see if $\underline{M}$ in some sense (say, location or dispersion) "belongs" to the pair of $\underline{D}$ maps, or more generally whether the three maps belong to some parent population. In

situation (b) we literally hit rock-bottom in the number of maps to be intercompared, where we inquire whether the map of $\underline{D}$ and the map of $\underline{M}$ belong to some common population.

It is quite possible that the situation in (a) can arise where $\underline{D}$ is for example the result of two very expensive GCM runs simulating January SLP (recall §6J). What can be done in this case? Since we have now only 3 points in E_p, a PPP-type approach will yield only 3 points in the sample space. Clearly this is not a feasible approach in this case.

One interesting possibility that may yield something to work with in case (a) is the following. Imagine the values of the first row of $\underline{D}$ to stick up like toothpicks of different lengths from a table top. These are the values of the field at the p points. Similarly for the second row of $\underline{D}$. Then randomly permute the vertical toothpicks in the first map. Do likewise with the second map. In this way form two new (permuted) maps. If the euclidean distances between the maps of $\underline{D}$ and $\underline{M}$ are of interest then find the distance between these two permuted maps of $\underline{D}$. Do this permutation and distance determination r ($\cong$ 100) times* and form the cdf of the distances. Then see where on this cdf axis the two distances between the map of $\underline{M}$ and each of the maps of $\underline{D}$ fall. If there is any real discrepancy between the values of the rows of $\underline{D}$ and the row of $\underline{M}$, these two distances will fall in the right tail of the cdf and this procedure may show it up.

Implicit in this procedure is the assumption that the individual values of the $\underline{D}$ maps are drawn from some common univariate population. If the nature of this population is assumed gaussian, e.g., then a gaussian bell shaped curve fit can be made to the p values of each row of $\underline{D}$ (thought of as independent

---

* If p = 5, then there are already 120 such permutations for each map of $\underline{D}$, and so on the order of $(120)^2$ = 14,400 points for the cdf.

samples) and the two gaussian curves intercompared. If there is reasonable agreement between the two, then some standard classical tests clearly could be applied to the rows of $\underline{D}$ and that of $\underline{M}$ for consideration of membership of field values (singly or in map form) in a common population.

When we turn to case (b), it is clear that a similar tactic of map-value permutation can be carried out on the values of the map of $\underline{D}$, with euclidean distances between these maps and the original $\underline{D}$ map noted and arranged in a cdf. At the same time, distances between the map of the original $\underline{M}$ and each new permuted $\underline{D}$ map can be noted and arranged in a cdf. Then the two cdfs can be compared. If the median of the M-produced cdf lies beyond the α level of the $\underline{D}$-produced map then (as in APP) we would decide that $\underline{D}$ and $\underline{M}$ are significantly separated.

Another possibility for case (b) is to make some working assumption about the population from which the values of $\underline{D}$ were drawn, the kind of assumption that abounds in routine, cook-book approaches to statistics. This would lead to the use of classical type tests for the intercomparison of the two maps. (Hence the generic name for the present procedure: CIP = Classic Intercomparison Procedure.) It should be noted that once *any* clearly stated working assumption of this kind is made, then the intercomparison is readily and rigorously (but not necessarily meaningfully) executable. ***It is therefore the choice of the working assumption and its defence by the chooser that lies at the base of a Classic Intercomparison Procedure in the present setting.*** The likelihood of the success of any such Classic Intercomparison Procedure rests heavily on prior knowledge by the investigator of the statistics of the values of $\underline{D}$ (as in the instance where the row or rows of $\underline{D}$ are the result of GCM runs, or where $\underline{D}$ has strong thermodynamic or dynamic constraints placed on its values by a priori reasoning). With these remarks we leave this unfinished matter here.

8. Power Curves of Trinity Statistics via Classical Sampling Procedures, APP and PPP

A. The Notion of a Power Curve

There is a way of testing a new statistical procedure (as, e.g., SITES via APP), before it is applied to real data, to obtain a preliminary impression of how "powerful" it is in discerning whether some given null hypothesis H_o is false. Suppose we have two very populous swarms in E_p (see, e.g., Fig. 2.1(a) - but now there are, say, hundreds of times more points than $\underline{M}$ and $\underline{D}$'s 8 points in the respective populations that contain $\underline{M}$ and $\underline{D}$). One of these we will call the "$\underline{D}$-population," the other the "$\underline{M}$-population." To begin a power test, say for SITES via APP, we would place the centroids of the two populations together, i.e., the $\underline{D}$-population and the $\underline{M}$-population have the same average value. Next we would sample n-point swarms $\underline{M}$ and $\underline{D}$ randomly from each population, find SITES($\underline{M}$,$\underline{D}$), and the associated cdf for SITES via APP. We would then be able to make a decision to accept or reject H_o at a given significance level α, $0 < \alpha < 1$. We would repeat this random drawing and arrive at the H_o-decision altogether, say, ten times and note how many times out of ten H_o is rejected. If the statistical procedure is a flawless one, then H_o (which says that, as far as SITES is concerned, the $\underline{D}$- and $\underline{M}$-populations are the same) would never be rejected under the present centroid separation condition. However it is clear there is a definite built-in probability that H_o could be rejected under these conditions. This is the significance level α, the probability of a Type I error. Next, the $\underline{M}$- and $\underline{D}$-population centroids are separated by various amounts (so that H_o is false) and the entire testing procedure repeated, obtaining for each centroid separation the fraction of times out of ten that H_o is rejected. The probability β of accepting H_o when H_o is false measures the Type II error. Now, of two statistical procedures

being tested (say, SITES via APP vs. SITES via PPP) the one that on average has a higher probability of H_o-rejection (a smaller β), when H_o is false, is the more powerful of the two. The power P we measure, below*, is defined as $1-\beta$. With this general overview of the notion of a power test, we consider some specific cases of interest.

B. Power Curves for DIST2 when Sampling Populations are Purely Random

(i) (DIST2) As a first experiment we considered the case of $p = 2$, which is relatively easy to visualize as taking place in a plane. Two gaussian populations were generated, each of unit variance and zero mean at the outset. Then one was displaced so that the distance between centroids was 1.75 units (in the same scale as the variance was measured). Next, $n = 50$ points were randomly drawn from each population, and the procedure, outlined in par A, carried out. In all, ten such trials were made using DIST2 as the statistic. Four out of those 10 times H_o was rejected with confidence $1-\alpha = 0.9$. This determined the point (1.75, 0.4) in the power curve of DIST2 shown in the upper left part of Fig. 8.1. Three more sets of ten trials were made, namely for a spatial offset Δx of centroids by amounts 1.85, 1.90, and 2.00. Once again decisions were made on the $1-\alpha = 0.90$ confidence level, as in all the tests below. The resultant power curve for DIST2 is faired in as shown in Fig. 8.1 (upper left). This test is disappointing, as regards DIST2 as a measure of separation of two swarms. The results say e.g., that we must move

---

* In some practical applications of power tests, the probability β is used instead of P. Sometimes β is called the *operating characteristic* of a statistical procedure (Crow, Davis, Maxfield, 1960). A basic principle in statistics is: of two tests having the same α, that which has the smaller β (greater power) is to be preferred.

the $\underline{D}$ and $\underline{M}$ populations 1.75 σ-units apart before we can, on average, detect with probability 0.4 that they are indeed apart. We conclude that the power of DIST2 under straightforward random sampling is too low for practical applications.

(ii) (DIST2) As a second experiment we considered the case p = 2, n = 50 once again, but now we multiplied each member of one of the swarms by a scale factor, in effect decreasing the swarm's radius (its standard deviation). The lower left curve of Fig. 8.1 shows the results of 5 experiments, each consisting of ten trials of the kind outlined in par A. It was not until after we contracted the radius (std dev) of one of the populations by a factor 0.30, that DIST2 woke up to the fact that the two swarms had different radii. Once again DIST2 was disappointing in its power.

(iii) (DIST2) As a third experiment, we multiplied only one of the components (the x component) of each point of one of the two dimensional swarms by a scalar factor e (the "ellipticity"). The results of the power test of DIST2 in this case are shown by the three points of the upper right curve of Fig. 8.1.

(iv) (DIST2) The fourth experiment consisted of random samples of n = 50 from each population and then rotating one of these samples by various amounts before applying the power test procedure. The results are shown in the lower right curve of Fig. 8.1.

The net conclusion of the four power tests described above is that the test based on DIST2, under random sampling conditions, is a low power test for differences in location, spread, and shape of two data sets.

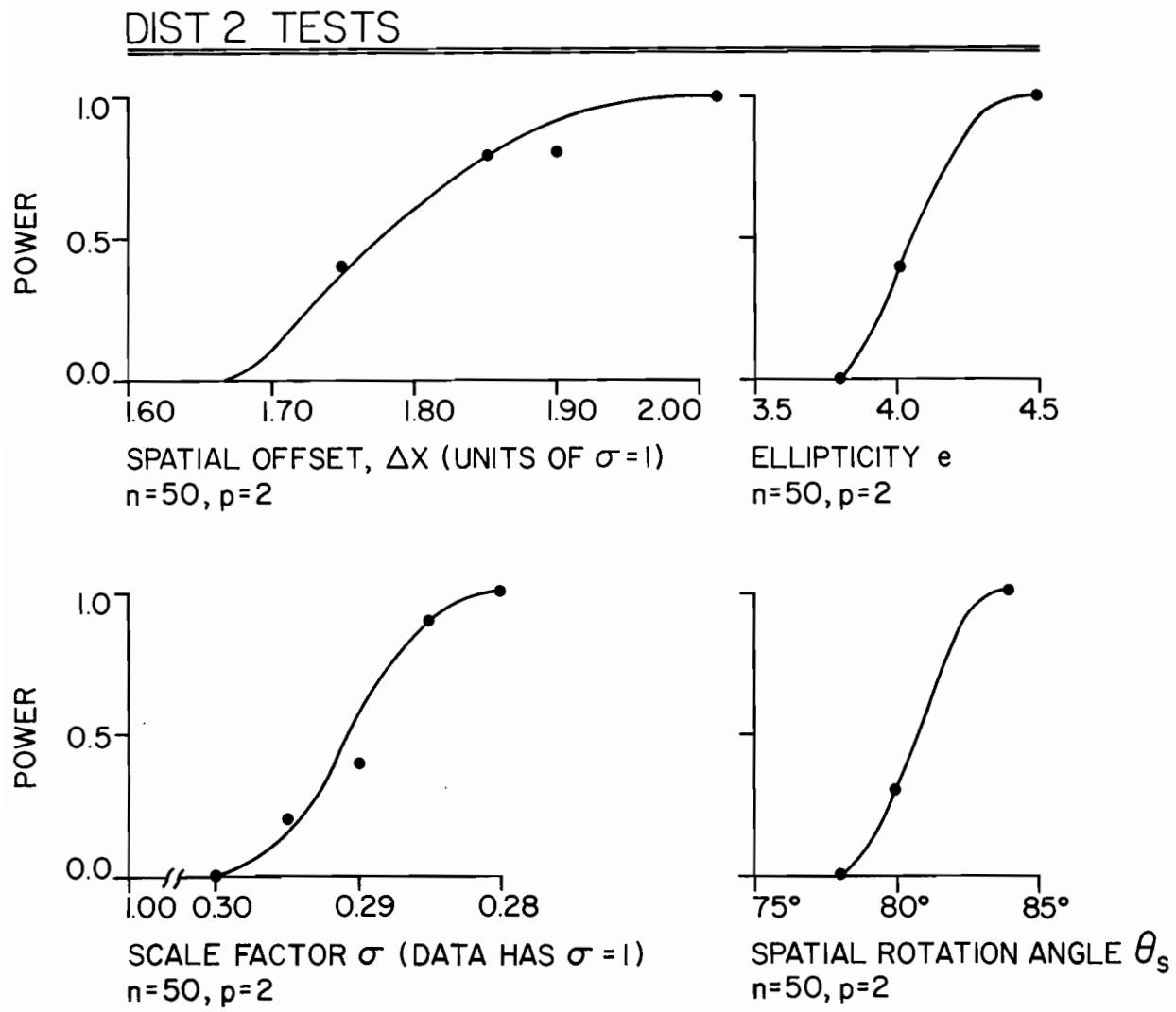
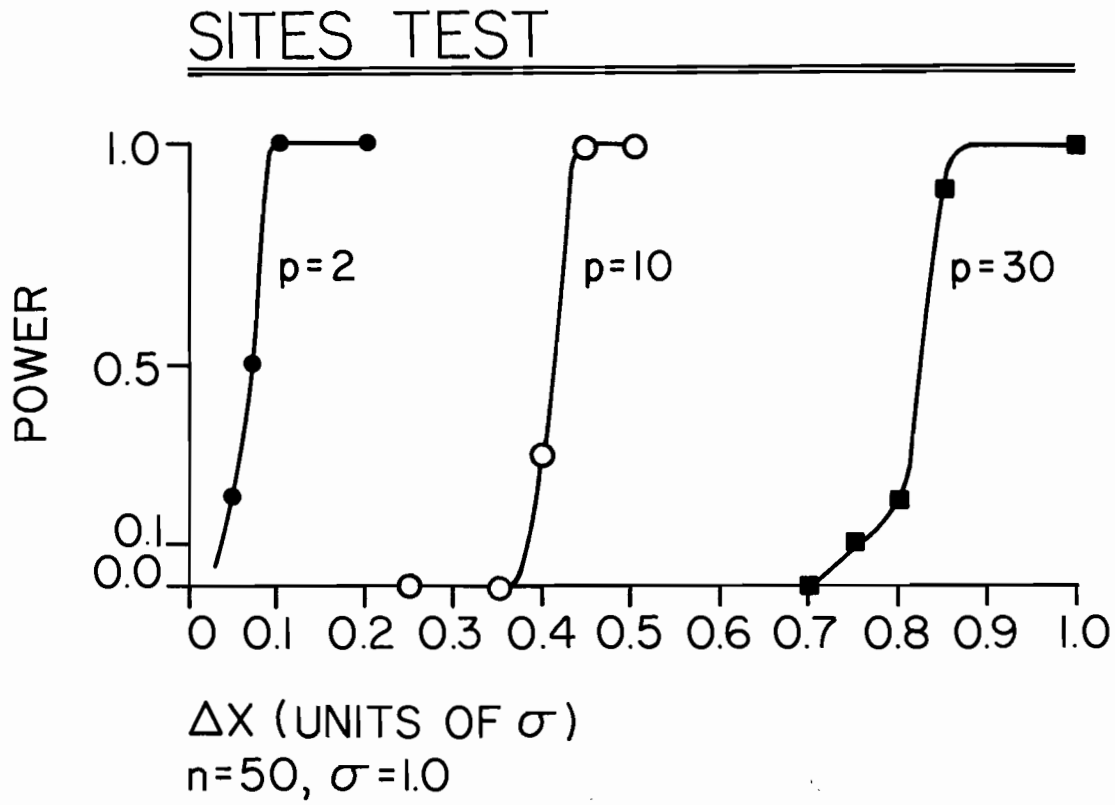

DIST 2 TESTS
POWER
1.0
0.5
0.0
1.60
1.70
1.80
1.90
2.00
SPATIAL OFFSET, ΔX (UNITS OF σ = 1)
n=50, p=2
3.5
4.0
4.5
ELLIPTICITY e
n=50, p=2
POWER
1.0
0.5
0.0
1.00
0.30
0.29
0.28
SCALE FACTOR σ (DATA HAS σ = 1)
n=50, p=2
75°
80°
85°
SPATIAL ROTATION ANGLE θs
n=50, p=2

Fig. 8.1

This conclusion motivated the search for alternate measures of differences in location, spread, and shape and resulted in the trinity of statistics SITES, SPRED, and SHAPE defined in §2. The results of power tests for these statistics, under various n-sampling of gaussian populations, of various p-dimensions, are shown in Fig. 8.2, 8.3, 8.4. We will now discuss each of these in turn.

C. Power Curves for SITES when Sampling Populations are Purely Random

SITES power curves are shown in Fig. 8.2. These curves were produced following the same procedure used for the DIST2 curves. These are all based on $\alpha = 0.10$, as also are all the tests below. For $p = 2$, we see in the upper figure that unit power of SITES already exists when Δx (the population centroid separation) is 0.1. In the corresponding case for DIST2 (upper left, Fig. 8.1) Δx had to be increased to 2.00 before unit power of DIST2 was attained. It seems that the act of splitting SITES off from DIST2 was an effective move toward a more powerful location test. Observe, however, that increasing p to 10 and then to 30 resulted in a marked decrease in power of SITES for a given Δx. Yet, even for $p = 30$, SITES has unit power when, under the same sampling ($n = 50$) conditions, DIST2 still has zero power. In the lower panel of Fig. 8.2, we briefly explore the n-dependence of the power of the SITES test. Here $p = 10$ and, as may be expected, power decreases (for a fixed Δx) as the sample size goes down. Notice that the middle curve (for $n = 50$) in the lower panel is identical to the $p = 10$ curve in the upper panel (as it should be).


SITES TEST
POWER
1.0
0.5
0.1
0.0
p=2
p=10
p=30
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
ΔX (UNITS OF σ)
n=50, σ=1.0

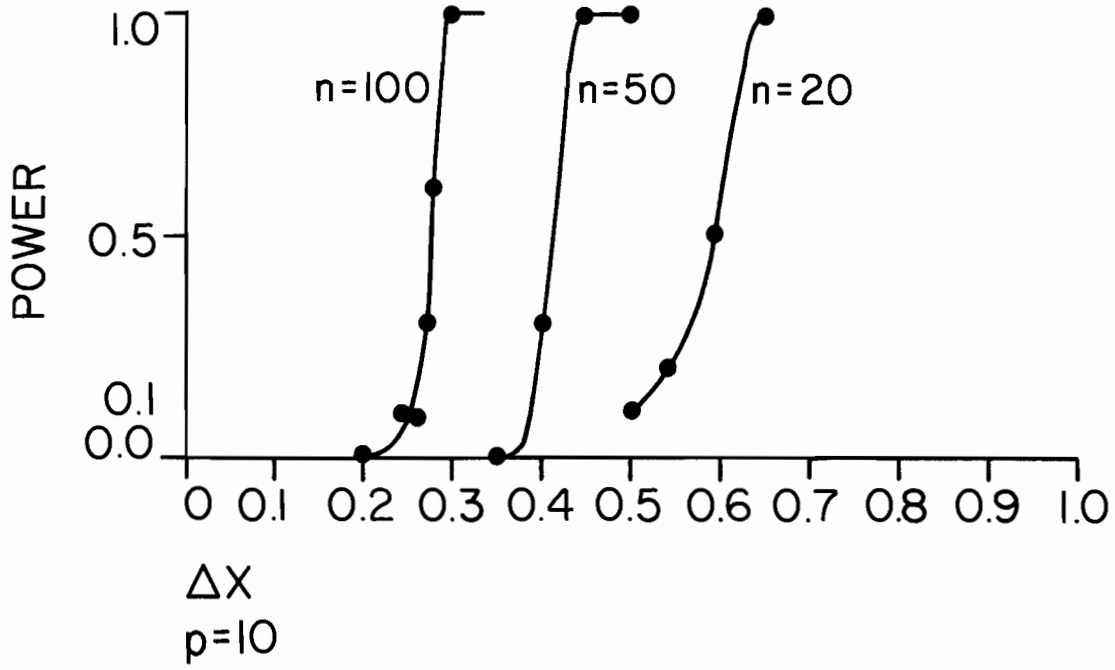
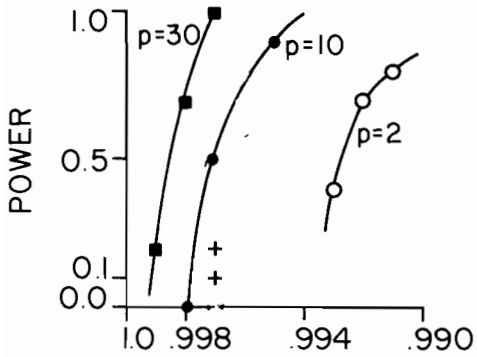

POWER
1.0
0.5
0.1
0.0
n=100
n=50
n=20
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
ΔX
p=10

Fig. 8.2

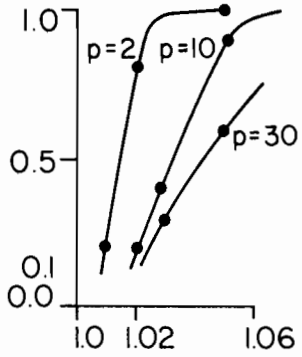
SPRED TEST

(FUNCTION OF σ)

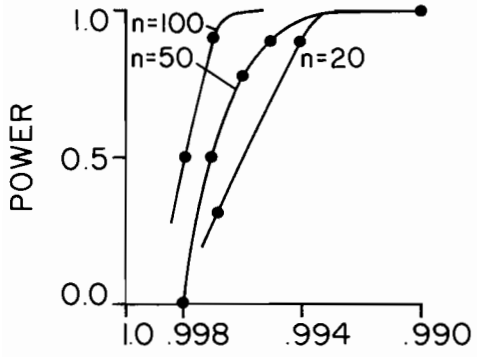


n=50 σ FACTOR APPLIED TO EACH DIRECTION IN E_p
+: n=50, p=30 BUT σ APPLIED TO 10 DIMENSIONS ONLY

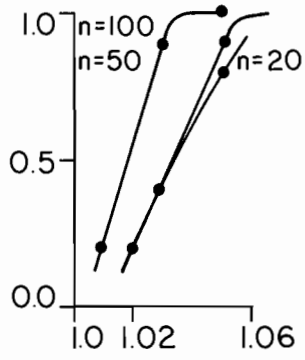
(FUNCTION OF e)



n=50



p=10, e=1.0



p=10

Fig. 8.3

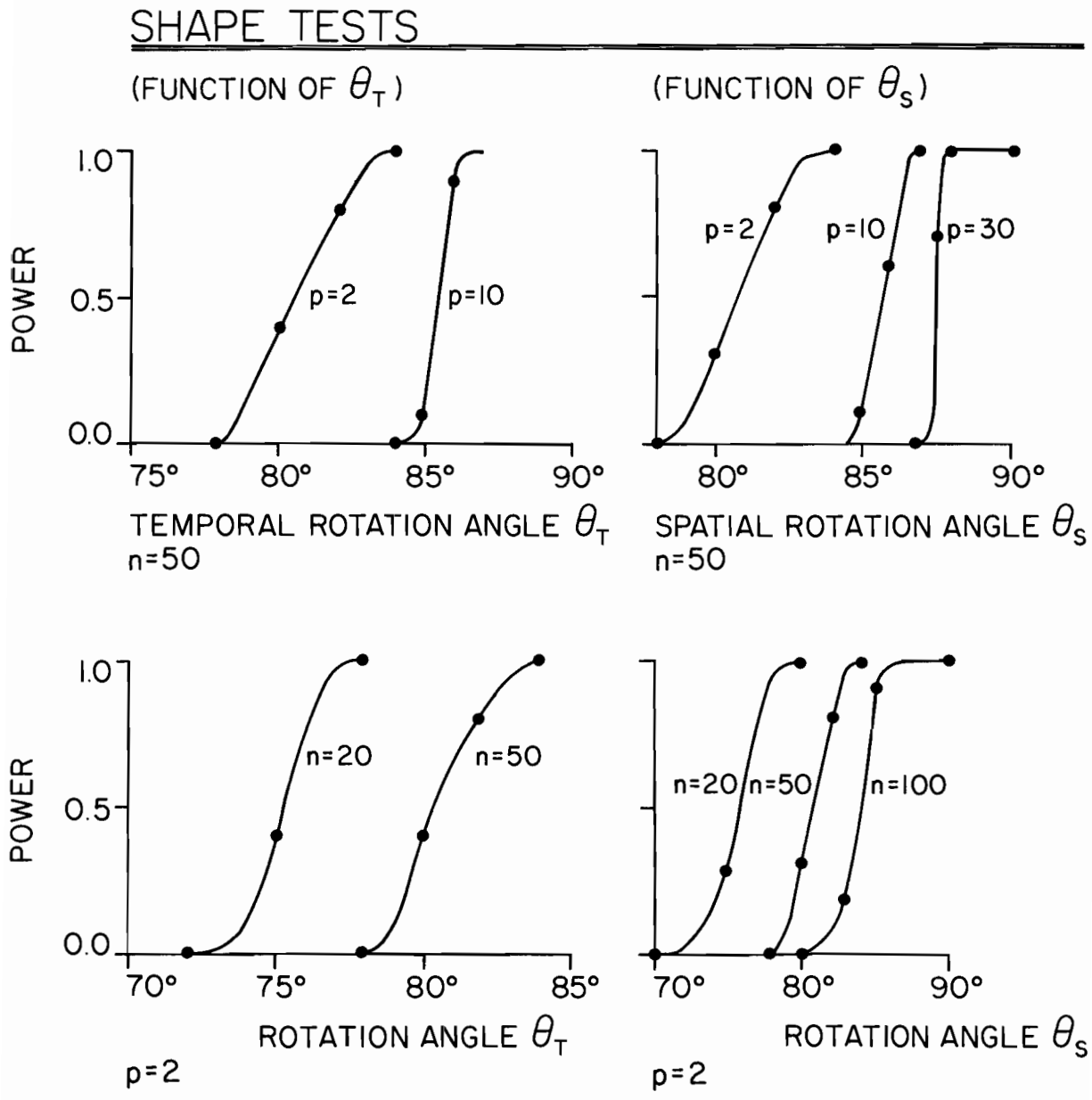

SHAPE TESTS
(FUNCTION OF θ_T)
(FUNCTION OF θ_S)
POWER
1.0
0.5
0.0
p=2
p=10
p=30
75°
80°
85°
90°
TEMPORAL ROTATION ANGLE θ_T
n=50
SPATIAL ROTATION ANGLE θ_S
n=50
n=20
n=50
n=100
70°
ROTATION ANGLE θ_T
p=2
ROTATION ANGLE θ_S
p=2

Fig. 8.4

D. Power Curves for SPRED when Sampling Populations are Purely Random

The differences in the populations drawn from were in their radii (variances). One was held fixed while the other was multiplied by a factor σ. SPRED power curves under various testing conditions are shown in Fig. 8.3. In the upper left panel we explore power as a function of σ for n = 50. Each curve is the result of a different choice of p. Notice how, in each curve, very little σ must be decreased before the power of SPRED reaches one. The '+' points are explained in the caption. Once again it is clear that in going from DIST2 to SPRED we have gained much power (compare the p = 2 curve in the upper left of Fig. 8.3 with the lower left curve of Fig. 8.1). Notice that the power of SPRED (for fixed σ) increases with p. This is inverse to the behavior of SITES with p (upper panel, Fig. 8.2). In the upper right panel of Fig. 8.3 we examine the power of SPRED when one of the spherical gaussian populations is expanded along a single direction by a factor e (ellipticity). Here, power (above a fixed e abscissa) decreases with increasing p. In the panel below, power of SPRED is seen to increase with n, for a given ellipticity e.

E. Power Curves for SHAPE when Sampling Populations are Purely Random

The SHAPE power curves are shown in Fig. 8.4. In this case the natural transformation needed to change the SHAPE of a swarm is not a centroid change nor a radial scale change. It is e.g., a rotation in p-space, or in n-space that is required. That is, if $\underline{M}$ is an n×p matrix, we can subject it to two different kinds of rotations.* For example, if $\underline{T}$ is an n×n orthogonal matrix then $\underline{T}\ \underline{M}$ has been rotated temporally away from $\underline{M}$. It is possible to characterize

* These will be considered in detail in DIT(III).

such a rotation by [n/2] rotation angles $\theta_1,\ldots,\theta_\ell$, $\ell = [n/2]$ within a suitable basis for E_n (here "[x]" denotes the largest integer in x). A simple instance of such a rotation is a *homogeneous* rotation wherein $\theta_1 = \cdots = \theta_\ell$. This is the kind of rotation used to produce the left panels in Fig. 8.4.

In more detail, the curves in the left panels of Fig. 8.4 were found as follows. First of all, a reference distribution for SHAPE was made by generating 100 realizations of n×p matrices $\underline{M}$ and $\underline{D}$. Each realization consisted of n samples from $N_p(\underline{0},\underline{I}_p)$. If $\underline{M}^{(i)},\underline{D}^{(i)}$ are the ith realizations of $\underline{M}$ and $\underline{D}$, then SHAPE($\underline{D}^{(i)},\underline{M}^{(i)}$) was obtained and all 100 of these values arranged in ascending order to form the reference cdf. The null hypothesis then is that $\underline{D},\underline{M}$ are drawn from the same population. Next, ten more pairs of realizations of $\underline{M}$ and $\underline{D}$ were made in the same way and the $\underline{M}$-member of each pair was rotated temporally a fixed amount to produce $\underline{T}\ \underline{M}$. (The rotation was homogeneous and for a given θ, as indicated in the graphs.) Each of the ten SHAPE($\underline{D},\underline{T}\ \underline{M}$) values for a given θ was compared to the upper 10% critical value of the reference cdf for SHAPE constructed above, and the number of rejections of H_o were noted. For example, when θ was chosen as 80° and we had p = 2, n = 50, it happened that 4 out of 10 times H_o was rejected. This formed the point (80°, 0.4) on the curve in the upper left panel of Fig. 8.4.

Notice how the power of SHAPE decreases drastically with increasing p or n. Even more basically, notice how far a homogeneous rotation of one of the populations must be carried out before the power of SHAPE rises to 0.5!

The second kind of rotation is spatial and is effected by p×p orthogonal matrices $\underline{S}$ such that $\underline{M}\ \underline{S}$. The rotation $\underline{S}$ may be characterized by [p/2] angles $\theta_1,\ldots,\theta_\ell$, $\ell = [p/2]$ (with respect to a suitable basis for E_p) and we may specialize these to homogeneous rotations. Such rotations were used on a gaussian population to separate it rotationally from another, exactly in the

manner described above for temporal rotations. Samplings were then made to produce the curves in the right panel of Fig. 8.4. Once again the power of SHAPE is disappointingly low.

Before drawing our final conclusion here, and in fairness to the SHAPE test, we should observe that it was examined under the most stringent conditions: the populations from which the $\underline{D}^{(i)}$ and $\underline{M}^{(i)}$ were drawn were *spherical*. Therefore, on rotating $\underline{M}^{(i)}$ temporally or spatially we were producing, statistically, a very slight difference. Fuller studies of SHAPE must still be made wherein the populations $N_p(\underline{0},\underline{\Sigma})$ have $\underline{\Sigma}$ other than $\underline{I}_p$. Undoubtedly the power of SHAPE will be greater, all other factors the same, when the ellipsoid associated with $\underline{\Sigma}$ has greater eccentricity. However, there are further indications, beyond the examinations made here, that SHAPE does indeed have relatively low-power tests. This matter will be discussed further in DIT(III).

In conclusion, the power of the SHAPE tests under random sampling conditions from spherical gaussian populations is low for both homogeneous temporal and homogeneous spatial rotations.

This conclusion has led to a search for more powerful tests than the SHAPE tests and which continue to look for differences in attributes other than location and scale. The search has resulted in the S-phase and T-phase tests, and their descendants, to be reported on in DIT(III). Consequently, no further discussion of the SHAPE statistic or its reference distributions is made in the present work.

F. Power Curves for SITES and SPRED via APP

Further power tests of SITES and SPRED were made using APP and PPP. Rather than using classical random samplings from gaussian populations, as in the preceding paragraphs, we generated a data set that represented a mix of deterministic and random activity. With APP and PPP involved, this was a somewhat more expensive undertaking than the earlier experiments. For this puspose we used the data generator described in Appendix A. The parameters n, p, α, along with μ, σ were selected, and corresponding data values for $\underline{D}$ and $\underline{M}(\mu,\sigma)$ generated. From these, 10 power curve trials were made to produce each point on the curves below. Ten trials constitute an experiment. Recall from §5 that APP also requires random permutations ϕ of the first n integers. One hundred permutations were required for each trial. These permutations were computed and stored for use prior to the tests. Moreover, the $\underline{R}_1,\underline{R}_2$ matrices (see Appendix A) were computed at the outset and used later when variations in μ and σ were needed. In this way (pre-computing and storing) we could gain a general initial impression of the power curves as functions of μ and σ without having to generate totally new realizations for permutations and random sets for each choice of μ and σ in each trial of the experiment. The resultant curves are still representative of the results obtainable under fully rigorous sampling conditions.

For those readers intending to reproduce our results and perhaps go beyond them, we summarize here the activity required to produce a typical power curve point in the present setting. Consider the n = 24 curve in Fig. 8.5(a). First of all, at the outset of our experiments, we constructed three master matrices: a 3600 × 24 matrix $\underline{F}$ generated as described in Appendix A, and two independently produced 3600 × 24 random matrices $\underline{R}_1,\underline{R}_2$ obtained by repeated random drawings from N(0,1). (These matrices were to be used also for the

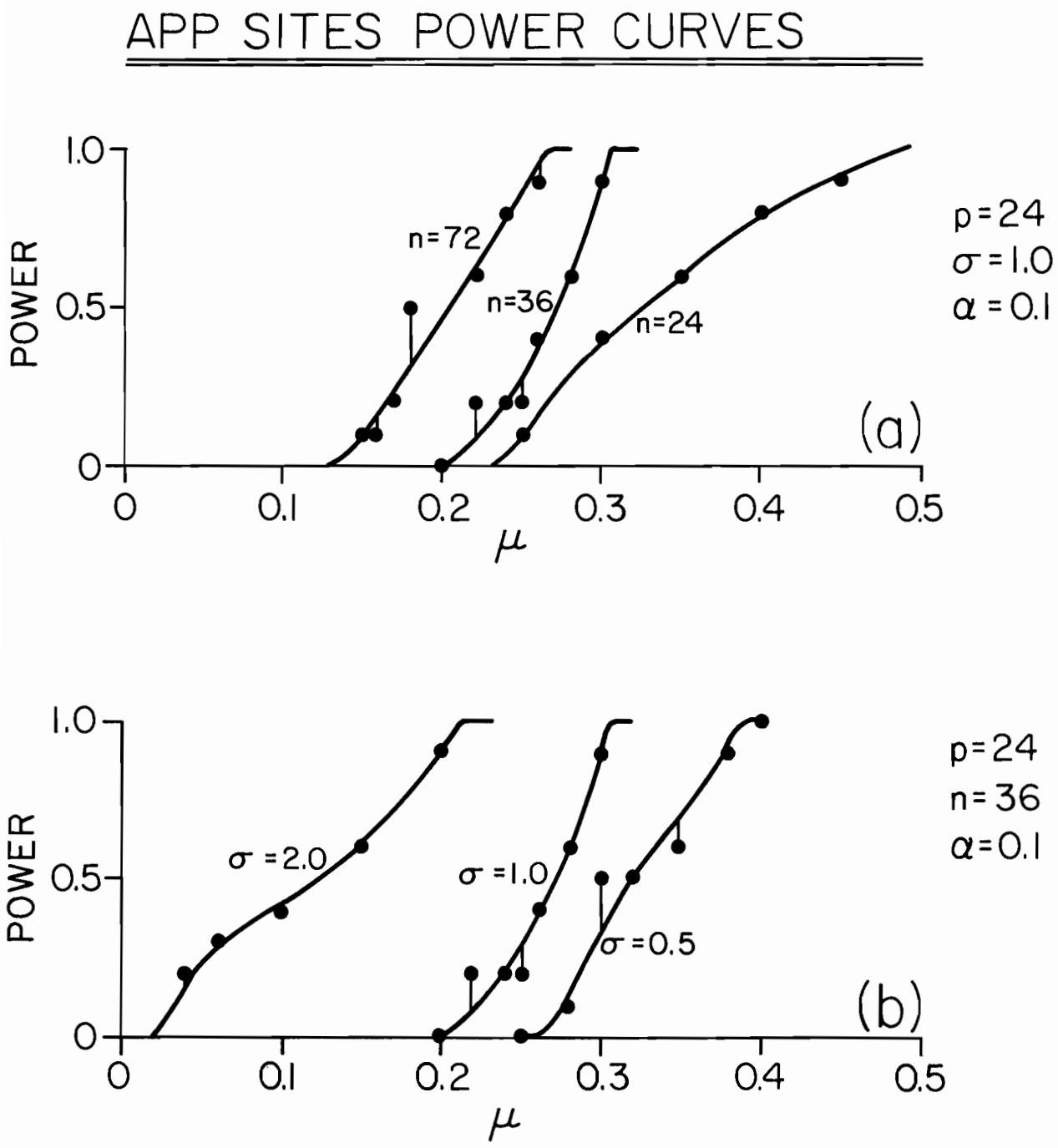

APP SITES POWER CURVES
POWER
1.0
0.5
0
n=72
n=36
n=24
p=24
σ=1.0
α=0.1
(a)
0
0.1
0.2
0.3
0.4
0.5
μ
POWER
1.0
0.5
0
σ=2.0
σ=1.0
σ=0.5
p=24
n=36
α=0.1
(b)
0
0.1
0.2
0.3
0.4
0.5
μ

Fig. 8.5

experiments in §9.) Next, setting $\sigma = 1.0$ and $\mu = 0.25$, and snipping off the top 240 rows of $\underline{F}$ and $\underline{R}_1, \underline{R}_2$, we found the 240×24 matrix $M^*(\mu,\sigma)$ as given in (A1.6), and the 240×24 matrix $\underline{D}^*$ as given in (A1.5). Starting at the top of $\underline{D}^*$ and $\underline{M}^*(\mu,\sigma)$ we successively separated off 24×24 disjoint matrices, obtaining in all ten 24×24 realizations of $\underline{D}, \underline{M}(\mu,\sigma)$. These were the basis for our ten trials. Each of the ten 24×24 matrix pairs $(\underline{D}, \underline{M}(\mu,\sigma))$ was put into the APP hopper for the statistic SITES. The result was one rejection of H_o, and nine acceptances. This resulted in the first point (0.25, 0.10) of the $n = 24$ curve in Fig. 8.5(a). All other points were found in exactly analogous ways. (For larger n values, we generated $10n \times 24$ matrices.) We now turn to a discussion of the results.

In Fig. 8.5(a) we have the power curves for SITES as generated by APP. Here $p = 24$, $\sigma = 1.0$ (in $\underline{M}(\mu,\sigma)$) and the significance level of the tests is $\alpha = 0.10$. Now, for $n = 24$, and a sequence of μ values the points of the curve $n = 24$ were generated (as described above). Each point is the result of 10 trials with 10 realizations of $\underline{D}$ and $\underline{M}(\mu,\sigma)$ using the pre-stored random matrices and permutations. It is seen that the power of SITES becomes non zero after $\mu = 0.20$, and rises to 1 by about $\mu = 0.50$. Once again, as in Fig. 8.2 (lower panel) power increases with increasing n.

In Fig. 8.5(b) some power curves for SITES via APP were generated for the three choices of σ. This represents a new wrinkle in the tests relative to the earlier ones. We wanted to see how SITES' power behaved when the radii of the swarms (as gauged by σ) changed. The power behavior is fairly smooth and it increases at a fixed μ as σ increases.

Fig. 8.6 shows SPRED power curves indexed by various n values. Comparison of this set of curves with those in the lower left panel of Fig. 8.3 is instructive.

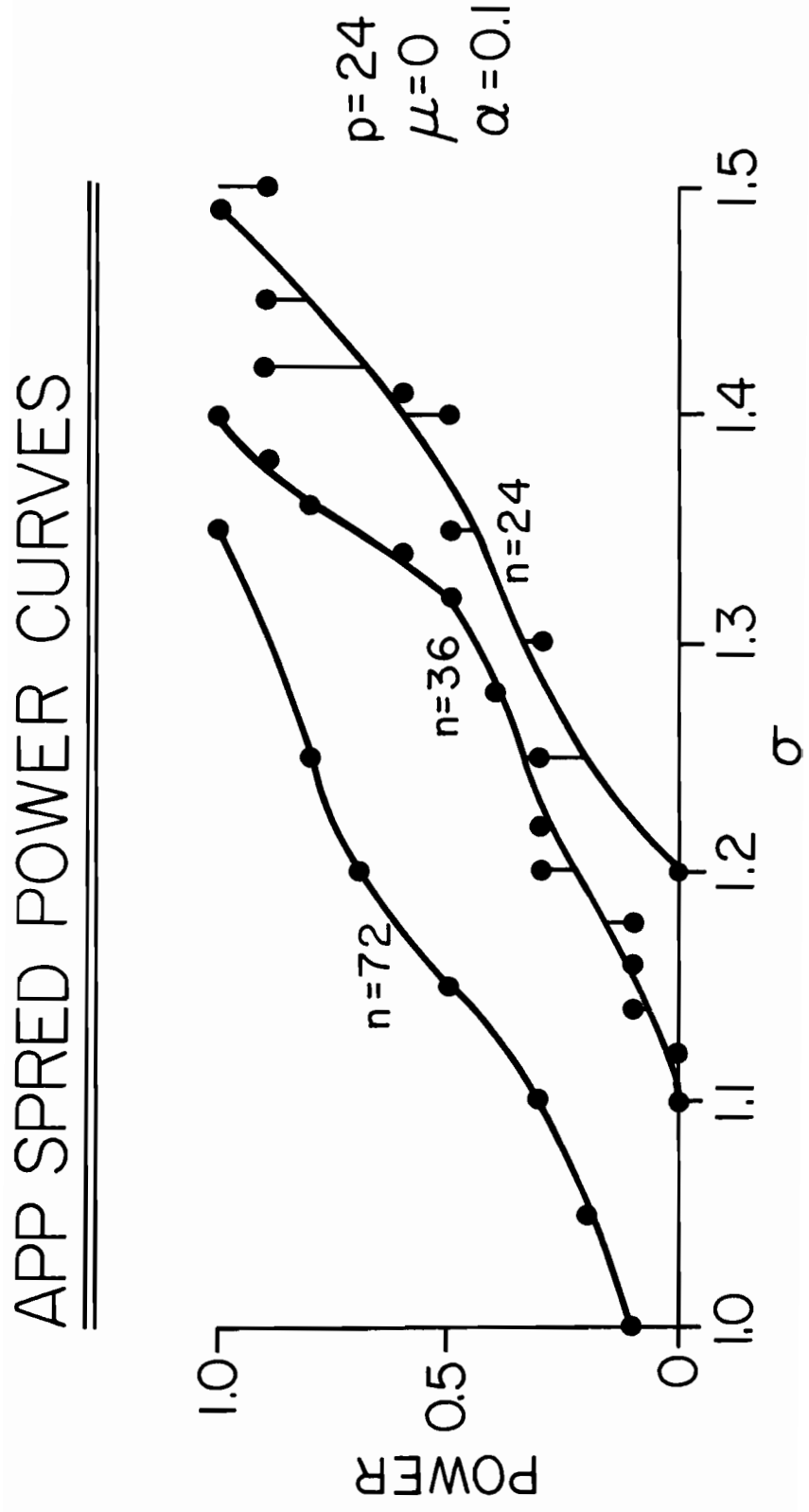

APP SPRED POWER CURVES
p=24
μ=0
α=0.1
n=72
n=36
n=24
POWER
1.0
0.5
0
1.0
1.1
1.2
1.3
1.4
1.5
σ

Fig. 8.6

The σ's of course are differently interpreted but observe the very much lower power of SPRED in the present more realistic sampling situation. There is even a Type I error in the n = 72 curve (i.e., H_o is true there, but it was rejected once in the 10 trials).

G. Power Curves for SITES and SPRED via PPP

The introductory remarks for the present set of curves are precisely parallel to those of APP in par F. Hence we may proceed directly to the discussion of the results. First of all we notice that Fig. 8.7 and Fig. 8.5 are constructed with exactly the same sets of parameters so as to facilitate intercomparisons of the two families of SITES power curves. Consider in particular those for n = 72 in the upper panels. Despite their different shapes, these two curves are fairly close in power content: they both start at about 0.12-0.14 and rise to unit power somewhere between 0.25, 0.30. Similar gross resemblances exist between the pairs for n = 36 and n = 24 in each Figure. We may thus conclude that APP and PPP produce SITES power curves of closely comparable strength. A similar conclusion is evident on comparing the two lower panels in Figs. 8.5, 8.7. The cross-over of the two curves in the lower panel of Fig. 8.7 (along with the plateau of points at 0.95 power) is most likely an artifact of the present somewhat economical power curve construction policy (explained in par F). The PPP power curves for SPRED are shown in Fig. 8.8. The curves of the top panel are comparable to those in Fig. 8.6. Once again the familial resemblance between the two sets of curves is evident and we conclude that APP and PPP generate SPRED power curves of comparable strength under the present test data conditions. The curves in the lower panel of Fig. 8.8 are the result of some curiosity on our part as to how

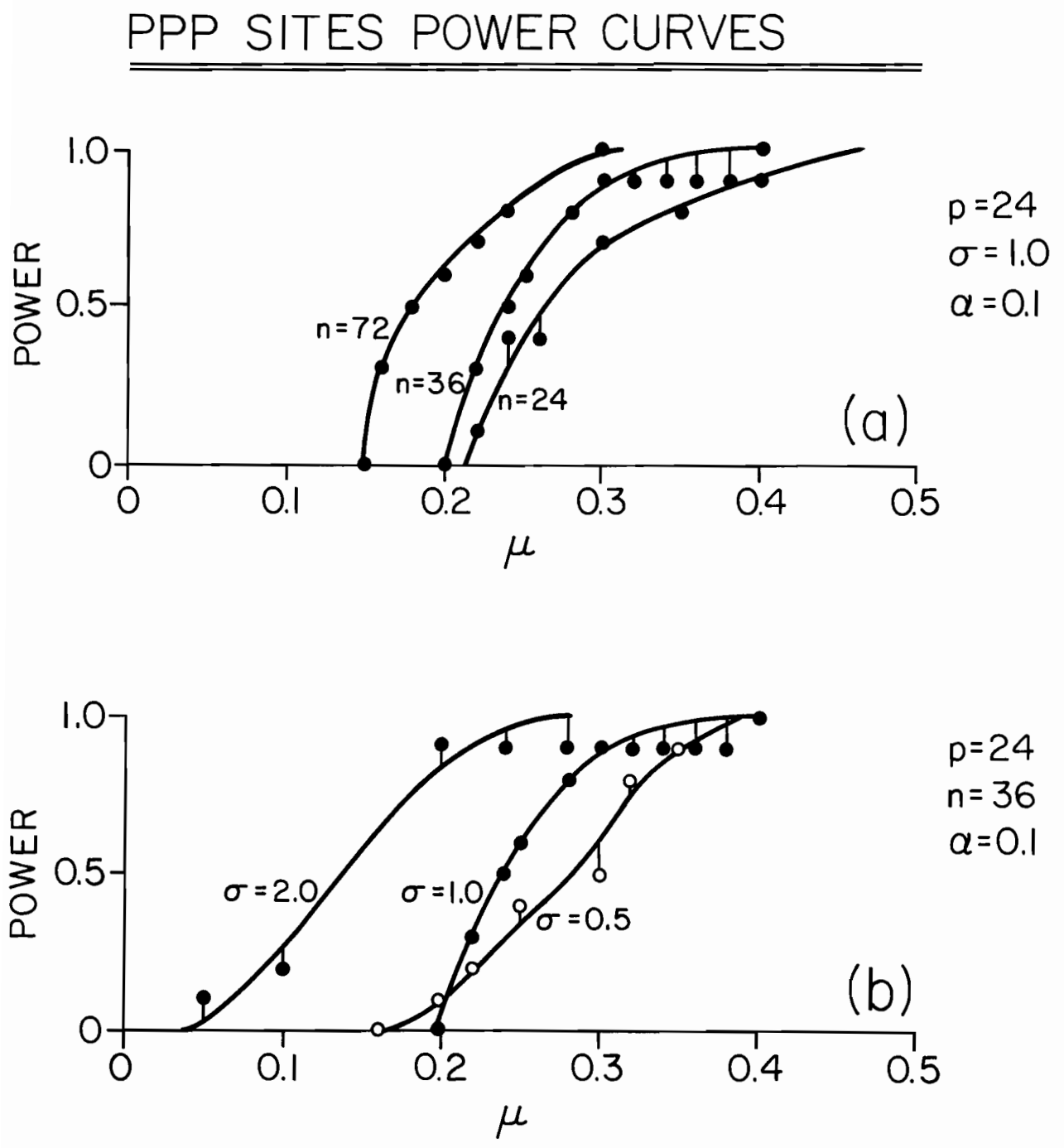

PPP SITES POWER CURVES
POWER
1.0
0.5
0
0
0.1
0.2
0.3
0.4
0.5
μ
n=72
n=36
n=24
(a)
p=24
σ=1.0
α=0.1
POWER
1.0
0.5
0
0
0.1
0.2
0.3
0.4
0.5
μ
σ=2.0
σ=1.0
σ=0.5
(b)
p=24
n=36
α=0.1

Fig. 8.7

PPP SPRED POWER CURVES

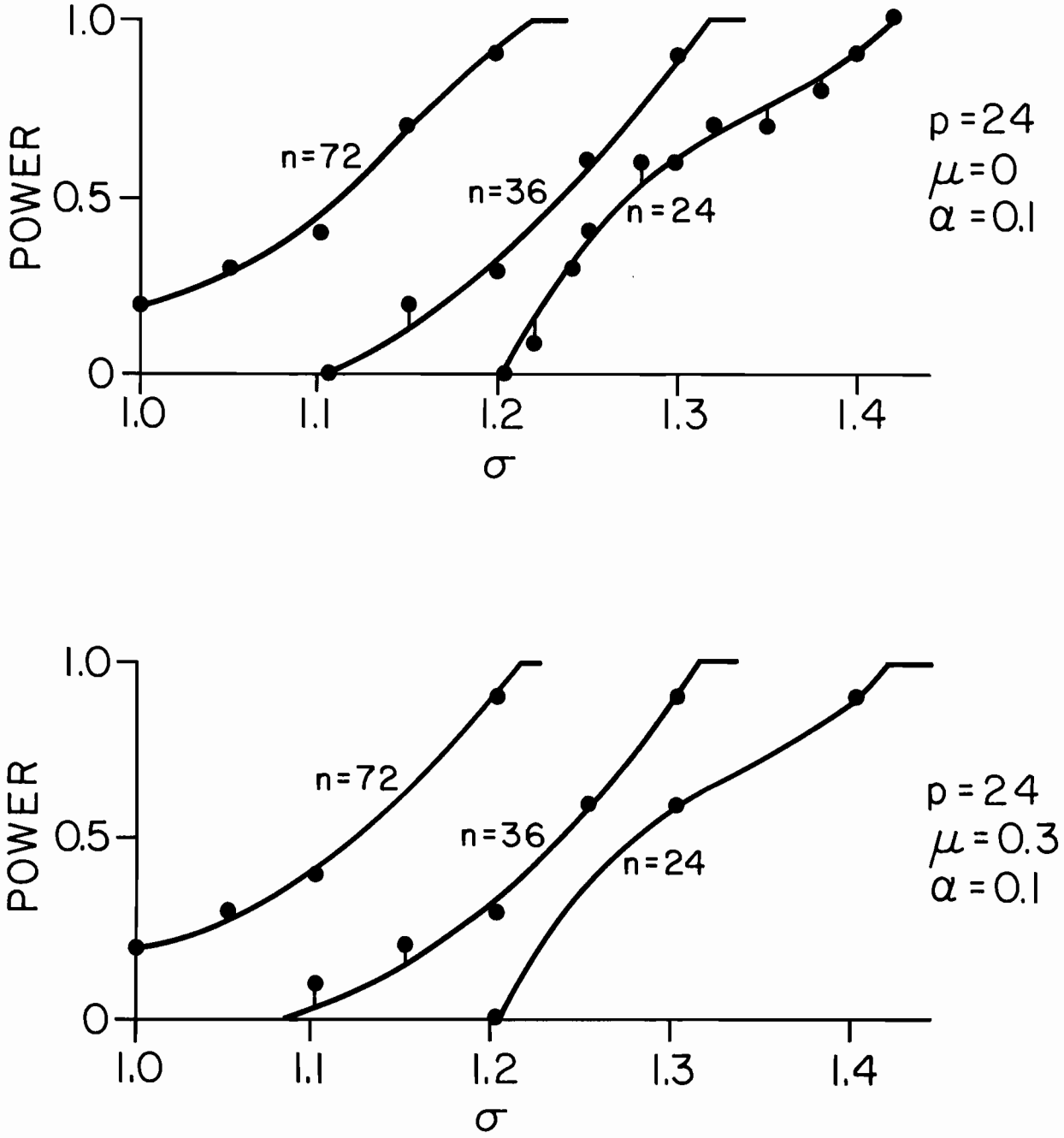


Fig. 8.8

SPRED would do if $\mu \neq 0$. The reader should note that there is no SPRED counterpart possible to such a curve in APP. Accordingly we set $\mu = 0.30$, and ran out the PPP-produced curves. These may be compared with those in the upper panel of Fig. 8.8. Only minor differences exist. Hence offsetting the two populations by $\mu = 0.30$ does not seem to affect PPP's SPRED power curves in any material way. A moment's thought (on reviewing the definitions of SPRED and PPP) shows that, in principle, such a μ-change can affect the SPRED power curves. But evidently under the present sampling conditions this effect is minor.

We conclude that the SITES and SPRED power curves for APP are comparable in strength to those for PPP under the present set of data conditions (Appendix A).

H. Power Curves for some Classical Statistics for Location (T^2) and Spread (M/b)

We examined the power of the classical T^2 test for location (Anderson, 1958; see also Appendix B in DIT(I)) using the same artificial data set as in pars F and G. In this way we could compare, under identical working conditions, the T^2 power curves with the APP- and PPP-produced power curves for SITES. The classical power curves are shown in Fig. 8.9(b). These are to be compared to those in Fig. 8.5(b) (for APP) and Fig. 8.7(b) (for PPP). At first glance, the T^2 curves seem to have higher power than their APP and PPP counterparts (see, e.g., the curve labeled "$\sigma = 2.0$"). But this high power is spurious since it is bound up with high Type I errors near $\mu = 0$. The classical T^2 test is known to have lower power and erratic behavior when the covariances of the two populations are not alike (hence the anomalous behavior of the $\sigma = 0.5$, 2.0 curves in Fig. 8.9(b)). When $\sigma = 1.0$, as in Fig. 8.9(a), all is well and the curves of the T^2 test are competitive with the $n = 72,36$ curves in Figs. 8.5(a), 8.7(a). However the T^2 test again has problems near $n = 24$ (where it doesn't

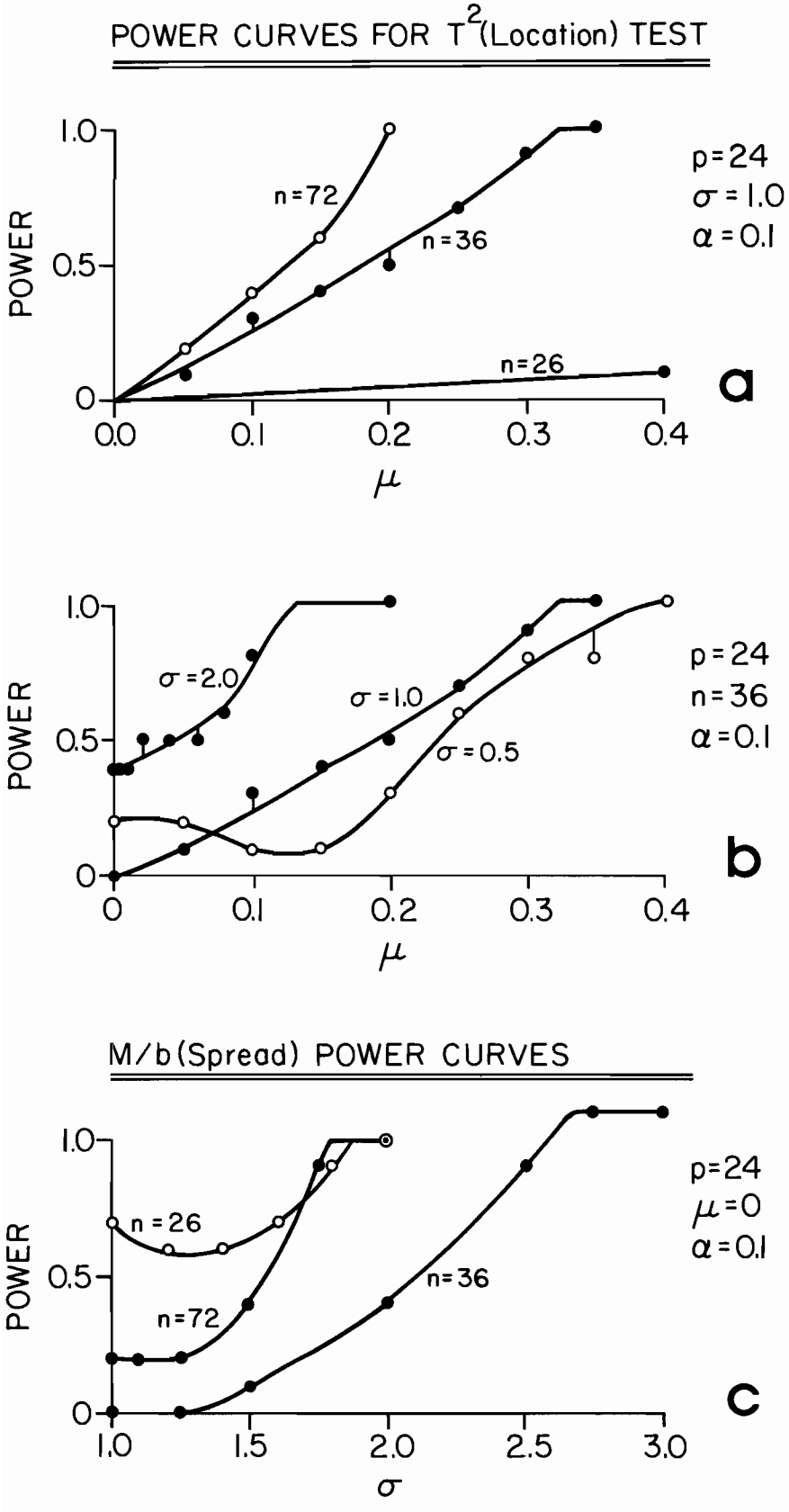

POWER CURVES FOR T^2(Location) TEST
POWER
1.0
0.5
0
n=72
n=36
n=26
p=24
σ=1.0
α=0.1
a
0.0
0.1
0.2
0.3
0.4
μ
σ=2.0
σ=1.0
σ=0.5
p=24
n=36
α=0.1
b
M/b(Spread) POWER CURVES
n=26
n=72
n=36
p=24
μ=0
α=0.1
c
1.0
1.5
2.0
2.5
3.0
σ

Fig. 8.9

exist). *In sum, we would generally tend to trust the APP and PPP SITES tests more than the classical T^2 test for location.*

Finally we looked at a much-cited classical test for detecting spread differences in data sets (Pearson, 1969; Korin, 1969; Box, 1949). Before we ran the power curves for this test we examined for validity the basic approximations underlying the test. First of all, the test is built around the statistic M where

$$M \equiv (n-1) \log |\underline{S}| - \sum_{i=1}^{k} N_i \log |\underline{S}_i| \tag{8.1}$$

and where $N_i = n_i - 1$, and the bars denote determinants. For our purposes, $k = 2$, since we are comparing two populations with sample covariance matrices $\underline{S}_1, \underline{S}_2$. Moreover, we set $n_1 = n_2 = n$. Under the hypothesis that we are sampling from the same population to form $\underline{S}_1, \underline{S}_2$, it is asserted (see above references) that, approximately,

$$M(1-A_1) \sim \chi_f^2 \tag{8.2}$$

where

$$A_1 = \frac{(2p^2+3p-1)(k+1)}{6(p+1)k(n-1)}$$

$$f = \tfrac{1}{2}p(p+1)(k-1)$$

It is claimed that this approximation to χ_f^2 is good for lower values of p. Since, often in oceanography and meteorology our range of p values is typically 24 or more, we were interested in the validity of the approximation (8.1) in this range. Accordingly, we conducted a Monte Carlo experiment in which $n = 26$, $k = 2$, and for values $p = 6, 12, 18, 20, 24$. The $\underline{M}$ and $\underline{D}$ matrices were obtained by random samplings from N(0,1). These matrices were then used

to determine $\underline{S}_1$ (for $\underline{D}$) and $\underline{S}_2$ (for $\underline{M}$), and from which $M(1-A_1)$ was found along with f. In each choice of p, fifty $M(1-A_1)$-values were found and we determined how many of these exceeded the 5% level of the χ^2_f distribution (we would expect 2.5 on the average). The table below gives the percentage of the 50 values exceeding the 5% critical value of χ^2_f.

p	6	12	18	20	24
% of $M(1-A_1)$ values exceeding 5% critical value of χ^2_f	4%	6%	10%	24%	98%

Thus for $p > 12$ the χ^2 approximation is not useful.

There is an associated test for spread differences using the F distribution, which was approximated using the same techniques as for the χ^2 distribution. Presumably, therefore, that too becomes useless for the range of p values usually occurring in oceanographic and meteorologic research (cf, e.g., Pearson, 1969, p. 220). Nevertheless, we computed some power curves for that statistic, which has the form:

$$M/b \sim F_{f,g} \tag{8.3}$$

where

$$g = (f+2)/(A_2-A_1^2)$$

$$A_2 = \frac{(p-1)(p+2)(k^2+k+1)}{6k^2(n-1)^2}$$

$$b = f_1[1-A_1-f/g]^{-1}$$

Once again we set $k = 2$ and all data sets were given the same $n_i = n$, $i = 1, 2$. Here $F_{f,g}$ is the F-distribution with f,g degrees of freedom (in the references above $f_1 = f$, $f_2 = g$).

Fig. 8.9(c) shows the results of power curve computations based on $\underline{D}$ and $\underline{M}(\mu,\sigma)$ of Appendix A, for the case p = 24, μ = 0, and significance level α = 0.10. Three choices of n are shown over a range of σ values. Notice the high Type I error at μ = 0 for n = 26. This set of curves may be compared to their APP and PPP counterparts in Figs. 8.6, 8.8. *Since APP and PPP generally agree on SPRED's power for the present data set (Appendix A), we conclude that the M/b statistic is inadequate to gauge significant spread differences between data sets with n and p values in the range of geophysical interest.*

I. On the Use of Power Curves in Practice

Our present approach toward power tests and their use may now be stated. When a new statistic (e.g. SITES, SPRED) is defined, and a means (e.g. APP, PPP) of constructing reference distributions for this statistic is devised, and a particular family of n×p data matrices (e.g. $\underline{D}_o,\underline{M}_o$) is to be intercompared, then proceed as follows. Construct an artificial-data-set family $\underline{D},\underline{M}(\mu,\sigma)$ in which n,p match the given n,p. Moreover, μ,σ are chosen such as to yield sets of centroids and spreads that are representative of and *embed* those of the given $\underline{D}_o,\underline{M}_o$ sets. Next, produce SITES and SPRED power curves via APP, PPP at some significance level α. Then choose the procedure (APP, PPP, e.g.) that produces for SITES the higher-power curves; and also the procedure (APP, PPP, e.g.) that produces for SPRED the higher power curves. Use these procedures on $\underline{D}_o,\underline{M}_o$ to reach the conclusions about significant values of SITES ($\underline{D}_o,\underline{M}_o$) and SPRED ($\underline{D}_o,\underline{M}_o$).

Knowledge of the approximate power of these SITES and SPRED statistics allows an estimate of the risk and cost of a conclusion based on a significance decision.* This kind of thinking on the part of oceanographers and meteorologists

---

* See, e.g. (Crow, Davis, Maxfield, 1960)

may yet be a long way off. But someday their understanding of air/sea/land/cryosphere interactions will have matured. Then predictions will have to be made of the values of important climatological variables bearing on fuel and food supplies, and people will want to know the risks and costs of a wrong decision. In the meanwhile, good statistical techniques, perhaps such as those studied here, or their modifications, will aid in formulating dynamical hypotheses about how these interactions work.

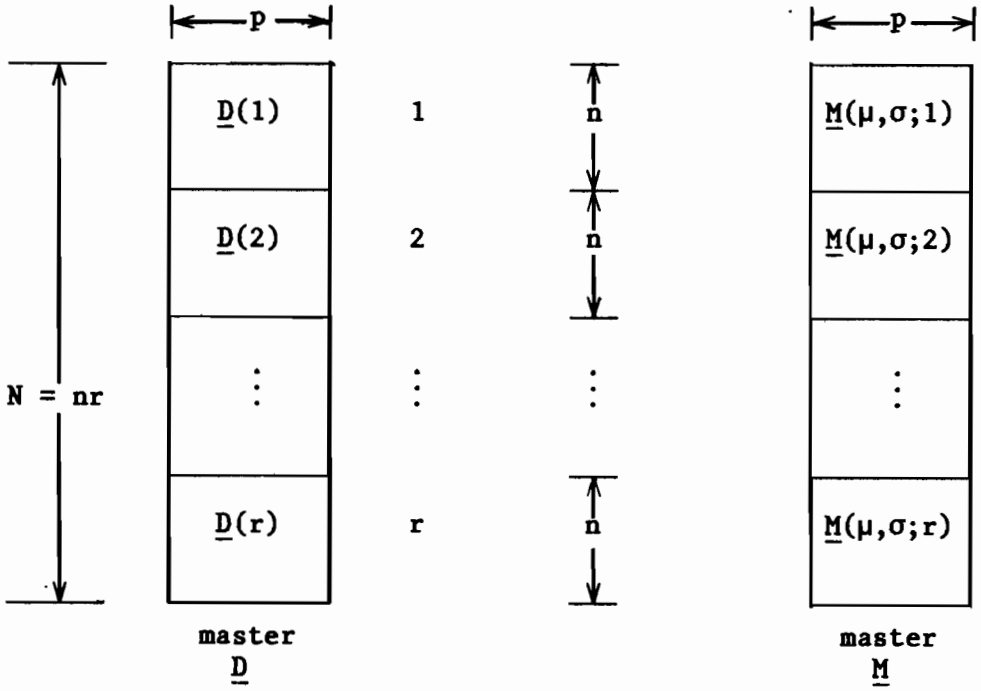
9. Comparison of APP and PPP Distributions with Reference MCP Distributions

A. Introduction

Suppose we had an adequate data setting such as that in §3, and we extracted from it relatively small n×p sample realizations of $\underline{D}$ and $\underline{M}$. Suppose further that we applied APP or PPP to these extracted $\underline{D}$,$\underline{M}$ and gauged their separation with respect to either SITES or SPRED. In these comparison procedures, as a matter of course, cumulative distribution functions (cdfs) are produced for SITES or SPRED. *The question then arises as to the visual comparison of these APP- (or PPP)-produced cdfs with those cdfs that would be produced by straightforward Monte Carlo procedures on the entire adequate data set.* In this section we present the results of a detailed study of this question. Our aim is to see how the APP- (or PPP-)produced cdfs compare (in the sense of location and dispersion) with those that would be generated using a relatively "exact" method, such as the Monte Carlo procedure (MCP) under adequate data conditions.

B. Definition of an Adequate Setting

As a preliminary step to the cdf constructions, we define the present form of an adequate data setting. We generate n×p matrix fields $\underline{F}, \underline{R}_1, \underline{R}_2$ and form the resultant n×p matrices $\underline{D}$ and $\underline{M}(\mu,\sigma)$, as described in Appendix A. We chose N quite large--3600 in this case--and then extracted many realizations of n×p data sets from these master sets. The master sets can be schematically visualized as sketched, below:



For our experiments we will choose n and r combinations shown in the table:

n	r
24	100
36	100
48	75
72	50
180	20

Except for the first combination of n and r in the table, we used all 3600

rows of the master matrix $\underline{D}$ and master matrix $\underline{M}$ in each experiment (so that $N = nr$). If each row of the master matrices $\underline{D}$ and $\underline{M}$ represented monthly averages of some field over p points, then these matrices would summarize 300 years of data. These data were generated by a statistically stationary process, as described in Appendix A, which is part deterministic, part random.

C. Comparing APP and MCP cdfs

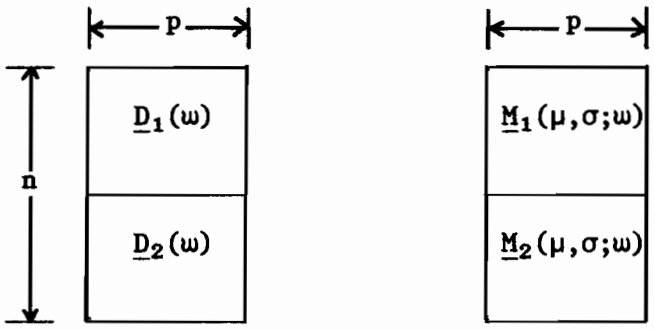
The results of our calculations are summarized in Figs. 9.1 to 9.6. The graphs intercompare the cdfs generated by APP and a Monte Carlo procedure (MCP). There were 24 runs made in all, and these are arranged 4 to a figure, with the run-number shown on each diagram along with the identification of the parameters chosen for each run. Thus, the symbols "MU," "SIGMA" and "COEF1" refer respectively to the parameters μ,σ and a, and which are defined in Appendix A. μ governs the centroid separation, while σ governs the radial scale, or spread separation of the data sets. Moreover the symbols "N," "P" and "R" refer respectively to the parameters n, p and r as also used in Appendix A and this section. Each run was executed in the following eight steps:

(i) Choose n and r (see table in par B above)

(ii) Choose μ,σ

(iii) For a given ω, $1 \leq \omega \leq r$, construct the $n \times p$ matrices $\underline{D}(\omega)$, $\underline{M}(\mu,\sigma;\omega)$ (see the diagram in par B above)

(iv) Partition $\underline{D}(\omega)$, $\underline{M}(\mu,\sigma;\omega)$ using a random permutation ϕ in accordance with APP (§5). The results are schematically defined as in the following figure



The identity transformation ϕ leaves $\underline{D}$ and $\underline{M}$ unchanged and as they appear in the above sketch. This fact will be used in (viii) below.

(v) For each ω in (iv) find the cdfs of SITES (or SPRED) using APP (§5). In particular, use $\underline{D}_1(\omega)$, $\underline{M}_2(\mu,\sigma;\omega)$ to generate the "DM" cdf of the APP.

(vi) Repeat (v) for $\omega = 1,\ldots,10$

(vii) Plot all 10 cdfs found in (vi)

(viii) Generate a ***Monte Carlo Procedure*** (MCP) cdf as follows: Find SITES (or SPRED) for each pair $\underline{D}_1(\omega)$, $\underline{M}_2(\mu,\sigma;\omega)$, $\omega = 1,\ldots,r$. These are the results of the identity permutation on $\underline{D}(\omega)$ and $\underline{M}(\mu,\sigma;\omega)$. Order these SITES (or SPRED) numbers and thereby find the ***resultant MCP cdf*** associated with the current choice of n, r, μ, σ.

The numbers N = 3600 and p = 24 are fixed throughout all the calculations.

The 24 runs in Figs. 9.1-9.6 fall naturally together into subgroups. The following Table gives an overview of the runs and the subgroups.

Table 9.1 Parameters for APP/MCP Runs 1-24 in Figs. 9.1-9.6

RUN	DATA SETS	STATISTIC	n	p	r
1	$\mu = 0, \sigma = 0, a = 0$	SPRED	24	24	100
2			180	24	20
3	$\mu = 0, \sigma = 1.25, a = 1$	SPRED	24	24	100
4			36	24	100
5			48	24	75
6			72	24	50
7			180	24	20
8	$\mu = 0, \sigma = 1.25, a = 1$	PSPRED	24	24	100
9			36	24	100
10			48	24	75
11			72	24	50
12			180	24	20
13	$\mu = 0, \sigma = 0, a = 0$	SITES	24	24	100
14			180	24	20
15	$\mu = 0.3, \sigma = 1.0$	SITES	24	24	100
16			36	24	100
17			48	24	75
18			72	24	50
19			180	24	20
20	$\mu = 0.3, \sigma = 1.0$	PSITES	24	24	100
21			36	24	100
22			48	24	75
23			72	24	50
24			180	24	20

One point of interest of the study was to see how changes in the parameter n--presumably increases in n--would improve the resemblances between the APP cdfs and the MCP cdfs. The way we have arranged the n, r variations is to start with a relatively small n and a large r. Thus in run 1 we have n = 24 and r = 100. We would expect in this circumstance to have a well-stabilized MCP cdf of SPRED, for there are r = 100 realizations of each $\underline{D}(\omega)$ and $\underline{M}(\mu,\sigma;\omega)$, $\omega = 1,\ldots,100$ from which to form the MCP cdf. On the other hand the APP uses a 24×24 version of $\underline{D}$ and $\underline{M}(\mu,\sigma)$ and works *within them* to generate the cdf for SPRED. This was done 10 times. These 10 APP-produced curves are displayed in

run 1 of Fig. 9.1. Nestled among these 10 curves is the MCP curve. Notice that in this run we are sampling from a spherical gaussian population located in E_{24} (since $\mu = 0$, $\sigma = 0$ and $a = 0$). When we increase the sample size n to 180 as in run 2, the number of Monte Carlo replications r necessarily reduces to 20. Now we would expect the APP to provide a well-sampled estimate of the SPRED cdf, while the MCP SPRED cdf would be less stable. It appears, however, that both types of curve are still in quite good agreement (note the scale of the abscissas in runs 1 and 2). *Hence, our first impression is that there is close agreement between APP and MCP SPRED cdfs over a wide range of random sampling conditions.*

The family of five runs 3 to 7 for APP- and MCP-produced SPRED cdfs is done for $\mu = 0$, $\sigma = 1.25$, $a = 1$. Now we have the deterministic field $\underline{F}$ taking part in helping to produce data variance. The MCP SPRED cdf continues to be nestled among the 10 samples of the APP SPRED cdf, from the smallest sample $n = 24$ to the largest $n = 180$. *We would then conclude that under the present mixed deterministic and random conditions the good agreement continues to hold between the APP and MCP SPRED cdfs.*

In runs 8 to 12 we use the *physical SPRED* statistic, namely

$$\text{PSPRED} \equiv |\sigma'_D - \sigma'_M| \tag{9.1}$$

where now

$$\sigma'_D = \left\{ \frac{1}{np} \sum_{t=1}^{n} \sum_{x=1}^{p} (d(t,x) - d(x))^2 \right\}^{\frac{1}{2}}$$

(so that $\sigma'_D = \sigma_D/(np)^{\frac{1}{2}}$). The intent was to see the numerical range of PSPRED values under the various sampling conditions without the presence of the sampling dependent normalizing term $\sigma_D\sigma_M$. On the basis of runs 3 to 7, we

would expect the location and dispersion of the APP and MCP cdfs to once again be closely related, i.e., that the curves closely intermingle, and have the same shape, and this is borne out.* Notice how in runs 8 to 12, the 0.50 level of the APP and MCP cdfs (the median of the density distribution) stays sensibly constant (around PSPRED = 0.08) while the dispersions of the APP cdfs (their individual left to right extents) considerably contract as n goes from 24 to 180, as would be expected under larger and larger samplings of a homogeneous population. The MCP cdfs, interestingly, undergo about the same dispersion contraction as the replication number r decreases from 100 to 20. *Thus even in the absence of the denominator product* $\sigma_D\sigma_{M'}$, *the APP-produced PSPRED cdf has a close resemblance to its MCP-produced counterpart, as regards location and dispersion properties under a wide range of sampling conditions.*

Runs 13 to 24 repeat the SPRED study in all details but now for SITES. Thus in runs 13, 14 we have cdfs generated by purely random noise, with good agreement in the variously produced cdfs for SITES.

In runs 15 to 19, where now $\mu = 0.300$, $\sigma = 1.000$, and $a = 1$, we observe a migration of both MCP and APP SITES cdfs down the SITES scale from about 0.20 for n = 24, to 0.08 for n = 180. At the beginning of the sequence, the APP curves are somewhat displaced from the reference MCP curve, although they have about the same shape. At the end of the run the agreement between the cdfs is quite good, after the 10 APP curves have essentially migrated as a group into the vicinity of the MCP reference curve.

---

* It may then be asked, "If PSPRED seems to be so intuitively close to SPRED, why not just choose PSPRED as a statistic to work with, since it is simpler?" Unfortunately, this is not possible. In some experiments with PSPRED, and particularly with PSITES, defined in (9.2), below, the APP is adversely affected. It was found that the DD and MD curves could reverse order, so that MD<DD, in the PSITES context. (This follows from the possibility in the PSITES context of having DD<<MM.)

In runs 20 to 24, we repeat the setting of runs 15 to 19, but now with the *physical SITES* statistic, namely

$$\text{PSITES} \equiv \left\{ \frac{1}{p} \sum_{x=1}^{p} (d(x) - m(x))^2 \right\}^{\frac{1}{2}} \tag{9.2}$$

where $d(x) \equiv n^{-1} \sum_{t=1}^{n} d(t,x), \quad m(x) = n^{-1} \sum_{t=1}^{n} m(t,x)$

In this way we can check on our expectation that for large sample size n, PSITES should center around 0.300 (since $\mu = 0.300$), which it does for both APP and MCP SITES cdfs. *Once again, now for PSITES, agreement in location of the APP-produced cdfs improves with increasing n and, throughout, the shapes of the cdfs remain remarkably alike.*

D. Comparing PPP and MCP cdfs

The results of our calculations are summarized in Figs. 9.7 to 9.12. This series of graphs intercompares the cdfs generated by PPP and a Monte Carlo Procedure (MCP). There are 24 runs in all: from 25 to 48, inclusive. The parameters used are largely those used for the APP study, and are summarized by Table 9.2.

Table 9.2 Parameters for PPP/MCP Runs 25-48 in Figs. 9.7-9.12

RUN	DATA SETS	STATISTIC	n	p	r
25	$\mu = 0, \sigma = 0, a = 0$	SITES	24	24	100
26			180	24	20
27	$\mu = 0.3, \sigma = 1.0$	SITES	24	24	100
28			36	24	100
29			48	24	75
30			72	24	50
31			180	24	20

32	$\mu = 0.3, \sigma = 1.0$	PSITES	24	24	100
33			36	24	100
34			48	24	75
35			72	24	50
36			180	24	20
37	$\mu = 0, \sigma = 0, a = 0$	SPRED	24	24	100
38			180	24	20
39	$\mu = 0, \sigma = 1.25$	SPRED	24	24	100
40			36	24	100
41			48	24	75
42			72	24	50
43			180	24	20
44	$\mu = 0, \sigma = 1.25$	PSPRED	24	24	100
45			36	24	100
46			48	24	75
47			72	24	50
48			180	24	20

On turning to the graphs we see that runs 25, 26 *show good agreement between the MCP and PPP SITES cdfs, over the entire range.*

In the sequence of runs 27 to 31 we see that there is a systematic difference in location of the MCP and PPP SITES cdfs. Whereas in the corresponding APP runs 15 to 19 the two classes of cdfs intermingled, here they remain apart over the whole range of n samples, and the gap is insensitive to n. The shapes of the two types of curves, however, are alike and they are broad for small n, narrow for large n. It is this feature of like shape for APP and MCP curves that is important to maintaining the power of the PPP at a workably high level, relative to the reference curves of MCP.

The disparity between the PPP and MCP SITES cdfs is maintained in the run sequence 32 to 36 (note the change in abscissa scale). Observe how the MCP curve homes in on PSITES = 0.30, as expected, but the PPP settles at a definitely smaller estimate of PSITES. While there was some discrepancy in PSITES locations of the APP and MCP cdfs (runs 20 to 24), it decreased there with increasing sample size, as expected, unlike the present discrepancy.

In runs 37, 38 (a random setting) we have good agreement of SPRED cdfs as produced by MCP and PPP. The SPRED statistic's cdf tends to have a long right tail as found in either procedure. For larger samples such as n = 180, the natural tendency of the cdfs is to move nearer to the origin, with median values around 0.001, which therefore are down by an order of magnitude from the n = 24 case.

The cdf graphs of the SPRED runs 39 to 43 display a dramatic difference (relative to those witnessed in the preceding 38 runs) in the locations of the two cdfs. The MCP cdf medians migrate slightly, as n goes from 24 to 180, from about 0.005 to 0.007. By contrast, the PPP cdf medians move slightly toward the origin, widening the gap between the two families of cdfs. The runs throughout this sequence accordingly show two very different kinds of cdf belonging to the MCP and the PPP.

In runs 44 to 48 a general repetition of the preceding five runs is observed, now for the PSPRED statistic, showing once again disparate PSPRED cdfs produced by PPP and MCP remaining disparate throughout the entire sequence.

In sum, the APP distributions for both SITES and SPRED generally shared the same location and dispersion properties as the relatively trustworthy MCP distributions. The closeness of resemblance increased with increasing sample size n of $\underline{D}$ and $\underline{M}(\mu,\sigma)$ over the range of n, p, μ and σ values studied. On the other hand, the PPP distributions for both SITES and SPRED were generally of different location than the MCP distributions. The PPP SPRED distributions were, in addition, of different dispersion than the MCP distributions. In general, on the basis of these intercomparisons, we conclude that, under similar n, p, μ, σ, and statistically-stationary adequate-sampling conditions, the APP cdfs will more closely resemble the MCP (and presumably the 'true') cdfs than the PPP cdfs.

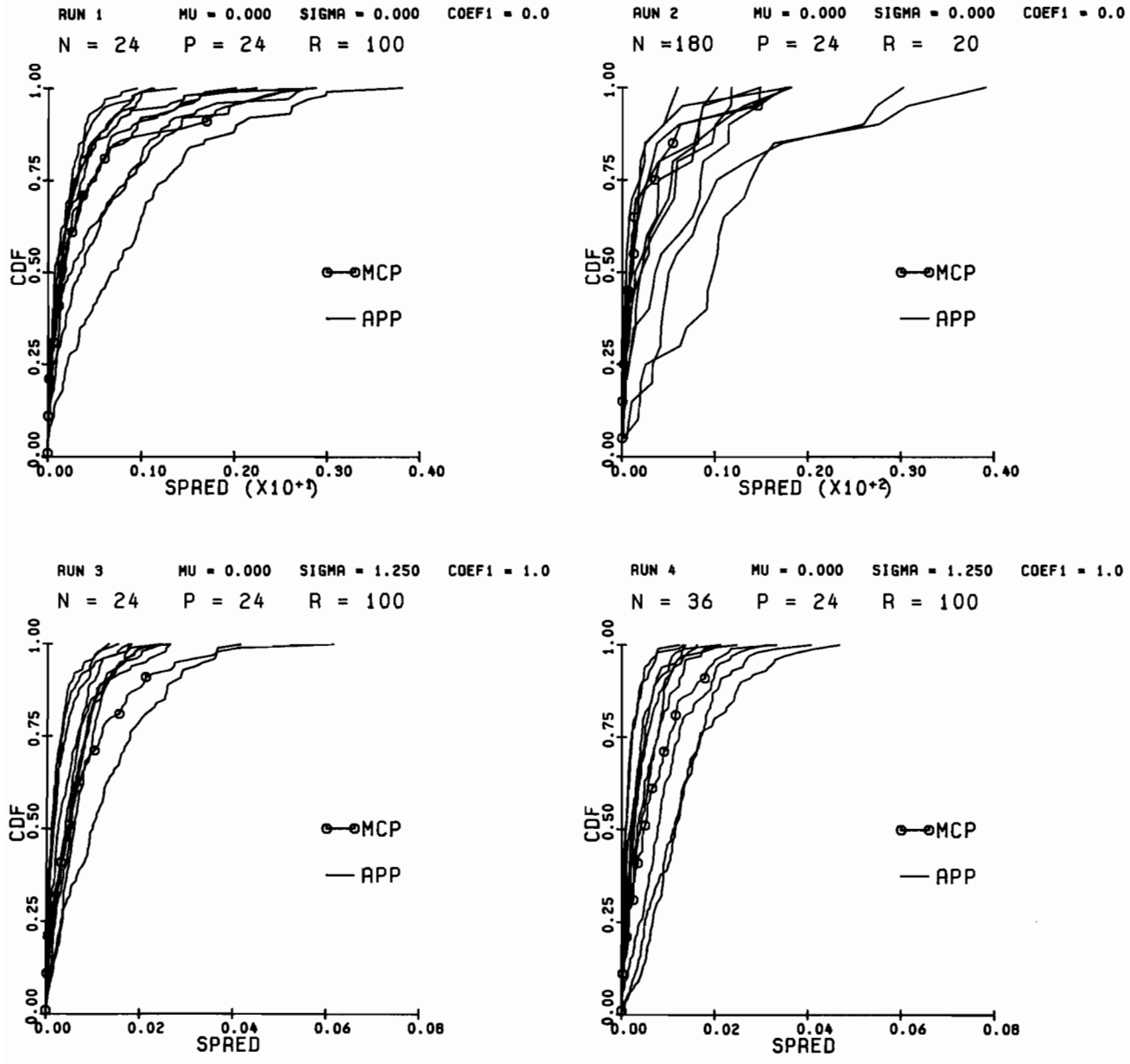

RUN 1 MU = 0.000 SIGMA = 0.000 COEF1 = 0.0
N = 24 P = 24 R = 100
CDF
SPRED (X10+1)
MCP
APP
RUN 2 MU = 0.000 SIGMA = 0.000 COEF1 = 0.0
N =180 P = 24 R = 20
CDF
SPRED (X10+2)
MCP
APP
RUN 3 MU = 0.000 SIGMA = 1.250 COEF1 = 1.0
N = 24 P = 24 R = 100
CDF
SPRED
MCP
APP
RUN 4 MU = 0.000 SIGMA = 1.250 COEF1 = 1.0
N = 36 P = 24 R = 100
CDF
SPRED
MCP
APP

Fig. 9.1

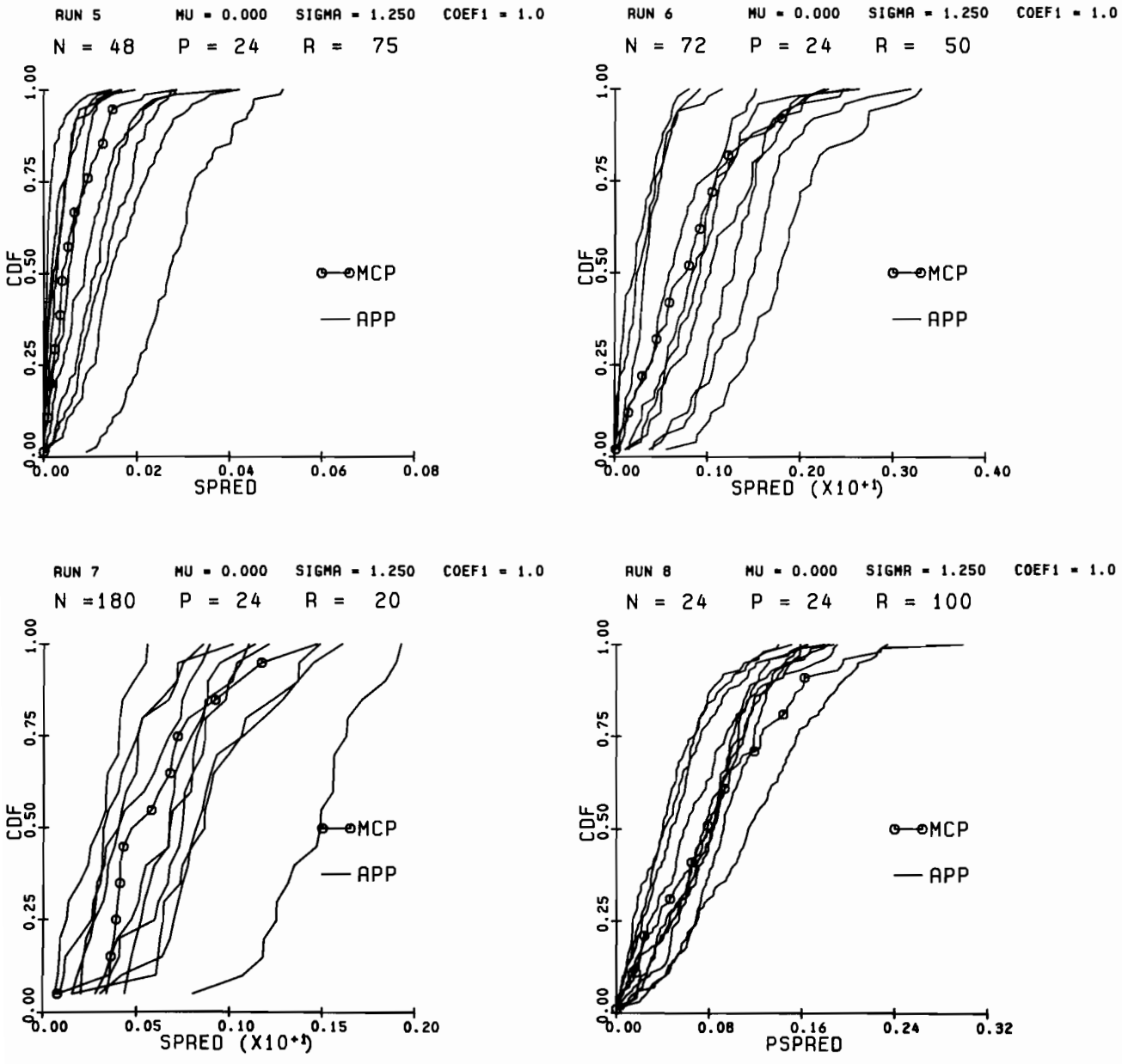

RUN 5 MU = 0.000 SIGMA = 1.250 COEF1 = 1.0
N = 48 P = 24 R = 75
CDF
SPRED
MCP
APP
RUN 6 MU = 0.000 SIGMA = 1.250 COEF1 = 1.0
N = 72 P = 24 R = 50
CDF
SPRED (X10+1)
MCP
APP
RUN 7 MU = 0.000 SIGMA = 1.250 COEF1 = 1.0
N =180 P = 24 R = 20
CDF
SPRED (X10+1)
MCP
APP
RUN 8 MU = 0.000 SIGMA = 1.250 COEF1 = 1.0
N = 24 P = 24 R = 100
CDF
PSPRED
MCP
APP

Fig. 9.2

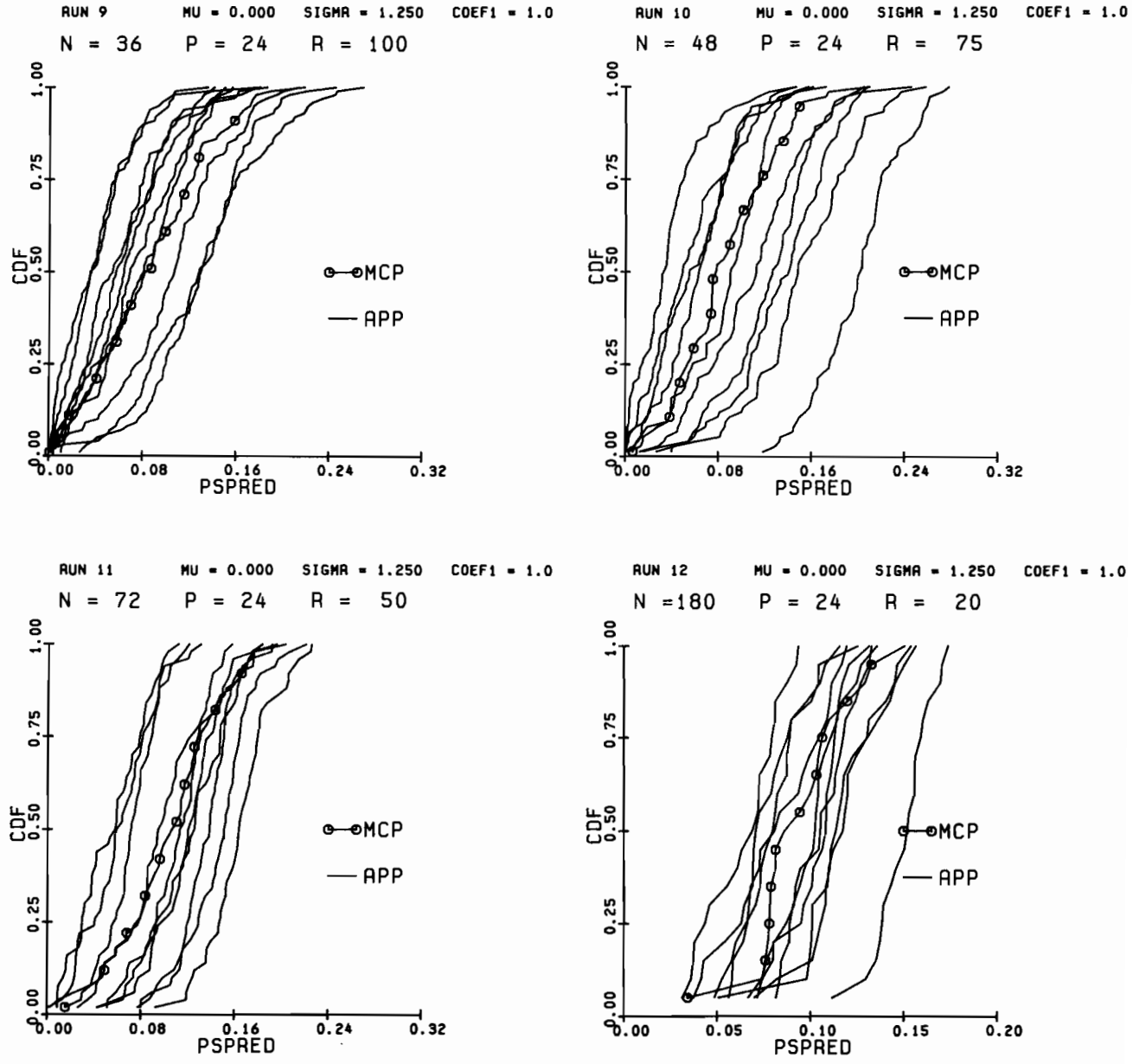

RUN 9 MU = 0.000 SIGMA = 1.250 COEF1 = 1.0
N = 36 P = 24 R = 100
RUN 10 MU = 0.000 SIGMA = 1.250 COEF1 = 1.0
N = 48 P = 24 R = 75
RUN 11 MU = 0.000 SIGMA = 1.250 COEF1 = 1.0
N = 72 P = 24 R = 50
RUN 12 MU = 0.000 SIGMA = 1.250 COEF1 = 1.0
N =180 P = 24 R = 20
CDF
PSPRED
MCP
APP
0.00
0.25
0.50
0.75
1.00
0.08
0.16
0.24
0.32
0.05
0.10
0.15
0.20

Fig. 9.3

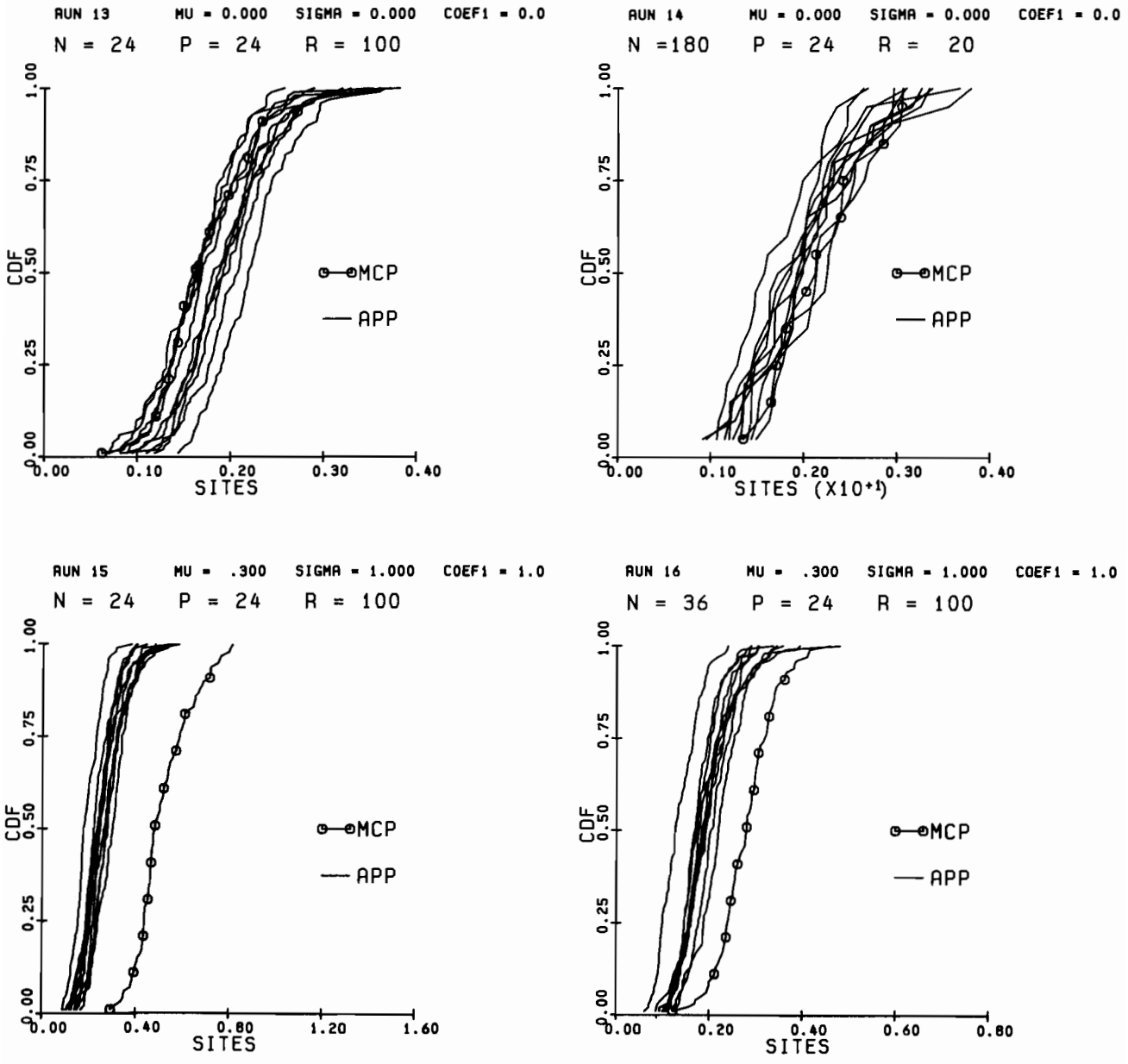

RUN 13 MU = 0.000 SIGMA = 0.000 COEF1 = 0.0
N = 24 P = 24 R = 100
CDF
SITES
MCP
APP
RUN 14 MU = 0.000 SIGMA = 0.000 COEF1 = 0.0
N =180 P = 24 R = 20
CDF
SITES (X10**1)
MCP
APP
RUN 15 MU = .300 SIGMA = 1.000 COEF1 = 1.0
N = 24 P = 24 R = 100
CDF
SITES
MCP
APP
RUN 16 MU = .300 SIGMA = 1.000 COEF1 = 1.0
N = 36 P = 24 R = 100
CDF
SITES
MCP
APP

Fig. 9.4

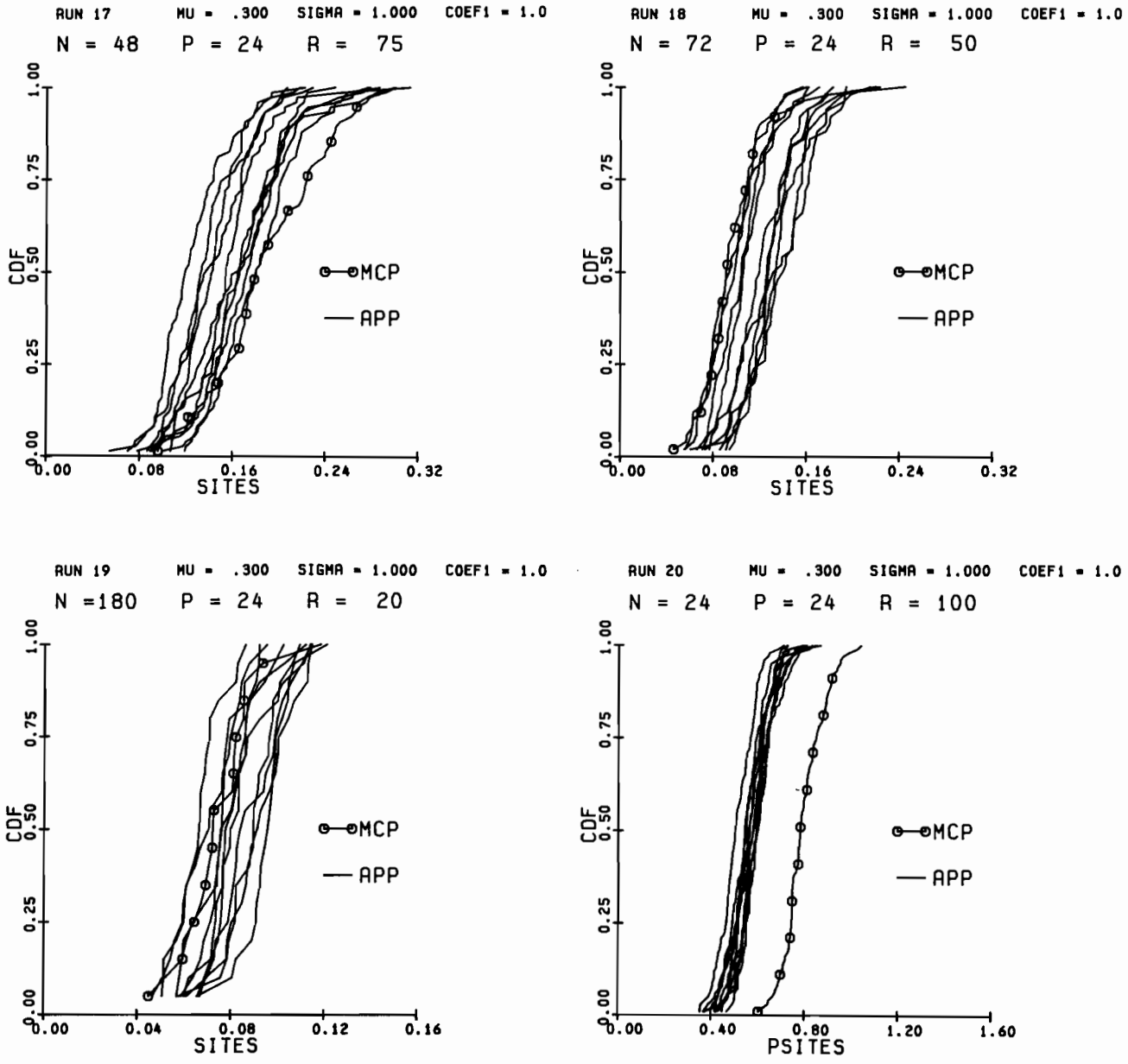

RUN 17 MU = .300 SIGMA = 1.000 COEF1 = 1.0
N = 48 P = 24 R = 75
RUN 18 MU = .300 SIGMA = 1.000 COEF1 = 1.0
N = 72 P = 24 R = 50
RUN 19 MU = .300 SIGMA = 1.000 COEF1 = 1.0
N =180 P = 24 R = 20
RUN 20 MU = .300 SIGMA = 1.000 COEF1 = 1.0
N = 24 P = 24 R = 100
CDF
SITES
PSITES
MCP
APP
0.00
0.25
0.50
0.75
1.00
0.08
0.16
0.24
0.32
0.04
0.12
0.40
0.80
1.20
1.60

Fig. 9.5

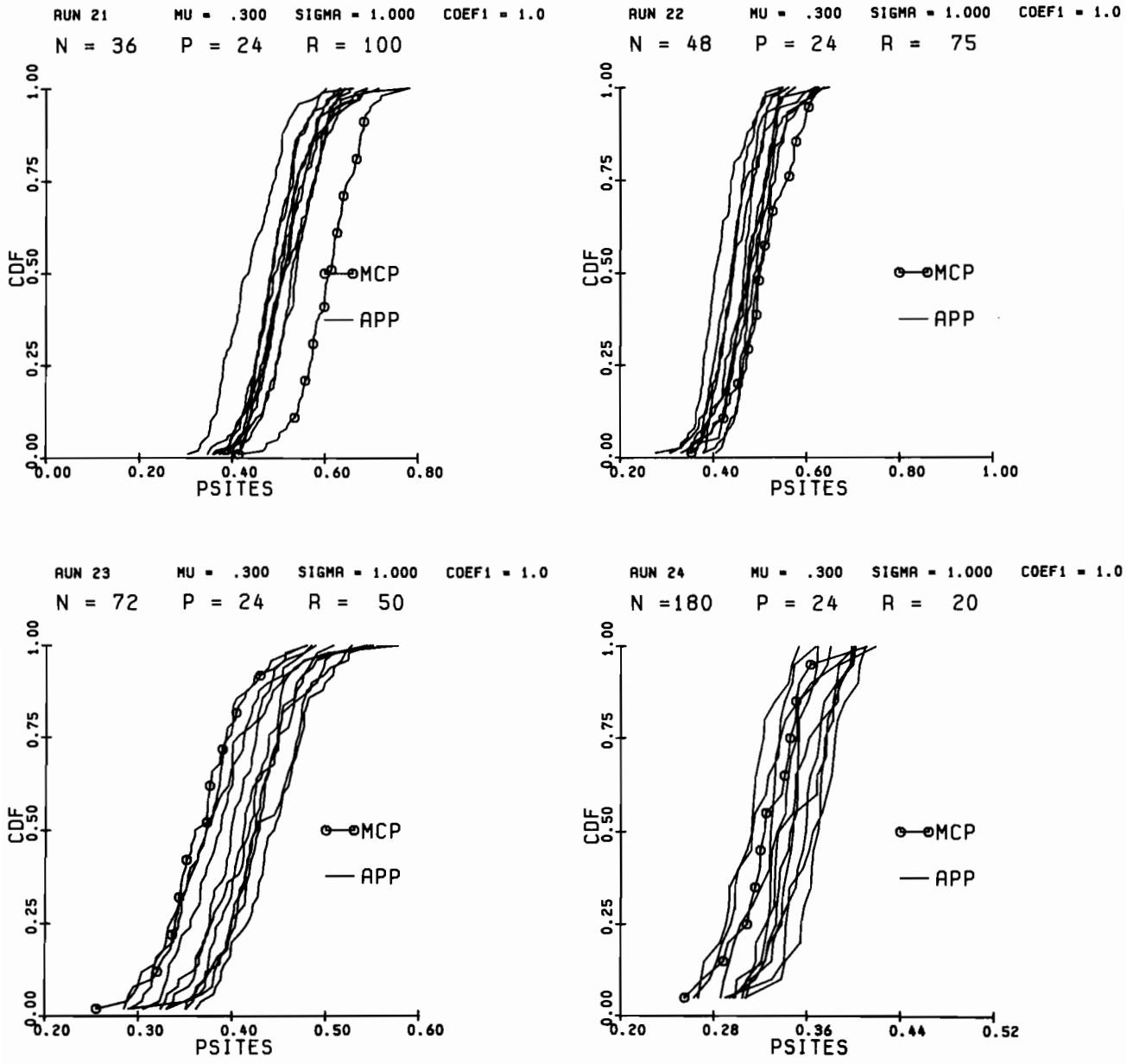

RUN 21 MU = .300 SIGMA = 1.000 COEF1 = 1.0
N = 36 P = 24 R = 100
RUN 22 MU = .300 SIGMA = 1.000 COEF1 = 1.0
N = 48 P = 24 R = 75
RUN 23 MU = .300 SIGMA = 1.000 COEF1 = 1.0
N = 72 P = 24 R = 50
RUN 24 MU = .300 SIGMA = 1.000 COEF1 = 1.0
N =180 P = 24 R = 20
CDF
PSITES
MCP
APP

Fig. 9.6

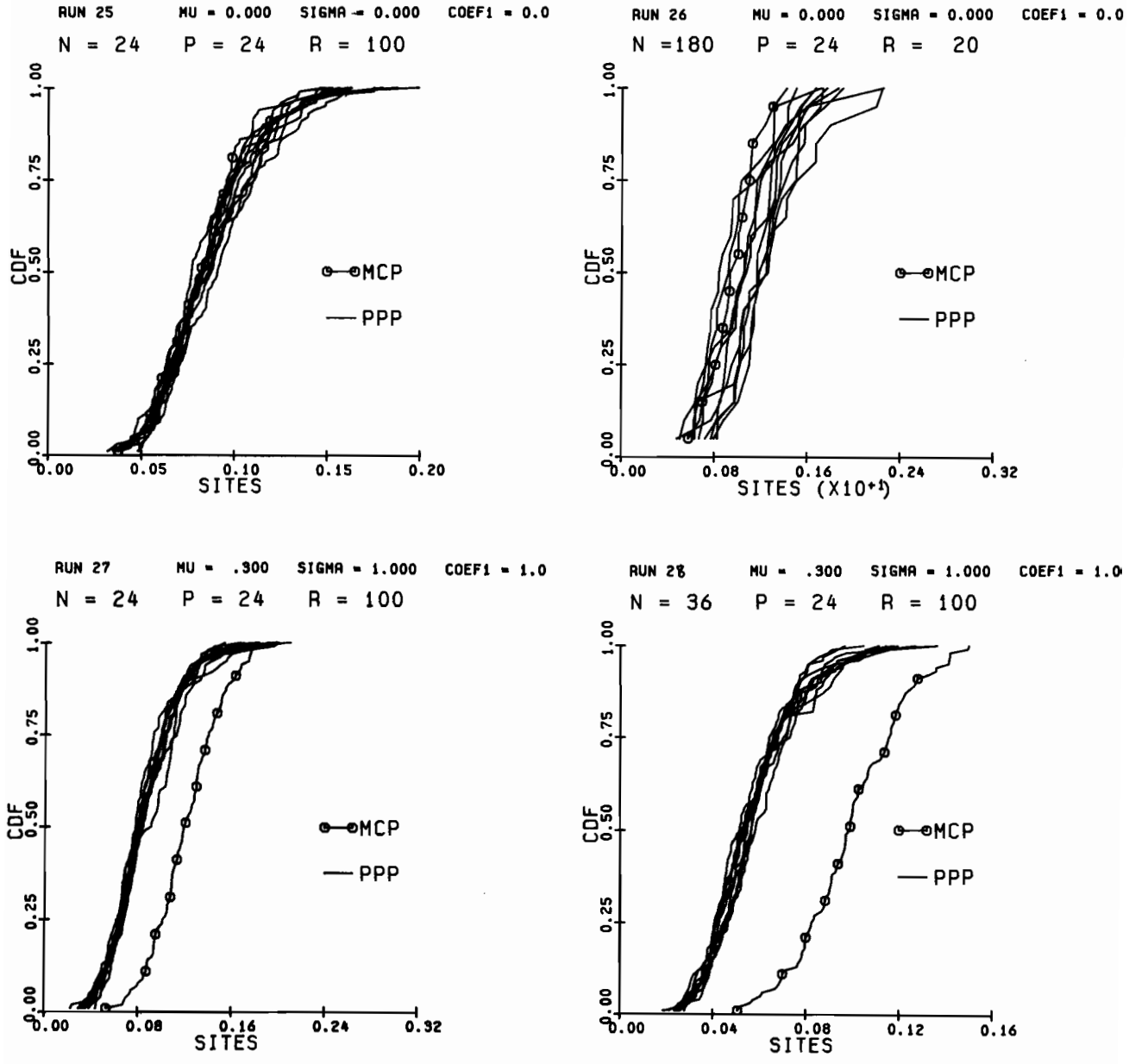

RUN 25 MU = 0.000 SIGMA = 0.000 COEF1 = 0.0
N = 24 P = 24 R = 100
CDF
SITES
MCP
PPP
RUN 26 MU = 0.000 SIGMA = 0.000 COEF1 = 0.0
N =180 P = 24 R = 20
SITES (X10**-1)
RUN 27 MU = .300 SIGMA = 1.000 COEF1 = 1.0
N = 24 P = 24 R = 100
RUN 28 MU = .300 SIGMA = 1.000 COEF1 = 1.0
N = 36 P = 24 R = 100

Fig. 9.7

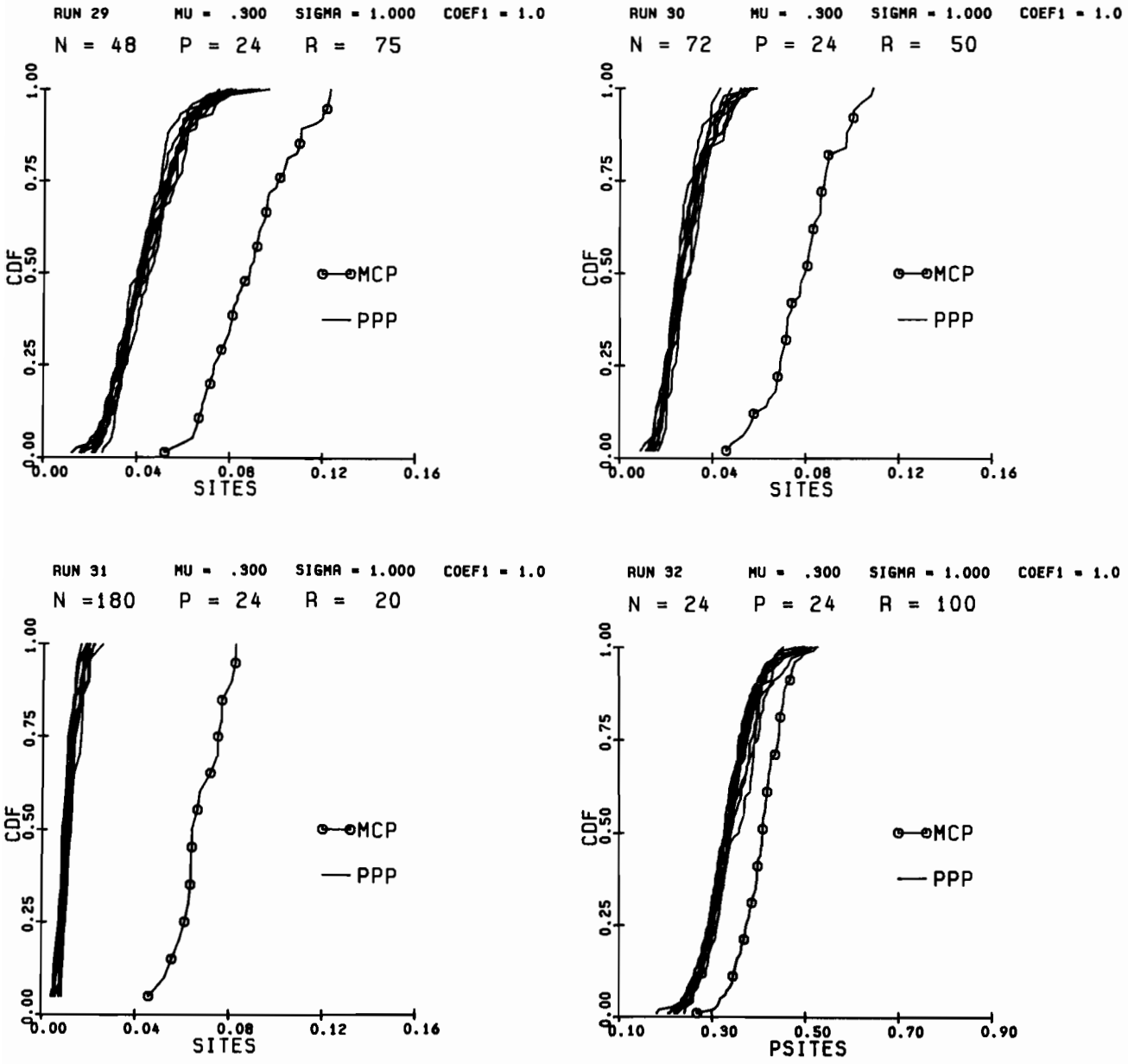

RUN 29 MU = .300 SIGMA = 1.000 COEF1 = 1.0
N = 48 P = 24 R = 75
RUN 30 MU = .300 SIGMA = 1.000 COEF1 = 1.0
N = 72 P = 24 R = 50
RUN 31 MU = .300 SIGMA = 1.000 COEF1 = 1.0
N =180 P = 24 R = 20
RUN 32 MU = .300 SIGMA = 1.000 COEF1 = 1.0
N = 24 P = 24 R = 100
CDF
SITES
PSITES
MCP
PPP
0.00
0.25
0.50
0.75
1.00
0.04
0.08
0.12
0.16
0.10
0.30
0.50
0.70
0.90

Fig. 9.8

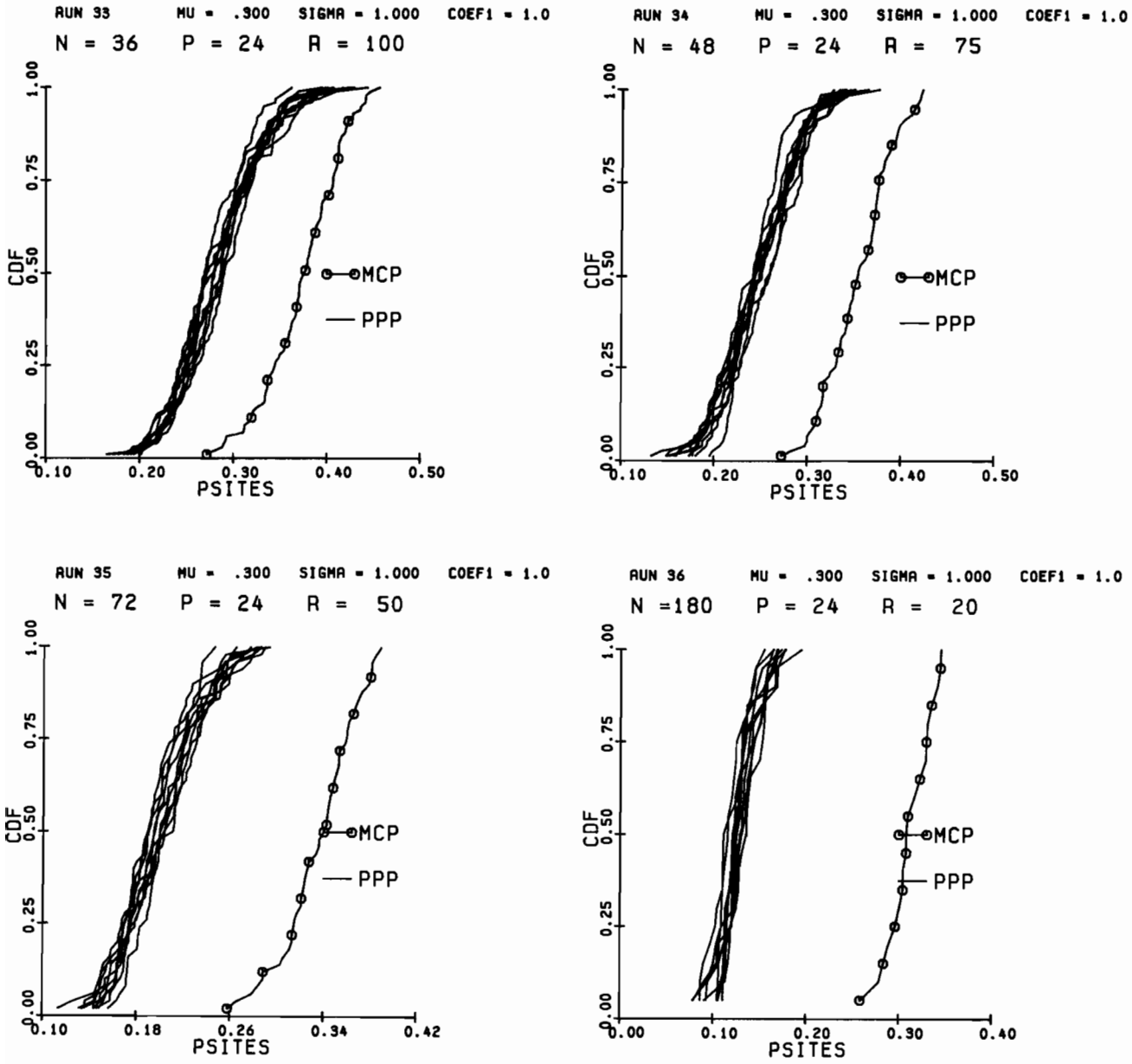

RUN 33 MU = .300 SIGMA = 1.000 COEF1 = 1.0
N = 36 P = 24 R = 100
RUN 34 MU = .300 SIGMA = 1.000 COEF1 = 1.0
N = 48 P = 24 R = 75
RUN 35 MU = .300 SIGMA = 1.000 COEF1 = 1.0
N = 72 P = 24 R = 50
RUN 36 MU = .300 SIGMA = 1.000 COEF1 = 1.0
N =180 P = 24 R = 20
CDF
PSITES
MCP
PPP

Fig. 9.9

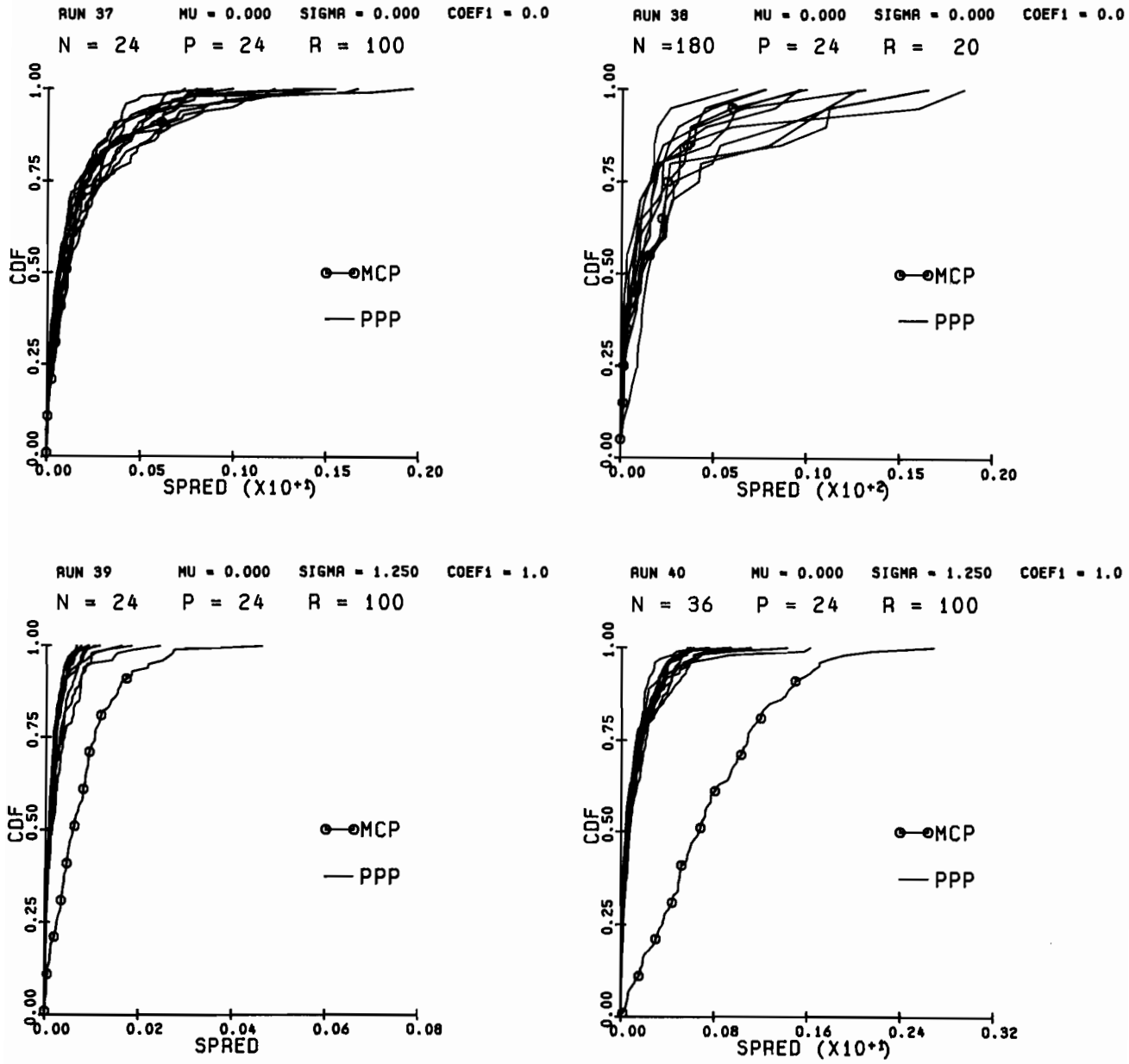

RUN 37 MU = 0.000 SIGMA = 0.000 COEF1 = 0.0
N = 24 P = 24 R = 100
CDF
SPRED (X10+1)
MCP
PPP
RUN 38 MU = 0.000 SIGMA = 0.000 COEF1 = 0.0
N =180 P = 24 R = 20
CDF
SPRED (X10+2)
MCP
PPP
RUN 39 MU = 0.000 SIGMA = 1.250 COEF1 = 1.0
N = 24 P = 24 R = 100
CDF
SPRED
MCP
PPP
RUN 40 MU = 0.000 SIGMA = 1.250 COEF1 = 1.0
N = 36 P = 24 R = 100
CDF
SPRED (X10+1)
MCP
PPP

Fig. 9.10

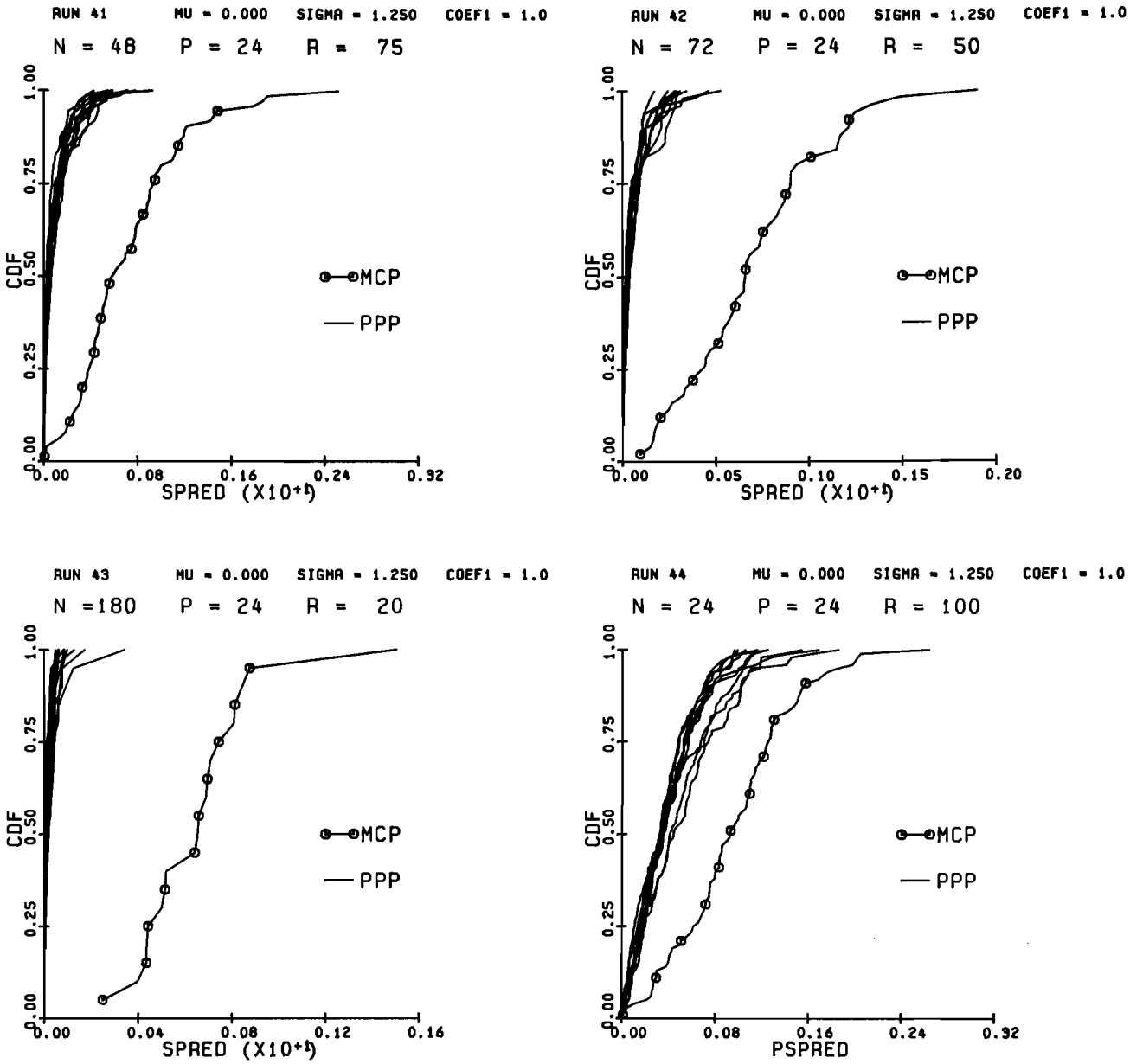

RUN 41 MU = 0.000 SIGMA = 1.250 COEF1 = 1.0
N = 48 P = 24 R = 75
CDF
SPRED (X10+4)
MCP
PPP
RUN 42 MU = 0.000 SIGMA = 1.250 COEF1 = 1.0
N = 72 P = 24 R = 50
CDF
SPRED (X10+4)
MCP
PPP
RUN 43 MU = 0.000 SIGMA = 1.250 COEF1 = 1.0
N =180 P = 24 R = 20
CDF
SPRED (X10+4)
MCP
PPP
RUN 44 MU = 0.000 SIGMA = 1.250 COEF1 = 1.0
N = 24 P = 24 R = 100
CDF
PSPRED
MCP
PPP

Fig. 9.11

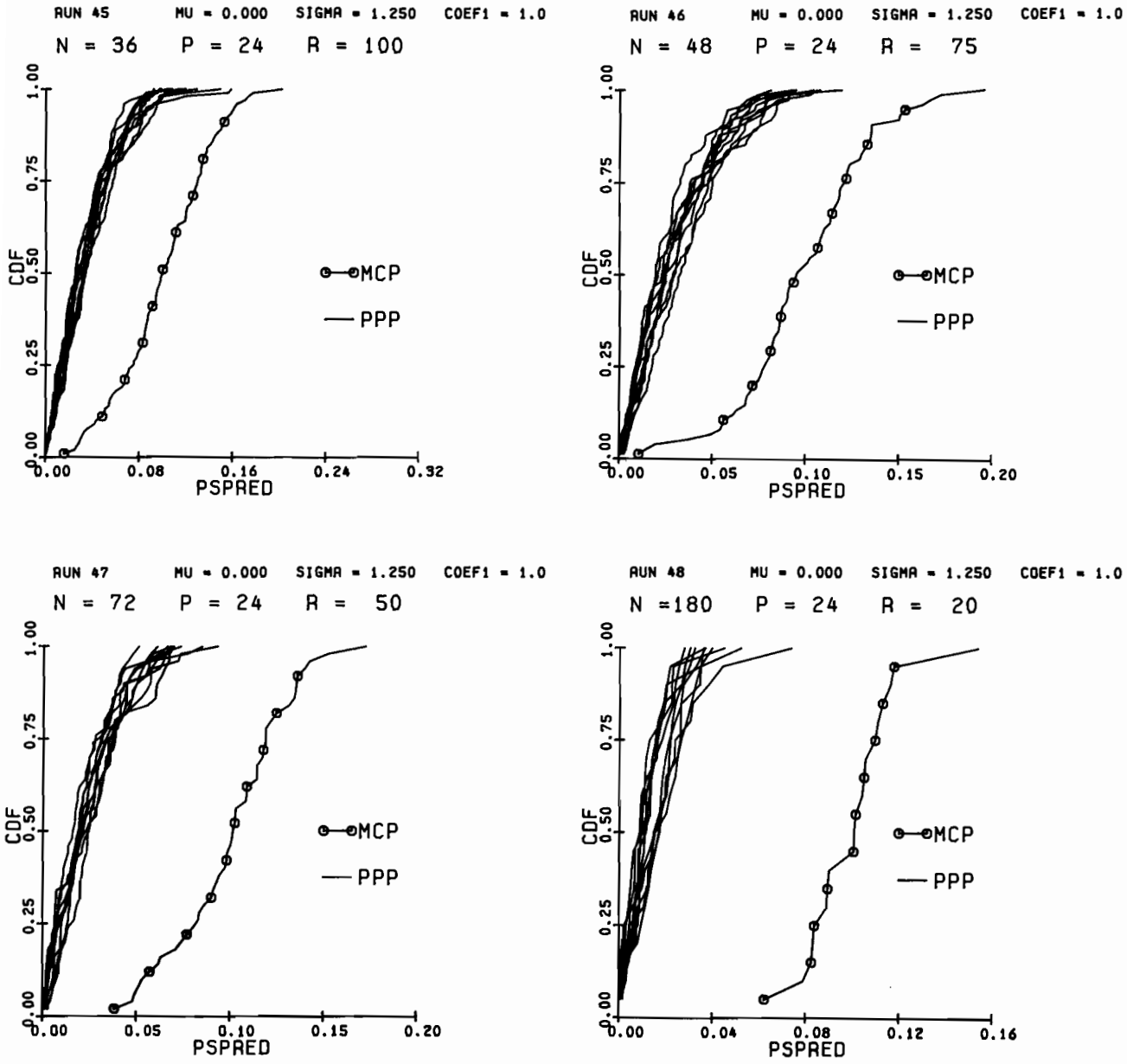

RUN 45 MU = 0.000 SIGMA = 1.250 COEF1 = 1.0
N = 36 P = 24 R = 100
RUN 46 MU = 0.000 SIGMA = 1.250 COEF1 = 1.0
N = 48 P = 24 R = 75
RUN 47 MU = 0.000 SIGMA = 1.250 COEF1 = 1.0
N = 72 P = 24 R = 50
RUN 48 MU = 0.000 SIGMA = 1.250 COEF1 = 1.0
N =180 P = 24 R = 20
CDF
PSPRED
MCP
PPP

Fig. 9.12

10. References

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

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.

Anderson, T.W. (1958) An Introduction to Multivariate Statistical Analysis, Wiley, N.Y.

Box, G.E.P. (1949) "A General Distribution Theory for a class of likelihood criteria." Biometrika 36, 317.

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

Crow, E.L., F.A. Davis, M.W. Maxfield (1960) Statistics Manual, Dover, N.Y.

Korin, B.P. (1969) "On testing the equality of k covariance matrices." Biometrika 56, 216.

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

Liu, C-T. (1982) Analysis of Tropical Pacific Sea Surface Temperatures for 1975 to 1980. Tech. Memo ERL PMEL-34, Pacific Marine Environmental Laboratory, Seattle, WA.

Pearson, E.G. (1969) "Some comments on the accuracy of Box's approximations to the distribution of M." Biometrika 56, 219.

APPENDIX A

A Model/Data Matrix Generator for Controlled Experiments with Statistical Techniques

1. The data-set generator is defined via

$$d(t,x) = f(t,x) + r_1(t,x) \qquad \text{(A1.1)}$$

for $t = 1,\ldots,n$, and $x = 1,\ldots,p$, where

$$f(t,x) = a \sum_{j=1}^{p} u_j(x)v_j(t) \qquad \text{(A1.2)}$$

$$u_j(x) = [\frac{2}{(p+1)}]^{\frac{1}{2}} \sin(\frac{j\pi x}{p+1})$$

$$v_j(t) = \cos\left(\frac{2\pi r_j t}{M}\right)$$

$$r_j = \sin\,(j\pi/2(p+1))/\sin(p\pi/2(p+1))$$

Here M is an adjustable parameter fixed throughout these constructions as $M = 12$. The constant a adjusts the signal-to-noise ratio. The present system has many combined periods (via the various values of r_j). The shortest natural period is

$$T \equiv \pi/\sin(\frac{p\pi}{2(p+1)}) \qquad \text{(A1.3)}$$

Choosing $M = 12$ then divides the time axis into 12ths of T. The time index t then gives the multiples of T by which time progresses in this system. Hence the quickest natural oscillatory changes of the system are still resolved into 12 instants. Finally, $r_1(t,x)$ is a random sample, at t and x, drawn from a

normal population of zero mean and unit variance. That is, for each t and x, $r_1(t,x) \sim N(0,1)$.

B. The model-set generator is defined via

$$m(t,x) = \mu + \sigma f(t,x) + r_2(t,x) \qquad \text{(A1.4)}$$

$$t = 1,\ldots,n \quad , \quad x = 1,\ldots,p.$$

Here μ, σ are specifiable parameters. By varying μ, we can make the centroids of the sampled $\underline{D}$ and $\underline{M}$ swarms in E_p change distance. By varying σ, the radius of the sampled $\underline{M}$ swarm can be changed. Here $r_2(t,x) \sim N(0,1)$, and is a random variable sampled independently of $r_2(t,x)$.

C. We can write (A1.1), (A1.4) in n×p matrix form:

$$\underline{D} \equiv \underline{F} + \underline{R}_1 \qquad \text{(A1.5)}$$

$$\underline{M}(\mu,\sigma) \equiv (\mu\underline{E} + \sigma\underline{F}) + \underline{R}_2 \qquad \text{(A1.6)}$$

where $\underline{E}$ is an n×p matrix of unit entries. By setting $\mu = 0$, $\sigma = 0$, and $a = 0$, we obtain

$$\underline{D} = \underline{R}_1 \qquad \text{(A1.7)}$$

$$\underline{M} = \underline{R}_2 \quad , \qquad \text{(A1.8)}$$

two random n×p matrices representing independent samples of n points each in E_p drawn from $N_p(\underline{O},\underline{I}_p)$, i.e., a multivariate unit spherical gaussian population in E_p, centered on the origin.

D. In general, we may write $\underline{D}$ as $\{\underline{d}(1),\ldots,\underline{d}(n)\}$, i.e., as an n-point swarm (or set) in E_p. Each $\underline{d}(t)$, $t = 1,\ldots,n$, is a vector:

$$\underline{d}(t) = [d(t,1),\ d(t,2),\ldots,\ d(t,p)]^T \qquad \text{(A1.9)}$$

the elements being formed via (A1.1). Similar notation holds for the vectors $\underline{m}(t)$ associated with $\underline{M}$.

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, Glenn 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 particulate 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. Overland (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. Hayes (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), Alaskan 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)
Hydrographic 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.