Naval Oceanographic and Atmospheric Research Laboratory
Technical Note 284
July 1992

A Horizontal Refractivity Depiction Product: An Evaluation

G. N. Vogel
Forecast Guidance and Naval Systems Support Division
Atmospheric Directorate
Monterey, CA 93943-5006

ABSTRACT

The results from an independent validation of the Naval Western Oceanography Center (NWOC) Horizontal Refractivity Depiction (HRD) product, based on a 12 month North Pacific data set, are presented. The average capability of the HRD product (analysis and 36 hr prognosis) in correctly assessing observed refractive structure (either ducting or nonducting) was near 65%; HRD forecasts of ducting were reliable 79% of the time. Based on the Hanssen and Kuipers skill score (the difference between the hit and false alarm rates), the capabilities of the HRD product in the assessment and short-range forecasting of ducting were determined to be statistically better than those using an electromagnetic propagation conditions climatology. Any potential operational use of the NWOC HRD product in regions climatically distinct from the eastern North Pacific would likely require some modification of existing forecast thumbrules. The implementation of an automated synoptic-satellite inference rule base (as per an expert system) for the HRD, which would likely result in a more objective and consistent refractivity product, is recommended.

ACKNOWLEDGMENTS

The support of the sponsor, Commander Naval Oceanography Command, CDR J. Jensen, USN, Program Element O&M,N-2, is gratefully acknowledged.

TABLE OF CONTENTS

A HORIZONTAL REFRACTIVITY DEPICTION PRODUCT: AN EVALUATION

1. INTRODUCTION

In response to meteorological requirements by Navy commanders for refractive forecasting products in support of Fleet operations as well as command, control and communications, various Navy command centers and detachments have developed and tested regional refractivity products within the past decade. In 1990, the Naval Western Oceanography Center (NWOC) in conjunction with the Pacific Missile Test Center (PMTC) at Pt. Mugu, California commenced development of a new Horizontal Refractivity Depiction (HRD) product. During the period April - September 1991, an "in-house" operational evaluation of this product was performed (NWOC, 1991). Although results for this trial period were favorable, an independent validation and verification of the HRD is required before the product can be approved for full operational status and Fleet-wide dissemination. This NOARL Technical Note presents the results of such an evaluation.

The verification of the NWOC HRD product is by direct comparison with shipboard upper air observations. The data used for this validation include that previously analyzed by NWOC as well as additional data from the October 1991 - March 1992 time period. Various statistical indices are computed in order to assess the forecast capabilities of the HRD product. In particular, a determination is made as to how well the HRD product is able to assess and forecast the most important anomalous propagation phenomenon to impact naval operations - ducting. In order to assess the degree of skill and usefulness of the HRD product, direct comparisons are made between the HRD and a readily available electromagnetic (EM) propagation conditions climatology. Apart from the detailed statistical analysis, this report briefly addresses several other issues, including the scientific accuracy of the HRD forecasting thumbrules and the product's suitability for other ocean basins (i.e., besides the North Pacific).

2. HRD DESCRIPTION

The following description of the NWOC Horizontal Refractivity Depiction chart, and the procedures used in its production, is based on detailed information provided by NWOC (1991). In general, NWOC operational procedures for refractivity assessment and forecasting are based on techniques developed at PMTC (Helvey and Rosenthal, 1983; Rosenthal et al., 1985). The manually produced HRD includes both an analysis and a 36 hr prognosis which display areas of normal, superrefractive and trapping conditions for the North Pacific (10°N to 50°N, 150°E to the western coast of North America), from the surface to 10,000 feet. It is intended as a tactical aid in the planning and conducting of Fleet operations which are sensitive to EM wave propagation conditions in the lower atmosphere. During the period covered by this study, NWOC produced the HRD chart twice daily (at 00Z and 12Z); for evaluation purposes, the 36 hr prognosis was considered to be valid from 25 to 36 hours after the initial (i.e., analysis) time. A sample HRD chart is shown as Figure 1.

Given the immense area of responsibility, the primary tool for the development of the HRD chart is satellite imagery. Visible and, to a lesser extent, infrared (IR) satellite imagery permit large-scale cloud patterns to be defined in terms of location, type and appearance (smooth, granular or cellular). In conjunction with thumbrules, which associate certain cloud patterns with specific refractive structures, visible and IR satellite data provide the analyst good open ocean estimates of duct probability and strength. If available, the HRD analyst can use the IR-duct technique (Lyons, 1985) to assess the duct topography over regional areas overlaid with stratiform clouds. The likelihood of ducting may also be inferred using water vapor satellite imagery, with very dark regions indicating possible ducts.

If satellite data are not available, synoptic information and observations are used as the primary input for the HRD. From synoptic surface and upper air analyses, the HRD analyst first locates frontal zones, cyclones and anticyclones, and areas of warm-air advection, subsidence and temperature inversions. Individual synoptic features are then associated with specific refractive structures as defined by operational thumbrules. Over the open ocean, estimates of duct height in the vicinity of anticyclones are made based on sea surface temperature. For coastal areas experiencing warm, dry offshore flow, specific guidelines apply for surface-based ducting. Where available, processed upper air ship observations provide

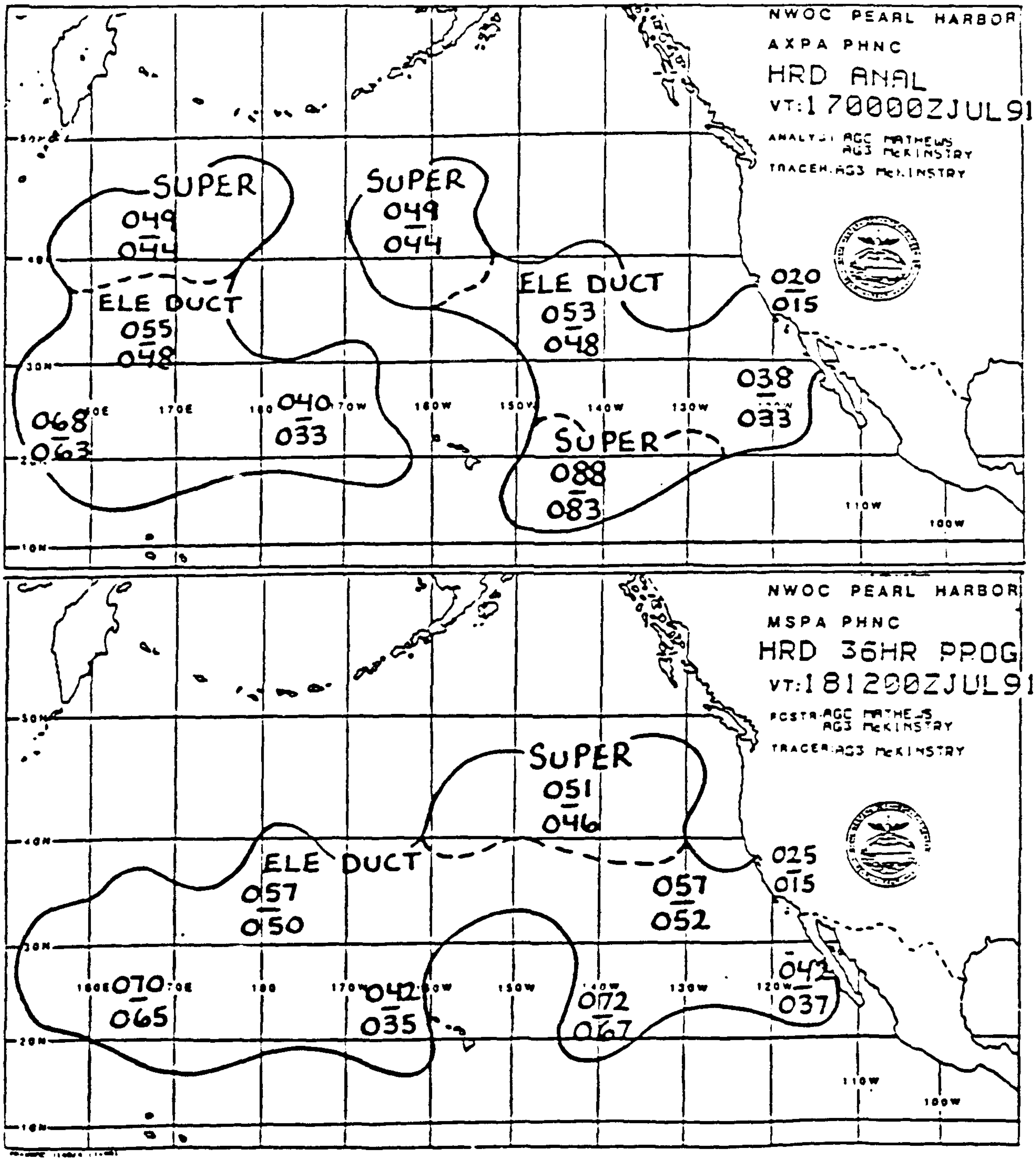


Figure 1. Sample HRD product (from NWOC, 1991). Refractive areas are identified as surface duct, elevated duct, superrefractive or normal; at selected locations, the top and bottom heights of nonstandard layers are given in hundreds of feet.

definitive refractivity information for the HRD analyst. Surface weather observations, in conjunction with available thumbrules, may be used to infer local refractive structure; for example, the presence of a probable inversion and duct can be inferred from a report of a haze layer.

Also available to NWOC meteorological office personnel during production of the HRD chart are numerical products issued by the Fleet Numerical Oceanography Center (FNOC). The "REFRACP" product provides refractive conditions (and M-gradients) at specified locations for six pressure levels (surface to 500 mb); these NOGAPS (Navy Operational Global Atmospheric Prediction System) extracted data are available to NWOC as short range forecasts in either tabular or graphical form. For a few select locations, high resolution vertical refractivity profiles (40 levels below ~ 600 mb) from the NABL (Navy Atmospheric Boundary Layer) forecast model are also available. Procedural guidelines for the HRD product indicate that the REFRACP and NABL products may be used for rough estimates and end product evaluation, but not as primary data sources. During the "in-house" evaluation of the HRD, the NABL model was found to be conservative in most cases, forecasting (at most) superrefractive when elevated ducting was observed (NWOC, 1991). As a last resort, in the event that no satellite or synoptic data are available, and the previous HRD forecast is no longer valid, NWOC guidelines call for the use of an EM propagation conditions climatology to estimate large-scale refractive structures, in particular, duct heights.

After the completion of the HRD analysis, the NWOC personnel use the synoptic situation depicted in the 36 hr prognosis blend to develop a refractive forecast. Synoptic thumbrules are used to forecast refractive condition changes. NWOC HRD production guidelines stress that continuity be maintained between the analysis and the 36 hr forecast; this is done by careful blending of expected movements and intensity changes of analyzed refractivity features.

3. VERIFICATION

A verification of a set of forecasts needs to determine the accuracy of the forecasts, the skill in forecasting and the operational value to the user. While accuracy and forecast skill can be evaluated statistically, the operational value of forecasts is much more difficult (if not impossible) to determine since it requires a knowledge of user strategy. For that reason, it is most often desirable to express verification results in some arbitrary quantitative manner so that the user of the forecast can then interpret the information in terms of his specific operations.

3.1 Data

The validation of the NWOC Horizontal Refractivity Depiction product is based on comparisons of the HRD analyses and 36 hour prognoses to shipboard radiosonde reports and to an electromagnetic propagation conditions climatological database available within the Tactical Environmental Support System (TESS). In effect, the validation with radiosondes is both a continuation and expansion of the original 6 month operational evaluation performed by NWOC (1991). On the other hand, the comparison of the HRD product against climatology is a completely new evaluation effort.

3.1.1 Radiosonde

During the period April 1991 - March 1992, the NWOC routinely received hundreds of upper air observations from ships within its operational area. When received, these radiosonde reports were processed into surface to 10,000 ft refractivity profiles suitable for comparison with the HRD product. Depending on the actual structure of the lower atmosphere, such refractivity profiles could indicate ducting, as well as superrefractive, standard and subrefractive propagation conditions.

Individual radiosonde reports may be classified according to their most significant propagation condition. For a sounding with both a duct and a superrefractive layer, the duct is considered more critical. A superrefractive layer is considered the most critical aspect of an otherwise standard refractive structure. Any sounding devoid of either a duct or superrefractive layer is classified as standard.

Figure 2 gives the monthly number of radiosonde reports (from April 1991 to March 1992) for various refractive categories. Significant fluctuations are noted in the monthly number of soundings with ducts and standard conditions. With the exception of one month, the number of radiosondes with ducts is greater than those classified as standard. Of the total 373 radiosondes used in this technical note, almost exactly two-thirds had ducts. In general, the number of soundings with multiple ducts and surface ducts, and the number classified as superrefractive, are all relatively small. The monthly number of surface ducts is noticeably higher during the winter period (October-March). On the contrary, soundings with multiple ducts and superrefractive conditions are seen to occur almost exclusively during the summer period (April-September). This peculiar distribution of multiple duct and superrefractive occurrences is believed not due to any real seasonal variability in frequency of occurrence, but rather to differences in NWOC data processing and reporting procedures between the first and second 6 month data periods.

In addition to refractive categories, radiosonde reports are grouped according to their time and location of occurrence. Day and night classifications were based on graphical time zone and sunrise/sunset information provided by Rudloff (1981). Observations located eastward of a line (from Alaska southward along the 150°W meridian to 40°N, then southeastward to 20°N, 130°W, then southward along the 130°W meridian) were classified as EPAC (east Pacific); those westward of this boundary, as CPAC (central Pacific) (see Figure 5). Latitudinal classifications (i.e., north and south) were not formed since almost three out of every four radiosondes were located within the subtropical band 20°-35°N. The monthly distribution of radiosonde reports based on temporal and geographical classifications is shown in Figure 3 Although these monthly data exhibit considerable variability, one can observe that the number of CPAC observations is markedly less than the number of EPAC observations over the second 6 month data period (October-March). When the entire 12 month study period is considered, the number of day and night observations is about the same.

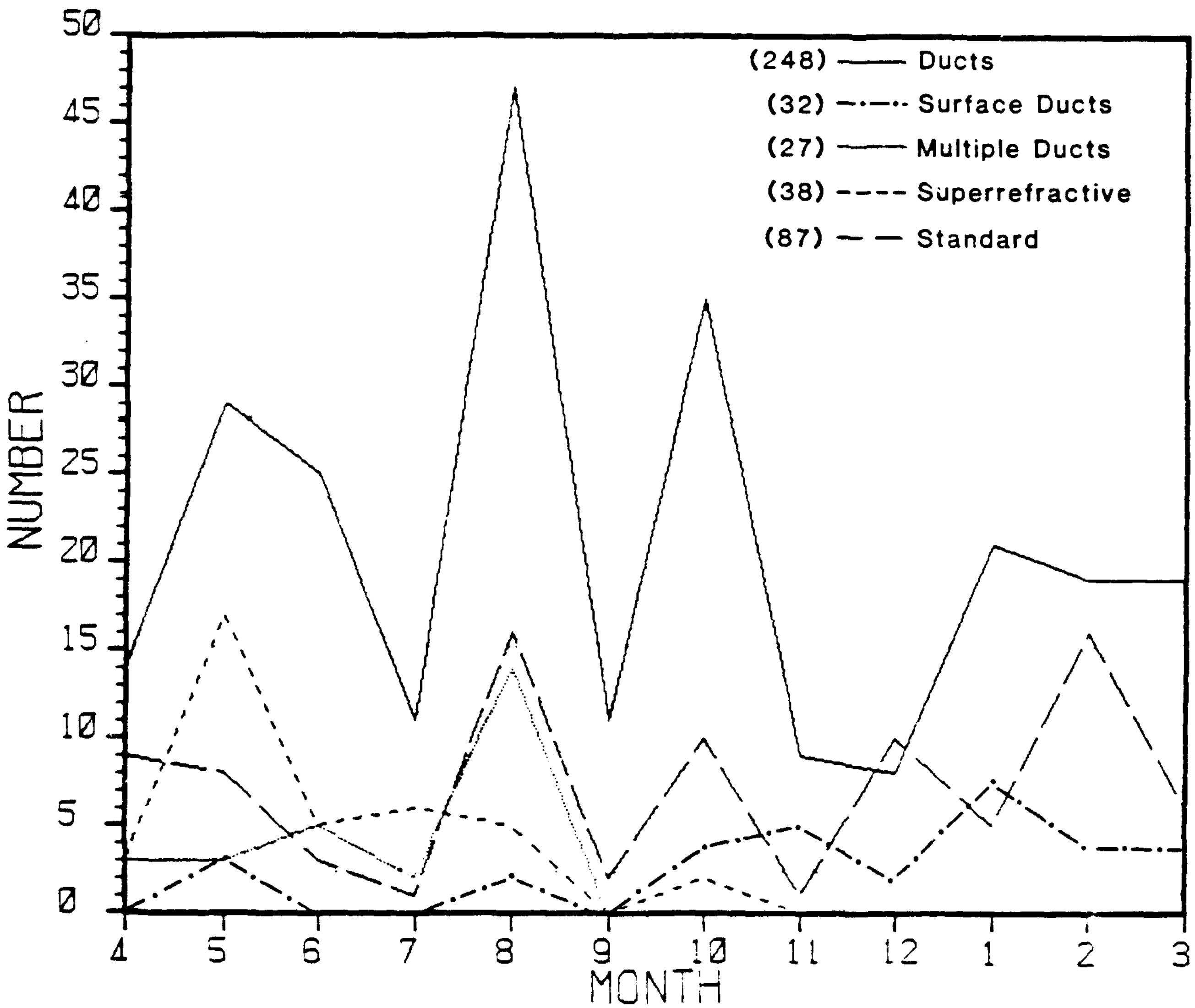


Figure 2. Monthly number of radiosonde reports based on refractive classifications. The total for each class over the April 1991 - March 1992 period is given by the number in parenthesis.

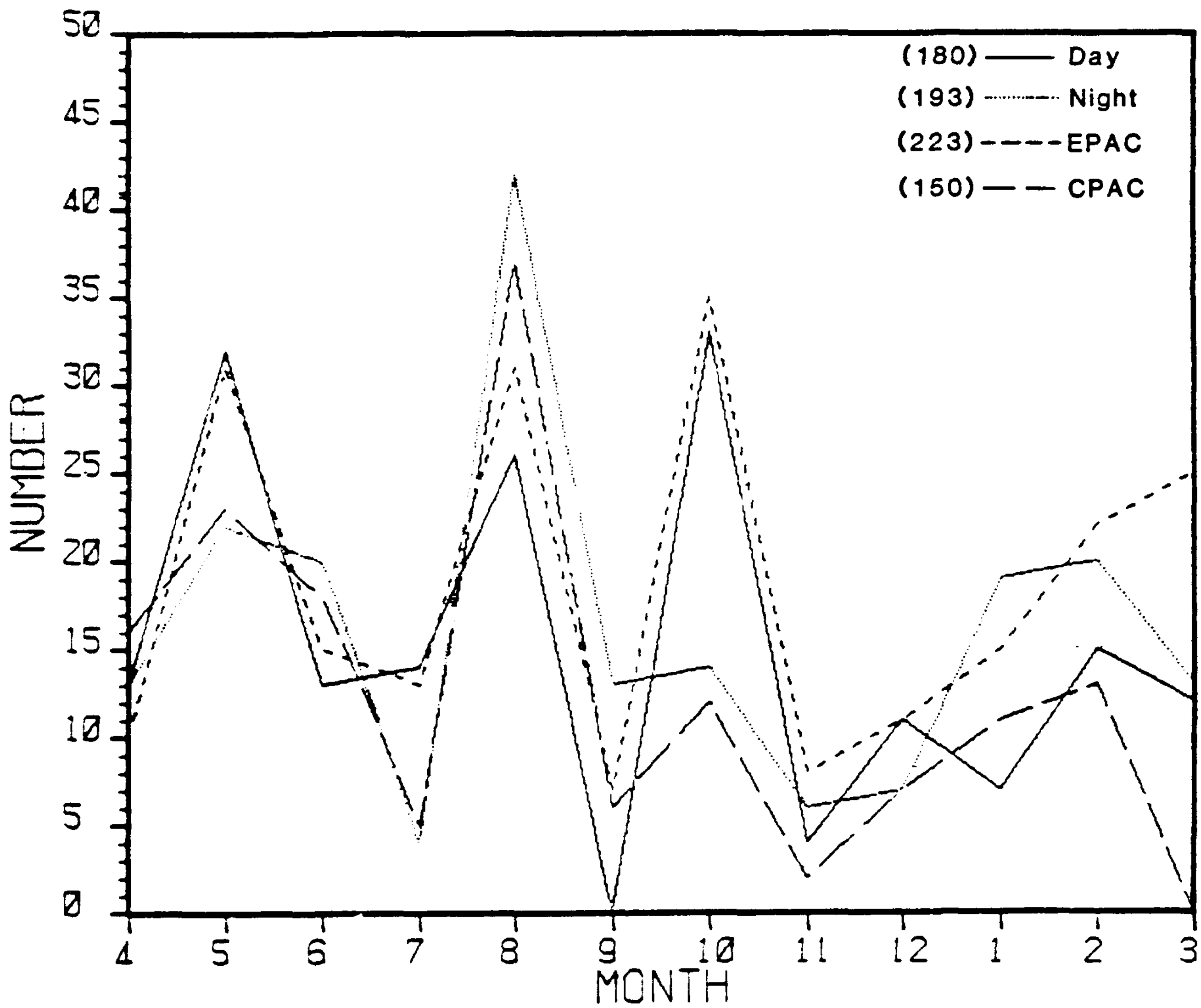


Figure 3. Monthly number of radiosonde reports based on temporal and geographical classifications. The total for each class over the April 1991 - March 1992 period is given by the number in parenthesis.

For this study, tabular radiosonde refractivity data , along with the corresponding HRD analysis and prognosis data, were provided to NRL Monterey by NWOC meteorological office personnel. Since the initial 6 month tabular data set was quite incomplete, the April through September radiosonde and HRD data were instead taken from bar graphs in NWOC (1991). However, even after this data extraction, the total number of radiosonde reports (212) still did not match the number reported by NWOC in their own HRD evaluation (220). Due to repetitions and obvious errors (e.g., ship reports over land), 6 observations with the second 6 month data set (October-March) were not used. Such data problems suggest that the processed radiosonde data used in this study are not of the highest possible reliability. On the other hand, given the relatively large sample size, it is likely that the data are of good enough quality to justify their use in verification.

3.1.2 HEPC Climatology

The climatology used in comparison against the NWOC HRD product is the Historical Electromagnetic Propagation Conditions (HEPC) Summary Function within the TESS. It is based on 5 noncontinuous years (between 1966 and 1974) of coastal, island and fixed-location station ship radiosonde reports. Pertinent to this study, the climatology provides monthly percent occurrence of both surface-based and elevated ducts, given in day, night and day/night (average) percentages (see Figure 4). Additionally, the average thickness of the surface-based duct and the average top and thickness of the elevated duct are given. Note that the HEPC climatology deals almost exclusively with enhanced propagation conditions; if so desired, percent occurrence of standard atmospheric conditions could be deduced from the available information. The HEPC climatology consists of data for the preceding month, the selected month, the following month and the average of the 3 months. Further information on the HEPC database is provided by Patterson (1987).

HISTORICAL PROPAGATION CONDITIONS SUMMARY

Specified location:	20.00 N	130.00 W	MS = 86	(* = Insufficient data)
Radiosonde source:	30.00 N	140.00 W	MS = 123	WMO number = 4YN
Coastal station:	FIXED SHIP, NORTH PACIFIC OCEAN			Hgt = 12 m
Surface observation source:	25.00 N	125.00 W	MS = 85	

Percent occurrence of enhanced surface-to-surface radar/esm/com ranges:

	Average			Jan			Feb			Mar		
Frequency	day	nite	d&n	day	nite	d&n	day	nite	d&n	day	nite	d&n
100 Mhz	3	<1	2	3	<1	2	4	<1	2	3	<1	2
1 GHz	12	3	7	11	3	7	14	2	8	10	4	7
3 GHz	17	4	11	16	4	10	21	3	12	15	6	11
6 GHz	47	26	37	45	26	36	49	24	37	47	29	38
10 GHz	71	55	64	70	55	62	72	52	63	73	58	66
20 GHz	90	83	87	88	82	85	90	82	86	91	85	88

Surface based duct summary:

	Average			Jan			Feb			Mar		
Parameter	day	nite	d&n	day	nite	d&n	day	nite	d&n	day	nite	d&n
Percent occurrence	22	5	13	19	5	12	28	3	16	18	7	13
Avg thickness Km			.09			.09			.09			.09
Avg trap freq GHz			.81			.73			.96			.75
Avg 1yr grad -N/Km			225			227			220			227

Elevated duct summary:

	Average			Jan			Feb			Mar		
Parameter	day	nite	d&n	day	nite	d&n	day	nite	d&n	day	nite	d&n
Percent occurrence	38	57	48	42	56	49	37	53	45	36	61	49
Avg duct top Km			1.6			1.4			1.6			1.7
Avg thickness Km			.17			.18			.17			.17
Avg trap freq GHz			.20			.21			.19			.20
Avg 1yr grad -N/Km			222			215			218			233
Avg 1yr base Km			1.5			1.3			1.5			1.6

Evaporation duct histogram in percent occurrence:

	Average			Jan			Feb			Mar		
Percent Occurrence	day	nite	d&n	day	nite	d&n	day	nite	d&n	day	nite	d&n
0 to 4 Meters	4	5	5	5	5	5	4	5	5	4	4	4
4 to 8 Meters	11	18	15	13	18	16	12	18	15	9	17	13
8 to 12 Meters	26	33	30	26	32	29	28	34	31	26	33	29
12 to 16 Meters	30	29	30	30	29	30	30	27	28	31	30	3[illegible]
16 to 20 Meters	19	12	16	19	12	15	17	11	14	20	14	17
20 to 24 Meters	7	3	5	6	3	5	7	3	5	7	2	5
24 to 28 Meters	1	<1	1	1	<1	<1	2	<1	1	2	<1	[illegible]
28 to 32 Meters	<1	<1	<1	<1	0	<1	<1	<1	<1	<1	<1	<1
32 to 36 Meters	<1	<1	<1	0	0	0	<1	0	<1	<1	<1	<1
36 to 40 Meters	0	0	0	0	0	0	0	0	0	0	0	0
above 40 Meters	0	<1	<1	0	<1	<1	0	0	0	0	0	0
Mean height Meters	13	11	12	13	11	12	13	11	12	13	12	13
# of observations	2k	2k	2k	2k	2k	2k	2k	2k	2k	2k	2k	2k

General meteorology summary:

	Average			Jan			Feb			Mar		
Parameter	day	nite	d&n	day	nite	d&n	day	nite	d&n	day	nite	d&n
# Accepted soundings	120	120	120	116	120	118	115	112	114	129	127	128
% occur EL&SB dcts			3			3			3			2
% occur 2+ EL dcts			3			5			3			2
Avg station N			339			343			337			337
Avg station -N/Km			48			50			48			45
Avg sfc wind m/s	7	7	7	7	7	7	7	7	7	7	7	7

Figure 4. Sample HEPC Function output (from Patterson, 1987).

Given a selected location, the HEPC Function selects the nearest radiosonde station in its database as the climatology for that site. Figure 5 depicts the location of all (28) HEPC climatological sites used in this study. A number at a site corresponds to the total number of ship radiosondes, over the April 1991 - March 1992 study period, that were assigned by the HEPC Function to that particular climatological site. Although ship observations covered a vast expanse of the North Pacific (from the Aleutian Islands to Central American waters and westward well past the dateline), the majority of the reports are concentrated along an axis extending from southern California to the Hawaiian Islands. Two southern Californian sites, San Diego and San Nicholas Island, account for about one-fourth of all reports. The site selected most often for climatology (65 times) was the fixed-location station ship 4YN, which happens to be the only site in the HEPC database located in the open ocean near the heavily traveled ship route between southern California and Hawaii.

In its present form, the HEPC climatology has some inherent problems. As previously stated, the climatology assigned to a ship observation is that determined from the HEPC database site closest to the ship location. Over data sparse regions (e.g., the central North Pacific above 30°N), this can result in data extrapolation over considerable distances, and a climatology at a particular ship location which may not be truly representative. For ships located along coastlines, climatology data may be assigned which comes from a HEPC station not truly representative of the open ocean (e.g., Seattle, WA), or even not in the same ocean! An example of the latter can be seen in Figure 5, where Veracruz, MEX and Isla de Cisne, HON were selected by the HEPC Function to represent climatology for several ship reports in Pacific coastal waters off Central America.

The HEPC climatological data may also be incomplete. In this study, the HEPC climatology for Tatoosh Island, WA (a site now nonexistent) was assigned to 22 ship observations off the northwestern U.S. coast. Unfortunately, this site has no useful surface and elevated duct data for most months of the year; as a result, only a few of these ship radiosonde reports could be used in the HRD comparison against climatology. Other HEPC database sites which provided inadequate (i.e., incomplete) data to this study are Seattle, WA, Los Angeles, CA, Isla de Socorro, MEX and Canton Island, Tokelau Islands.

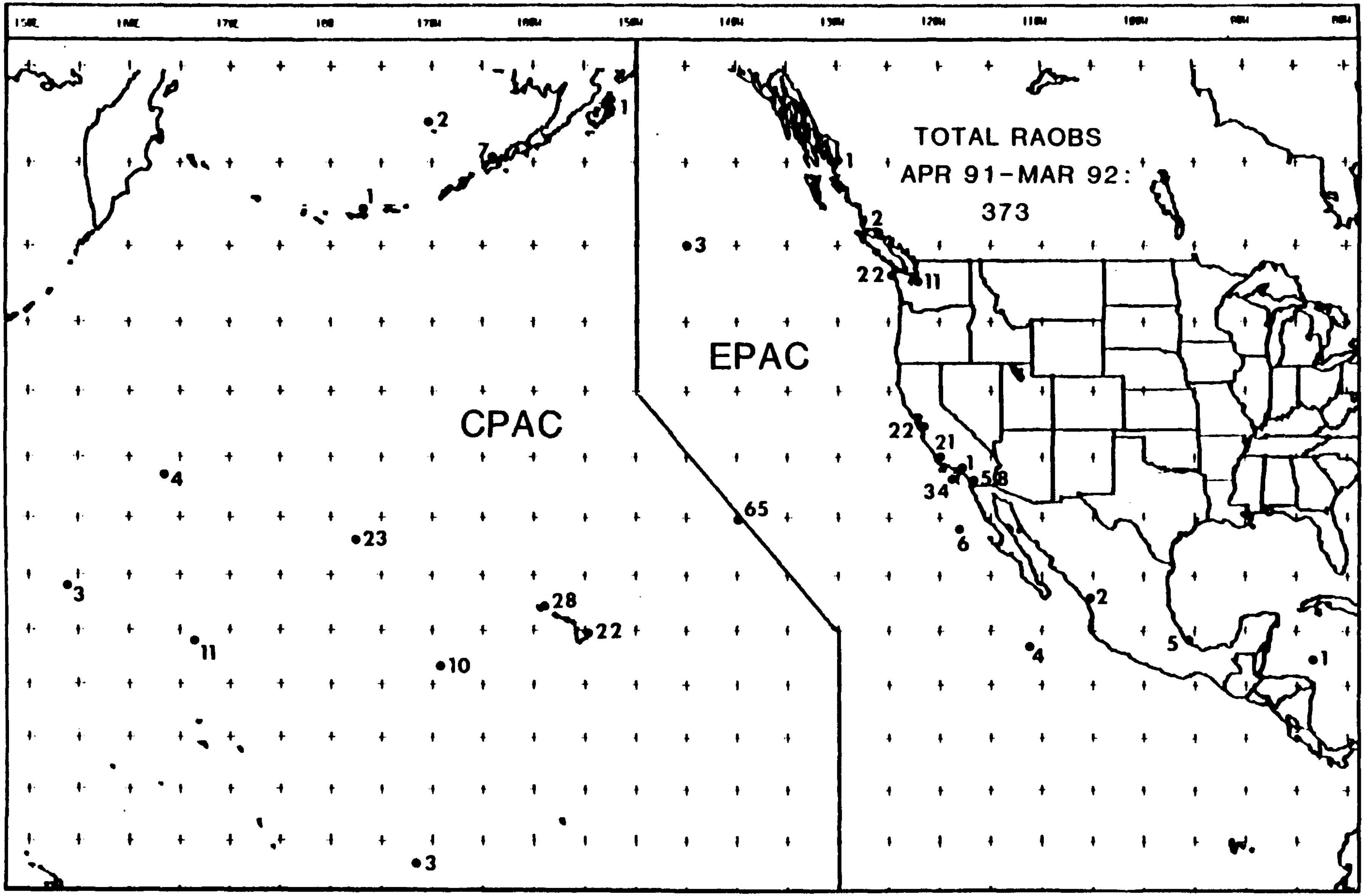


Figure 5. Location of HEPC climatological stations used in this study. A number at a site corresponds to the total number of radiosonde reports assigned by the HEPC Function to that particular climatological site.

Except for a few fixed-location ship stations, the HEPC climatology is determined from island and coastal sites. During certain times of the year, common near-surface refractive features at some land sites may not be truly representative of conditions found over nearby waters. This problem is likely more severe at high elevation coastal sites such as San Diego, San Nicholas Island and Vandenberg AFB, CA, and Seattle, WA, all at elevations above 300 feet. The adverse effect of elevation on HEPC climatological data is most apparent at San Nicholas Island which, at 502 ft msl, is the highest site used in this study. Here, during the months of July and August, the climatological average top of the surface-based duct (the station elevation plus the average surface-based duct thickness) is actually higher than the base of the average elevated duct. Additionally, the percent occurrence of surface-based ducts is greater than that for elevated ducts. Such climatological information might be quite misleading or confusing for a nearby ship located well below the radiosonde site of San Nicholas Island.

The ready availability of the HEPC climatology makes it a convenient database to utilize in the evaluation of a HRD product. As will be seen later, in spite of some serious shortcomings, it can be considered a valuable aid in the operational assessment of lower atmospheric refractive structure.

3.2 Techniques

The verification methods employed in this study include, and expand upon, those utilized by NWOC in their 6 month operational assessment of HRD capabilities in forecasting the occurrence and height of ducting. Statistical indices and skill score discriminants used in the evaluation of categorical forecast of discrete events such as ducting vs. no ducting are derived from two by two contingency tables, a prototype of which follows.

	Forecast (Predictor) 1	0
Observed (Predictand) 1	A	B
0	C	D

A	= no. of "type 1" events which are correctly forecast
B	= no. of "type 1" events which are incorrectly forecast
C	= no. of "type 0" events which are incorrectly forecast
D	= no. of "type 0" events which are correctly forecast
A+B	= no. of "type 1" events which actually occur
C+D	= no. of "type 0" events which actually occur
A+C	= no. of "type 1" events which are forecast to occur
B+D	= no. of "type 0" events which are forecast to occur
A+B+C+D=N	= no. of total events

For verification purposes, "type 1" and "type 0" events are classified as "ducting" and "no ducting," respectively. Unless otherwise indicated, category "A" correct forecasts of ducting are not type (surface-based or elevated) specific. From this table, several statistical indices, the prefigurance (PF), the postagreement (PA), the percent correct (PC) and the false alarm rate (f), are defined for "type 1" events as:

$$PF = A/(A+B) \tag{1}$$

$$PA = A/(A+C) \tag{2}$$

$$PC = ((A+D)/N) x 100 \tag{3}$$

$$f = C/(C+D) \tag{4}$$

The prefigurance, also known as the hit rate (h) or the Power of Detection (POD), is the capability of correctly forecasting an event (viz., ducting), while the postagreement is the reliability of the forecasts that were issued. Note that the false alarm rate as defined in (4) incorporates the correct forecast of non-"type 1" occurrences; this index can be looked upon as the probability that a "type 0" event will be incorrectly forecast.

Woodcock (1976) reviewed different skill score discriminants (used in the literature) and found that the Hanssen and Kuipers (1965) discriminant V provides an acceptable and unbiased measure of forecast accuracy for scientific purposes. The Hanssen and Kuipers discriminant has two propitious qualities not found jointly in other scores. First, V does not depend on the sample relative frequency of the predictand; that is, it is not biased wherein the occurrence of "type 1" events is not equal to the occurrence of "type 0" events. Second, any isopleth of the Hanssen and Kuipers skill score in f,h space has a slope which is unity; in some other skill scores, it is possible for forecasts with $h > f$ to score the same as forecasts with $f > h$. Unfortunately though, as with other scores, V can be quite sensitive to the sample size, with considerable fluctuations possible with slight partition changes of an event of small sample size. However, in this study, sample sizes are sufficiently large so as to minimize this undesirable characteristic.

The Hanssen and Kuipers skill score in contingency table elements is defined:

$$V = (AD - BC) / (A + B)(C + D) = h - f. \tag{5}$$

The score ranges from -1 to 1; -1 implies perfectly wrong forecasts, 0, random performance ($h = f$), and 1, perfect skill. As formulated, this skill score gives forecast successes and failures equal weight. In general, the greater the positive score, the greater the likelihood for high hit rates to be associated with low false alarm rates. While this quality is a widely accepted feature of good forecast skill for scientific purposes, it may indeed be inappropriate for an operational or economic evaluation in which forecast successes and failures are not weighed equally.

Hanssen and Kuipers derived the variance of V as

$$\sigma_V^2 = \frac{N^2 - (4(A + B)(C + D)\,V^2)}{4N(A + B)(C + D)} . \tag{6}$$

The standard deviation of V, σ_v is computed in order to access whether or not differences between predictors (the HRD analysis, the HRD prognosis and climatology) are statistically significant. Given values of V for two predictors n and m, the difference between them will be considered to be statistically significant provided the skill score difference is greater than the standard deviation in the difference times a confidence factor F. Here, the assumption is made that V_n and V_m are samples drawn from a large underlying population whose distribution can be characterized as normal. In mathematical notation, this test for significance is

$$|V_n - V_m| > \sigma_{(V_n - V_m)} * F = \sqrt{\sigma^2_{V_n} + \sigma^2_{V_m}} * F \, . \tag{7}$$

The factor F is determined from the normal probability function, and varies from 1.96 for a 95% confidence level to 2.576 for the 0.99 level of significance.

The probabilistic information available with the HEPC climatology permits a sequence of verification matrices (contingency tables) to be generated by stepping a decision or threshold probability p* through a range of values used in the forecasts. Individual climatological forecasts of ducting/no ducting would be based on the decision "cut-off" probability. For example, if p* was set at 50% and the climatological probability of ducting, as given by the percent occurrence, was greater (less) than 50%, then ducting would (would not) be forecast. A low decision threshold represents a bias toward forecasting occurrence and based on slight evidence; in this case, both the hit rate and the false alarm rate will be high. As the forecaster's decision threshold increases so that progressively stronger evidence is required for a positive forecast, both hit rate and false alarm rates decrease. For a given data set, optimum threshold probabilities may be found which maximize statistical indices or skill scores.

For this study, HEPC summary data are used for two distinct climatic predictors, hereafter designated CLIM1 and CLIM2. Given the forecast time (day or night), the predictor CLIM1 selects the larger of, the percent occurrence of surface-based ducts and the percent occurrence of elevated ducts, as its climatological probability. This value is then compared to the chosen threshold probability; if it is greater (less) than p*, ducting (no ducting) is forecast. For verification, a CLIM1 forecast of ducting is considered a hit (i.e., verifies) even if the observed duct is not the same type (surface-based or elevated) as that on which the CLIM1 predictor is based. In reality, such an occurrence is not common; within this study's data set, the duct type which determines CLIM1 matches that of the observed duct for the great majority of cases. The CLIM1 predictor will be evaluated for threshold probabilities of 30% to 50%, at intervals of 5%.

The CLIM2 predictor uses HEPC summary information corresponding to the day/night category. Its climatological probability is calculated as the sum of the occurrence of surface-based ducts (SD) plus the occurrence of elevated ducts (ED), minus the occurrence of (SD + ED) ducts. This formulation takes into account the fact that (SD + ED) ducts are reported in the HEPC climatology as both SD and ED occurrences. As an example of a CLIM2 calculation, consider Figure 4; here, the climatological probability for February is 16% + 45% - 3%, or 58%. Again, analogous to CLIM1, the calculated CLIM2 value is compared to the selected forecast threshold probability and, if is greater (less) than that value, ducting (no ducting) is forecast. The CLIM2 predictor, which essentially gives the climatological probability for any (type) ducting, will be evaluated for threshold probabilities from 50% to 80%, at intervals of 5%.

The verification of duct height is by means of comparison between observed duct height and thickness data and analogous data from HRD analyses and 36 hr prognoses, and HEPC climatology. Here, only duct observed/duct forecast data are used for verification. Three different duct height verification rates are utilized; one of these, based on the bell curve of Figure 6, was used by NWOC in their 6 month (April-September) evaluation. This particular verification rate gives 100% credit if the forecast duct height is within 1000 ft of the observed, 50% credit for a forecast within 2000 ft and no credit if off by 3000 ft or more. A second duct height verification rate determines the percentage of forecast ducts within 1000 ft of the observed. The final rate, the most rigorous of the three, gives the percentage of duct height predictions which overlap the observed. For observations with more than one duct, the duct closest to the predicted duct is used for verification. In the case of the HEPC climatology, which always gives two ducts - surface-based and elevated, the duct chosen for verification is that (type) which is

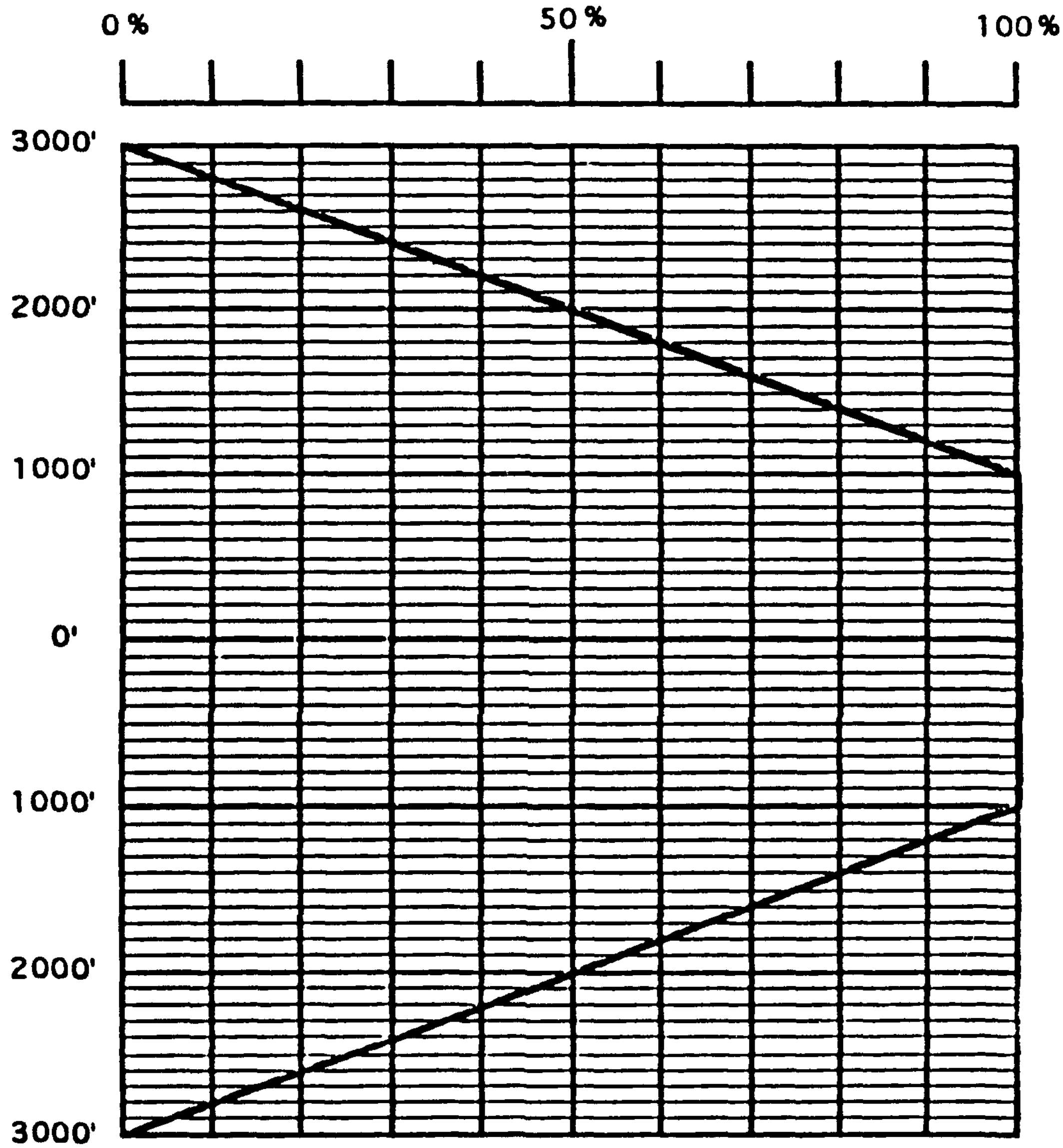


Figure 6. Bell curve used for duct height verification (from NWOC, 1991).

dominant (i.e., most frequent). Such a determination is made by a comparison of the percent occurrences for surface-based and elevated ducts in the day/night column of the monthly summary. Provided that the percent occurrences are equal, both ducts are considered, and the one closest to the observed is used for validation. Since average duct top and thickness information is only available in the HEPC climatology as monthly (day/night) averages, only CLIM2 predictions of duct height are evaluated, for threshold probabilities of 50% to 80%, at 5% intervals.

3.3 Results

In its operational environment, the NWOC HRD is subject to a wide variety of factors and circumstances which bear directly on its final production. While the relative importance of some of these factors (such as forecast location, season and time of day) may be explored statistically, other important factors, such as individual forecaster skill and day-to-day production procedures, are virtually impossible to quantify. Verification statistics based on a large sample of independent HRD products should be viewed in the context of "average" expected product accuracy and skill in an operational environment.

3.3.1 Duct Occurrence

Before evaluating the results for the full data set and its seasonal, geographical and temporal subsets, verification statistics for the HRD analysis and 36 hr prognosis computed on a monthly basis are presented (Figures 7a and 7b, respectively). The most salient result portrayed in this figure is the high degree of month-to-month similarity between the HRD analysis and the 36 hr prognosis verification statistics over the entire 12 month period. This desirable result strongly suggests that NWOC meteorological office personnel strictly adhered to one of the key points emphasized in the NWOC procedural guidelines for production of the HRD charts - the maintenance of continuity between the forecast and the analysis.

In general, since some monthly data samples are not sufficiently large or truly representative of the full (12 month) observational data set, sound statistical inferences can not be fairly drawn from month-to-month differences of individual performance indices depicted in Figure 7. To illustrate this point, consider the false alarm rate index for the months of September (1.0) and November (0.0); for these two months, only 2 and 1 "no ducting" events, respectively,

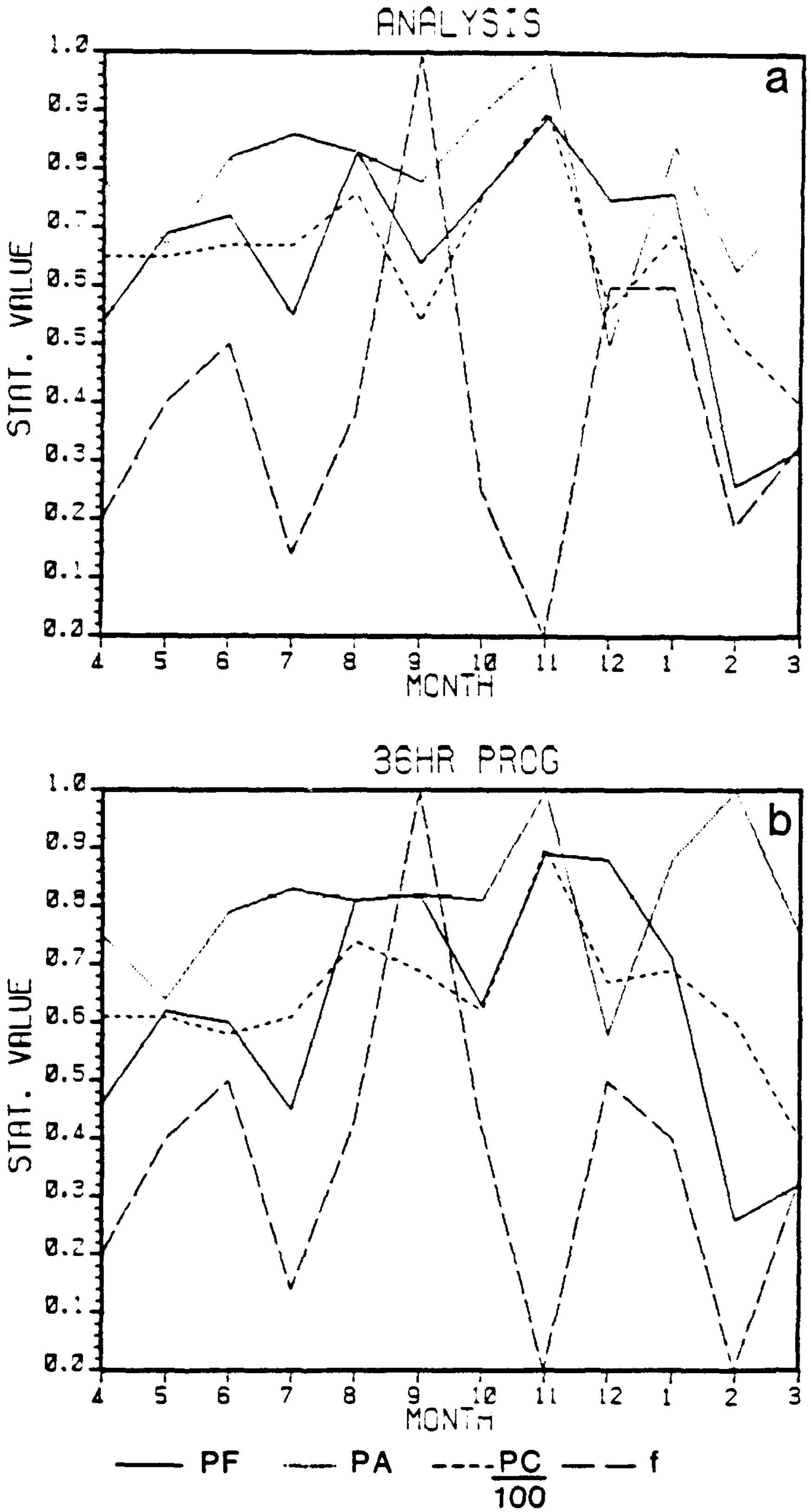


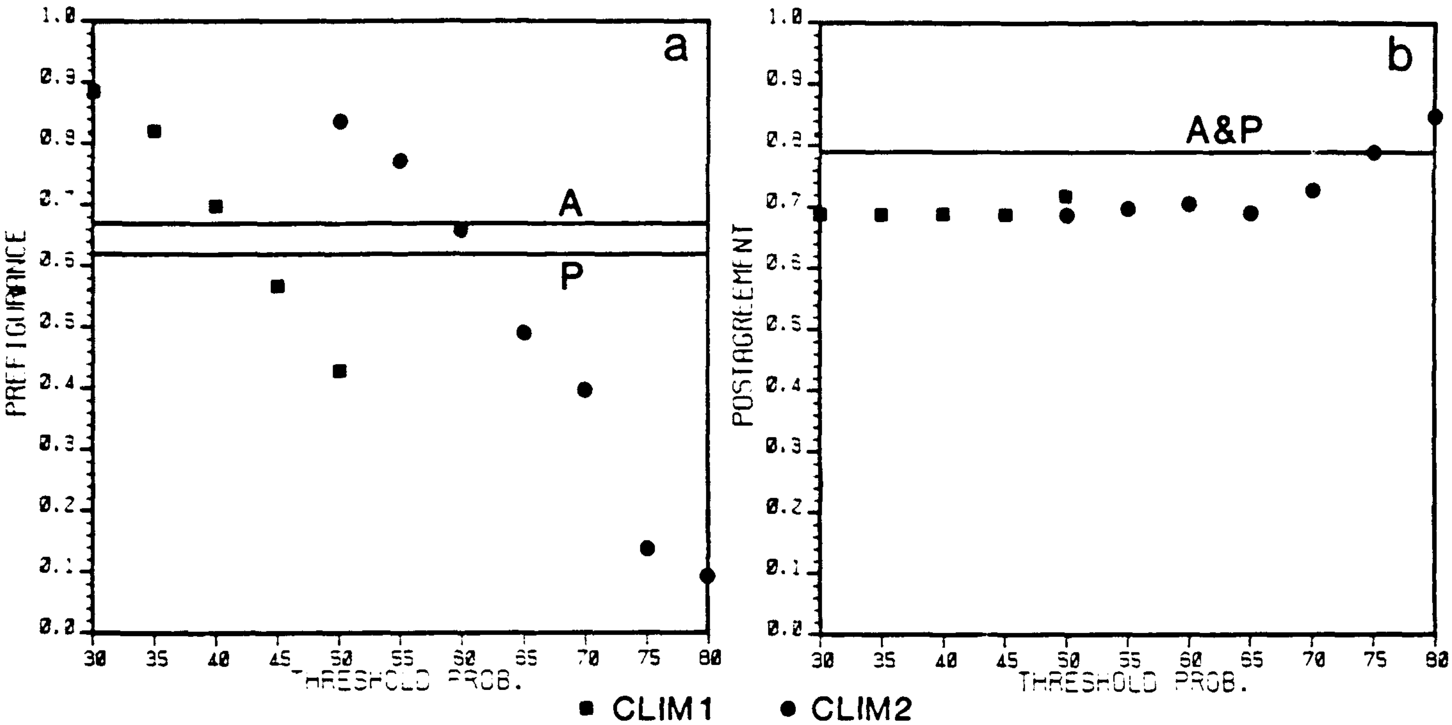
Figure 7. (a) Monthly HRD analysis prefigurance, postagreement, percent correct and false alarm rate indices over the period April 1991 - March 1992. (b) same as (a), except for the HRD 36 hr prognosis.

determined this index! An example of the adverse effect of non-representative monthly samples on statistical comparisons is best illustrated in Figure 7 by the sharp downturn of the prefigurance and the percent correct indices (for both the analysis and 36 hr prognosis) from January into February and March. Careful scrutiny of the February-March data reveals a considerable number of incorrectly forecast, abnormal (based on climatic expectations) "ducting" radiosonde observations in regions previously (April-January) either poorly sampled (i.e., the Pacific above 45N) or not sampled at all (i.e., offshore Central American waters).

Performance statistics for the HRD analysis and 36 hr prognosis, and the HEPC climatology, based on the full data set, are given in Figures 8a-d. The HRD analysis prefigurance and percent correct statistics are observed to be only slightly better than those for the HRD 36 hr prognosis, while the opposite is true for the false alarm rate. Both predictors have virtually the same postagreement values. The Hanssen and Kuipers skill scores for the HRD analysis and 36 hr prognosis are 0.31 and 0.28, respectively; these values are not significantly different.

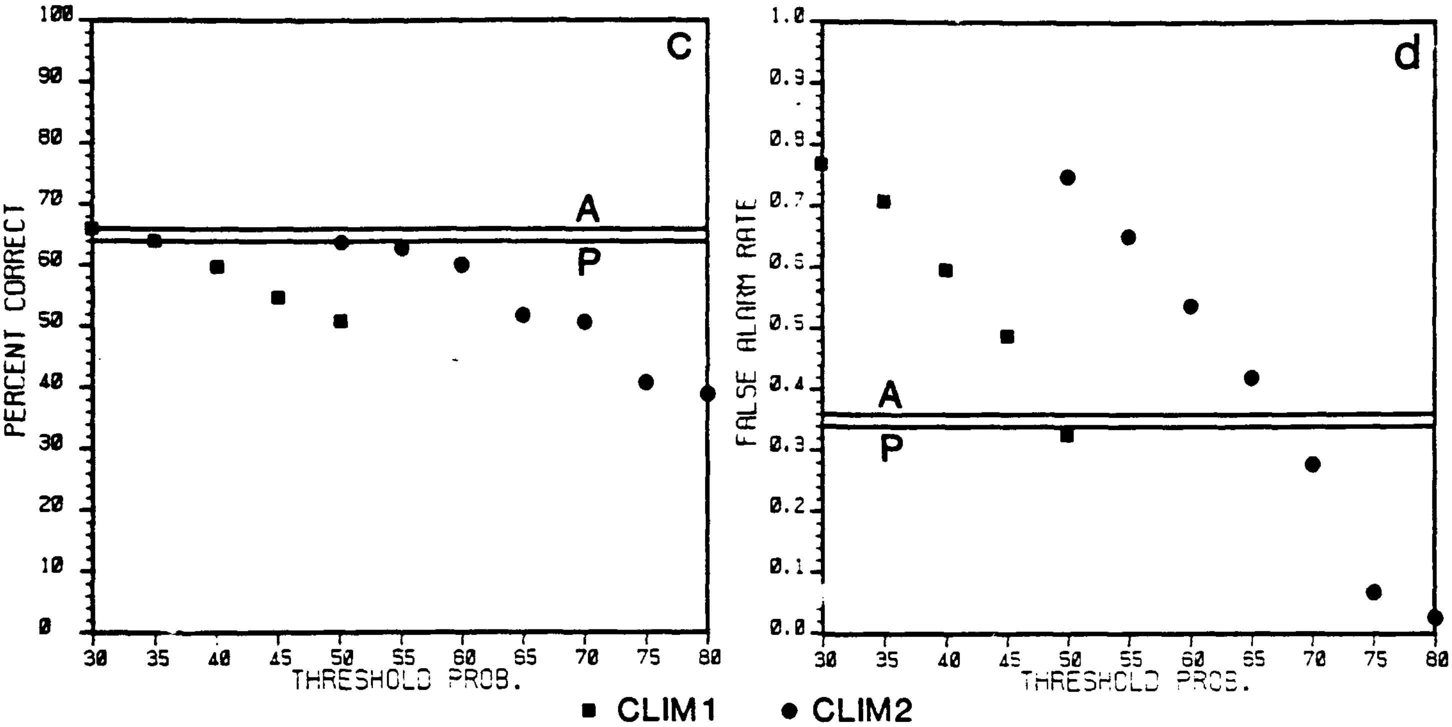
The two climatology predictors (CLIM1 and CLIM2) show marked variability, as a function of threshold probability, in three out of the four statistical indices depicted in Figure 8. The one exception is the postagreement index, which varies little over the given ranges for p^*. As expected, large (small) hit and false alarm rates correspond to low (high) decision thresholds. Comparisons of CLIM1 and CLIM2 performance statistics for the full data set do not indicate any clear preference of one over the other as a predictor. In terms of the Hanssen and Kuipers discriminant, optimum forecast skill is not found at $p^* = 50\%$ (i.e., simple probability); rather, the largest skill scores ($V = 0.12$ for both predictors) are found at $p^* = 30\%$ for CLIM1 and $p^* = 60\%$ for CLIM2.

A visual examination of Figures 8a-d indicates that, at any selected threshold probability p^*, no more than two of the four CLIM1 or CLIM2 statistical indices are better than the analogous HRD 36 hr prognosis values. At lower threshold probabilities (30 - 40% for CLIM1, 50 - 60% for CLIM2), climatology predictors have larger hit rates and comparable percent correct values to the HRD prognosis; additionally, the forecast reliability (the PA index) is not significantly lower than that for the HRD product. On the other hand, the false alarm rates for both CLIM1 and CLIM2 at these same threshold probabilities are much larger (by about twice)



PREDICTOR		PREFIGURANCE	POSTAGREEMENT
HRD ANALYSIS		164/246 = .67	164/208 = .79
HRD 36 HR PROG		154/247 = .62	154/196 = .79
CLIM1	p* ≥ 30%	205/231 = .89	205/297 = .69
	p* ≥ 35%	189/231 = .82	189/274 = .69
	p* ≥ 40%	161/231 = .70	161/232 = .69
	p* ≥ 45%	132/231 = .57	132/190 = .69
	p* ≥ 50%	99/231 = .43	99/138 = .72
CLIM2	p* ≥ 50%	198/235 = .84	198/289 = .69
	p* ≥ 55%	182/235 = .77	182/261 = .70
	p* ≥ 60%	156/235 = .66	156/221 = .71
	p* ≥ 65%	116/235 = .49	116/167 = .69
	p* ≥ 70%	94/235 = .40	94/128 = .73
	p* ≥ 75%	33/235 = .14	33/42 = .79
	p* ≥ 80%	22/235 = .09	22/26 = .85

Figure 8. (a) Prefigurance and (b) postagreement statistics for the HRD analysis and 36 hr prognosis, and for the HEPC climatology predictors CLIM1 and CLIM2 at selected threshold probabilities, based on the full (April 1991 - March 1992) data set.



PREDICTOR		PERCENT CORRECT	FALSE ALARM RATE
HRD ANALYSIS		(243/369) 66%	44/123 = .36
HRD 36 HR PROG		(235/370) 64%	42/123 = .34
CLIM1	p* ≥ 30%	(232/350) 66%	92/119 = .77
	p* ≥ 35%	(223/350) 64%	85/119 = .71
	p* ≥ 40%	(209/350) 60%	71/119 = .60
	p* ≥ 45%	(193/350) 55%	58/119 = .49
	p* ≥ 50%	(179/350) 51%	39/119 = .33
CLIM2	p* ≥ 50%	(228/356) 64%	91/121 = .75
	p* ≥ 55%	(224/356) 63%	79/121 = .65
	p* ≥ 60%	(212/356) 60%	65/121 = .54
	p* ≥ 65%	(186/356) 52%	51/121 = .42
	p* ≥ 70%	(181/356) 51%	34/121 = .28
	p* ≥ 75%	(145/356) 41%	9/121 = .07
	p* ≥ 80%	(139/356) 39%	4/121 = .03

Figure 8. (c) Percent correct and (d) false alarm rate statistics for the HRD analysis and 36 hr prognosis, and for the HEPC climatology predictors CLIM1 and CLIM2 at selected threshold probabilities, based on the full (April 1991 - March 1992) data set.

than the false alarm rate determined for the HRD 36 hr prognosis. Tests of statistical differences between the HRD analysis and 36 hr prognosis skill scores (V = 0.31 and 0.28, respectively) and those for the climatic predictors at their "optimum" (both V = 0.12) threshold probabilities (30% for CLIM1, 60% for CLIM2) indicate that such differences are significant at the 95% confidence level, but not at the 99% confidence level.

In order to investigate any seasonality in HRD product performance, the full data set was divided into two 6 month periods. The first period (summer - April to September) corresponds to that evaluated by NWOC (1991); the second (winter) period covers previously unevaluated data from October 1991 through March 1992. The summer period consists of 212 data which are geographically well distributed (107 EPAC, 105 CPAC); on the other hand, the winter period consists of 161 data, of which 116 are classified as EPAC and only 45 as CPAC. This nonuniformity in the geographical distribution of the summer and winter data sets is likely to adversely affect (i.e., compromise) the statistical integrity of the seasonal comparisons; as a consequence, such comparisons need to be viewed with some caution.

Figures 9a-d depict the seasonal performance statistics for the HRD product and the HEPC climatology. Prefigurance and percent correct indices are observed to be slightly higher for the HRD product during the summer period, while postagreement and false alarm rate indices are better for the winter period. In either season, statistical differences between the HRD analysis and 36 hr prognosis are quite small for all indices. The Hanssen and Kuipers skill score ranges from V = 0.34 for the HRD analysis during summer to V = 0.27 for the HRD analysis during the winter period; the score is the same (0.29) in both seasons for the HRD 36 hr prognosis.

In general, the prefigurance and percent correct indices for both CLIM1 and CLIM2 are higher in the summer period than during winter; the only exception to this is the percent correct index at the lowest threshold probability (30% for CLIM1, 50% for CLIM2). On the other hand, both the postagreement and false alarm rate indices for the climatological predictors are better for the winter period than in summer. Interestingly, these seasonal CLIM1 and CLIM2 statistical trends for all four indices are the same as those with the HRD product. Skill scores for the CLIM1 predictor are at least 0.10 larger at all threshold probabilities for the winter period than during the summer period; while the largest skill score for winter is 0.25

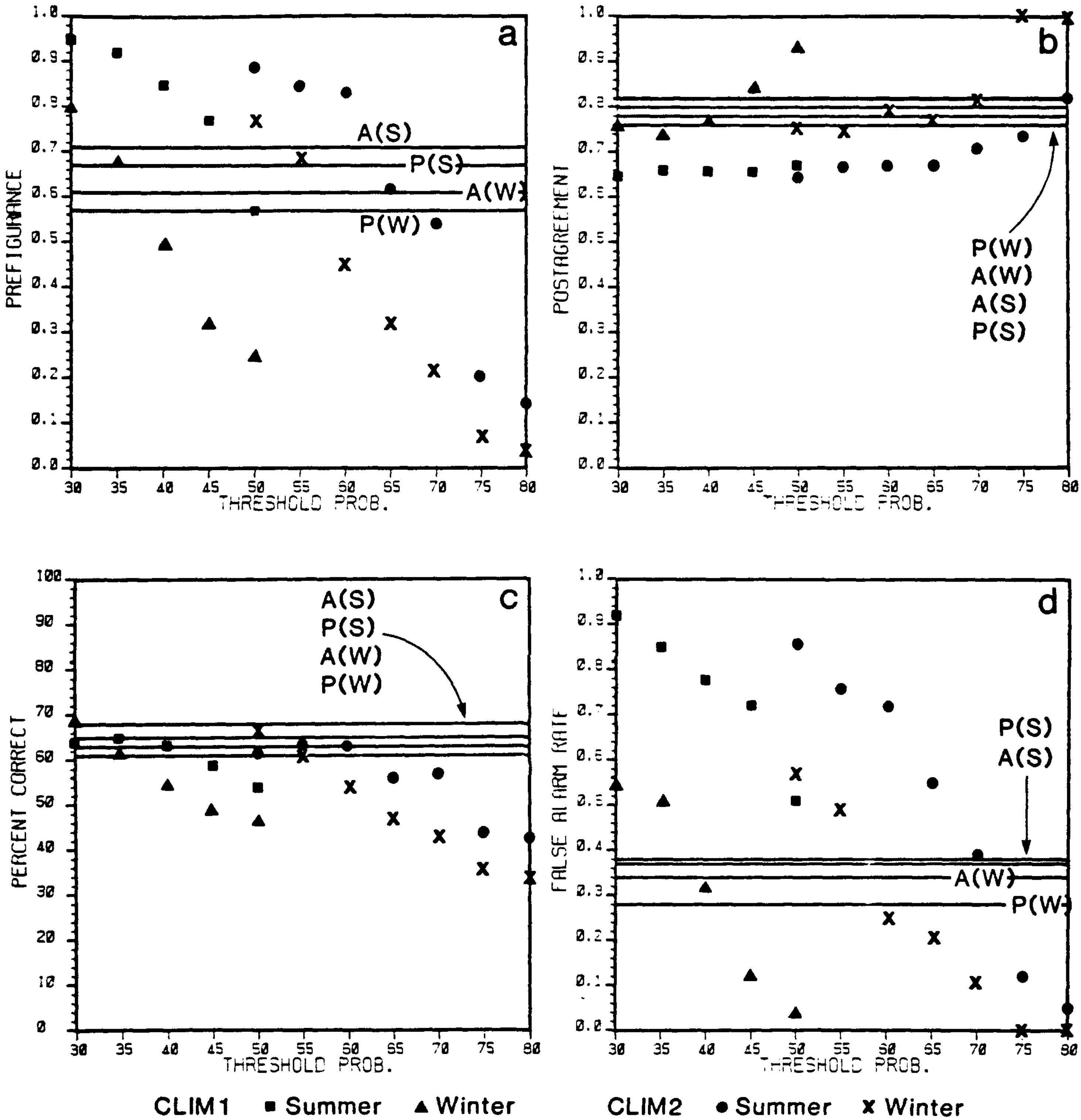


Figure 9. (a) Prefigurance, (b) postagreement, (c) percent correct and (d) false alarm rate statistics for the HRD analysis and 36 hr prognosis, and for the HEPC climatology predictors CLIM1 and CLIM2 at selected threshold probabilities, based on seasonal (summer and winter) data subsets.

(at p* = 30%), the largest for the summer is only 0.07 (at p* = 35% and 40%). CLIM2 skill scores for the winter period are better than those for summer at lower threshold probabilities (p* ≤ 65%), and slightly worst at higher probabilities (p* ≥ 70%) . Winter CLIM2 skill scores are between V = 0.17 and V = 0.20 for threshold probabilities of 50% to 60% . For the summer data set, V is largest (0.15) for CLIM2 at p* = 70%; at this threshold probability, the false alarm rate is about the same as that for the HRD product.

For the April - September period, differences in skill scores between the CLIM1 predictor (at all threshold probabilities) and the HRD analysis and prognosis are significant at the 95% confidence level. Differences in skill scores between the CLIM2 predictor and the HRD product for the same period are significant at the 95% confidence level at all threshold probabilities except p* = 70% and p* = 60% (HRD prognosis only). On the other hand, skill score differences for the October - March period are not significant between CLIM1 and the HRD product nor between CLIM2 and the HRD analysis, and are only significant (at the 0.95 level of confidence) between the HRD 36 hr prognosis and CLIM2 for a CLIM2 threshold probability of 80%.

Figures 10a-d depict performance statistics for the HRD product and the HEPC climatology based on regional subsets. As previously discussed, the EPAC radiosondes are well distributed throughout the April - March period, whereas fully 70% of the total for the CPAC region correspond to the summer (April - September) period. While not statistically significant, HRD prefigurance and postagreement statistics are noticeably larger for the eastern Pacific than for the central Pacific; on the other hand, false alarm rates are lower for the CPAC region. The percent of correct forecasts is almost the same for both regions (~66%). For the EPAC region, differences between the HRD analysis and 36 hr prognosis are quite small for all four performance statistics. The HRD analysis is noticeably better in assessment capability (i.e., prefigurance) than the HRD 36 hr prognosis for the CPAC data set (0.62 to 0.51); this results in a relatively low skill score for the HRD prognosis (0.20) compared to that for the analysis (0.31). Skill scores for the HRD analysis and 36 hr prognosis, based on the EPAC data set, are comparable (0.28 and 0.30, respectively).

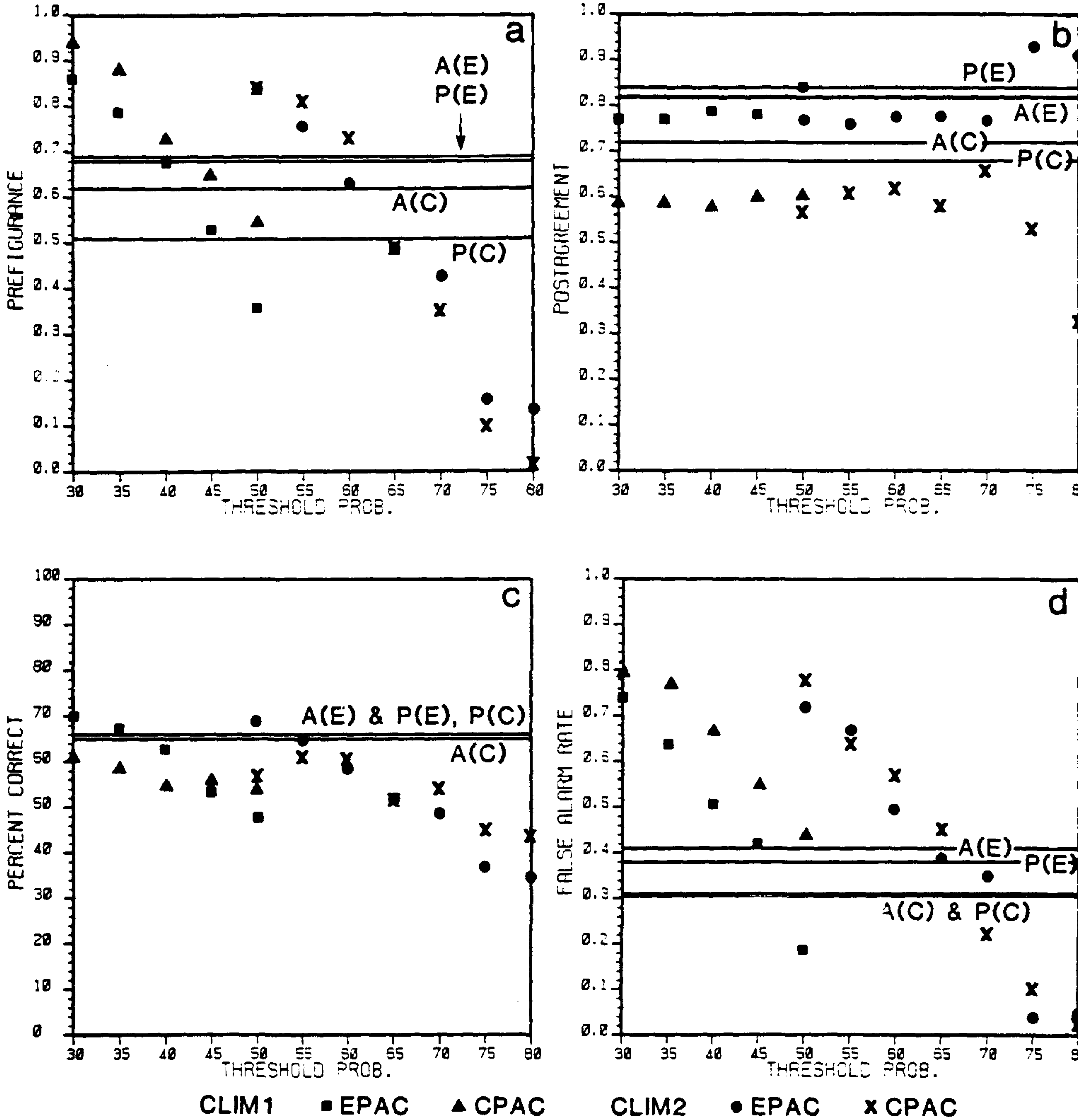


Figure 10. (a) Prefigurance, (b) postagreement, (c) percent correct and (d) false alarm rate statistics for the HRD analysis and 36 hr prognosis, and for the HEPC climatology predictors CLIM1 and CLIM2 at selected threshold probabilities , based on geographical (EPAC and CPAC) data subsets.

With respect to the climatology predictors, the most noteworthy statistic of Figure 10 is the postagreement; analogous to the HRD product, forecast reliabilities based on climatology are observed to be noticeably higher for the eastern Pacific region than for the central Pacific. At all threshold probabilities, CLIM1 hit and false alarm rates are higher for the CPAC region than for the EPAC region. Except at the lowest threshold probability ($p^* = 30\%$), CLIM1 skill scores are slightly higher for the eastern Pacific. While CLIM2 skill scores for the EPAC region are rather consistent ($V = 0.08$ to 0.13) over the range $p^* = 50\%$ to $p^* = 80\%$, CLIM2 scores for the central Pacific vary considerably, from $V = 0.17$ at $p^* = 55\%$ and 60%, to $V \leq 0.0$ (i.e., no skill) at $p^* \geq 75\%$.

In general, skill score differences between the HRD product and climatology (both CLIM1 and CLIM2) are not significant for the eastern Pacific data set; the only exception to this is the difference between CLIM2 (at $p^* = 70\%$) and the HRD 36 hr prognosis, which is significant at the 95% confidence level. For the central Pacific, differences between the HRD 36 hr prognosis and either CLIM1 or CLIM2 are not significant at any of the selected threshold probabilities. On the other hand, significant differences in skill scores (at the 0.95 level of confidence) are found between the HRD analysis and CLIM1 (at $p^* = 40\%$) , and between the analysis and CLIM2 at various scattered threshold probabilities ($p^* = 50\%$, 65%, 75% and 80%).

Performance statistics for the HRD analysis and 36 hr prognosis, and the HEPC climatology, based on day and night categories (180 and 193 total observations, respectively), are shown in Figures 11a-d. With the exception of the false alarm rate, statistical indices derived for the HRD analysis and prognosis from nighttime data are all better than analogous indices computed from "day only" events; for the HRD analysis, the nighttime prefigurance is considerably higher than that for daytime (0.72 to 0.60). For the day data set, all four performance statistics are virtually the same for the HRD analysis and 36 hr prognosis; for the night data set, the largest difference between the HRD analysis and prognosis is for the prefigurance (0.09). The Hanssen and Kuipers skill scores for the nighttime HRD analysis and 36 hr prognosis (0.34 and 0.29, respectively) are slightly better than analogous scores for the "day only" data set.

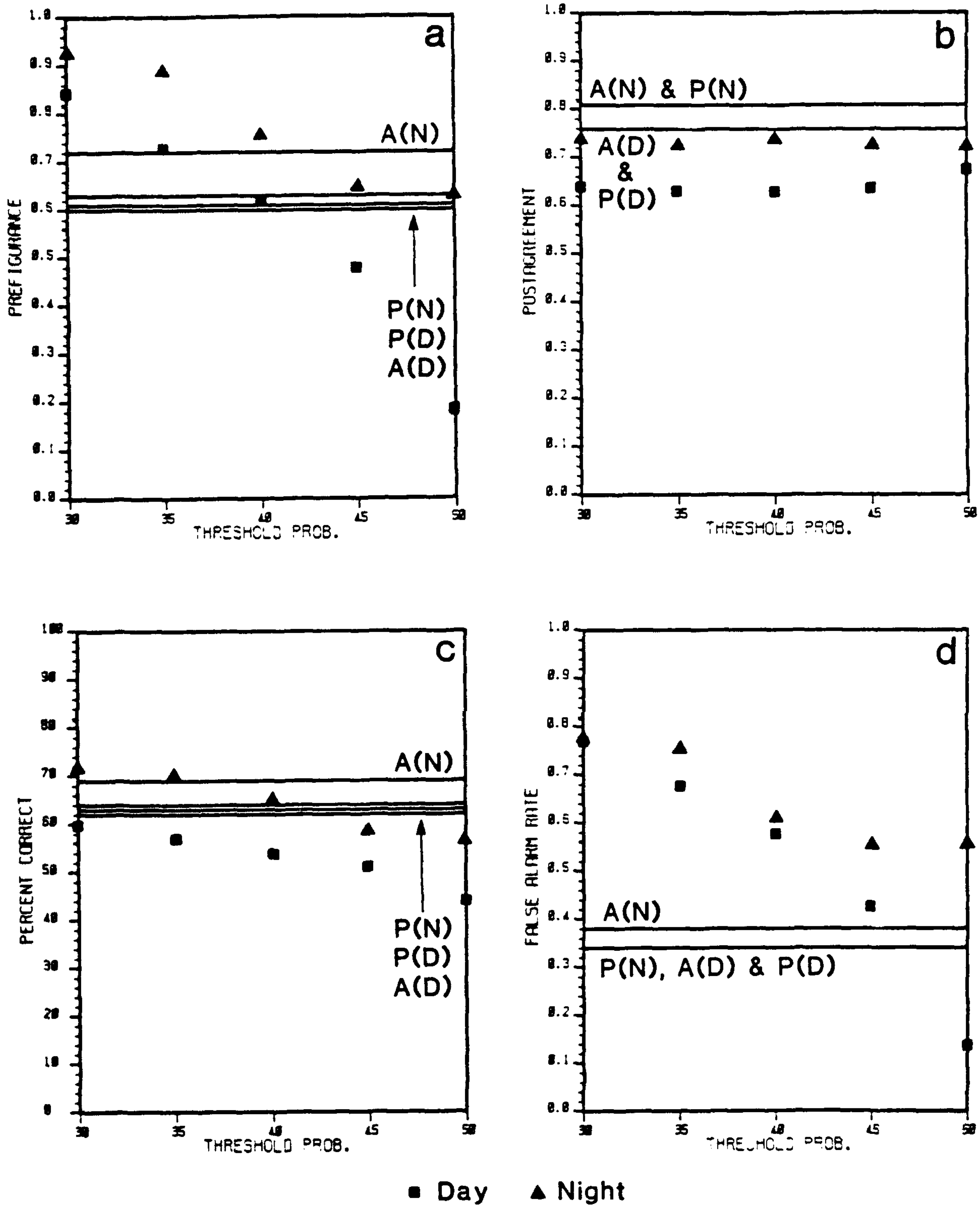
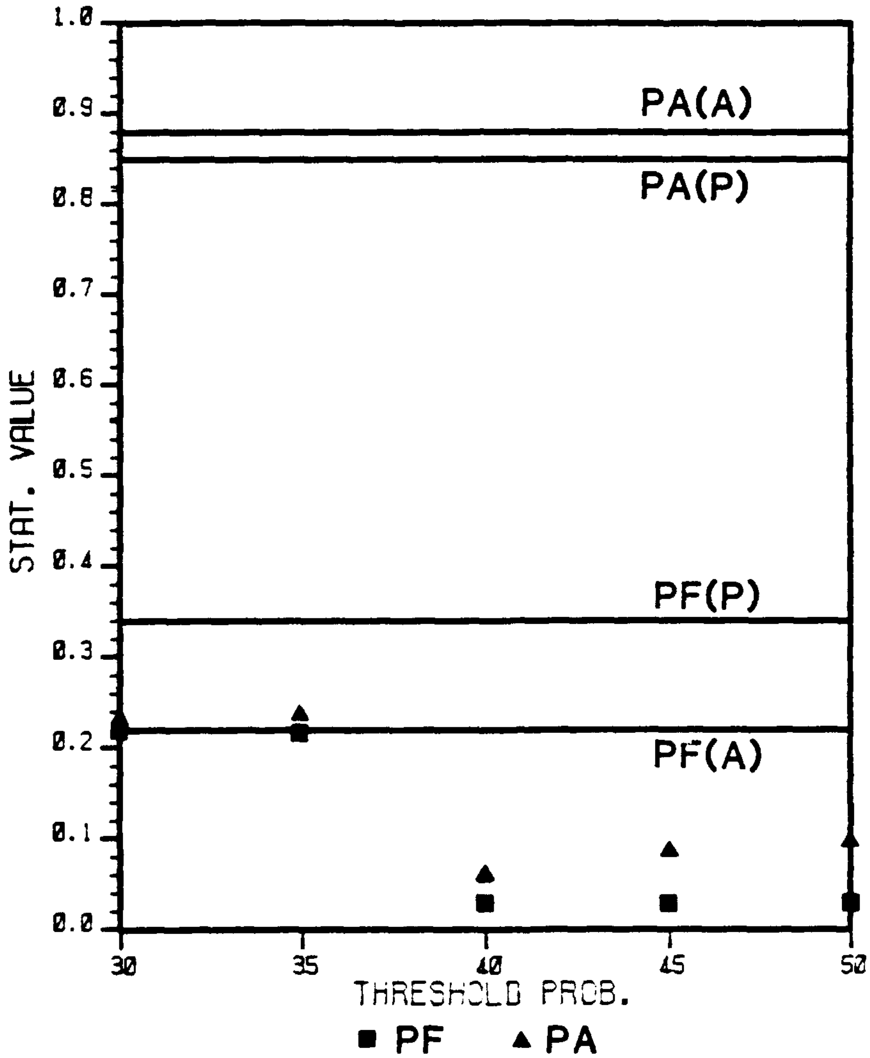


Figure 11. (a) Prefigurance, (b) postagreement, (c) percent correct and (d) false alarm rate statistics for the HRD analysis and 36 hr prognosis, and for the HEPC climatology predictor CLIM1 at selected threshold probabilities, based on temporal (day and night) data subsets.

At all selected threshold probabilities, CLIM1 prefigurance, postagreement and percent correct indices are determined to be better at nighttime than during the day, while the opposite is true for the false alarm rate index. These temporal trends in CLIM1 statistical indices are analogous to what is observed for the HRD product. CLIM1 skill scores at all threshold probabilities are higher for the nighttime data set; in fact, the lowest skill score for the nighttime data set ($V = 0.07$ at $p^* = 50\%$) is slightly higher than the best score for the "day only" data set (at $p^* = 30\%$) .

For night events, differences in skill scores between the CLIM1 predictor and the HRD prognosis are not significant; differences between the HRD analysis and CLIM1 are only significant (at a 0.95 level of confidence) at high threshold probabilities ($p^* \geq 45\%$). A comparison of daytime CLIM1 and HRD 36 hr prognosis skill scores indicates that differences are significant (at the 95% confidence level) at all threshold probabilities except $p^* = 30\%$. In spite of overall low CLIM1 daytime skill scores, differences between the HRD analysis and CLIM1 are only significant at two threshold probabilities ($p^* = 40\%$ and 50%) for the daytime data set.

Up to this point, statistical results for duct occurrence have been presented using combined surface-based and elevated duct occurrences; results are next presented for only surface-based ducting. Of the total 248 "ducting" radiosonde observations, only 32 were surface-based and most (25) occurred offshore from central and southern California. Figure 12 presents surface ducting prefigurance and postagreement statistics for the HRD and CLIM1 predictors. Although statistically firm conclusions can not be drawn from small sample sizes, the results depicted in Figure 12 do suggest the relative capabilities of each predictor for surface duct forecasting. The most striking aspect of Figure 12 is the large difference between HRD forecast reliability (PA) and hit rate (PF); although the HRD product was quite limited in its capability to correctly assess occurrences of surface ducting, on the relatively few occasions when ducts were forecast, such forecasts were very reliable. Statistical values for CLIM1 are observed to be highest at lower threshold probabilities ($p^* = 30\%$ and 35%). Finally, although CLIM1 forecast reliability is observed to be much lower than that for the HRD product, prefigurance values for CLIM1 are identical to those for the HRD analysis (viz., quite low) at $p^* = 30\%$ and 35%.



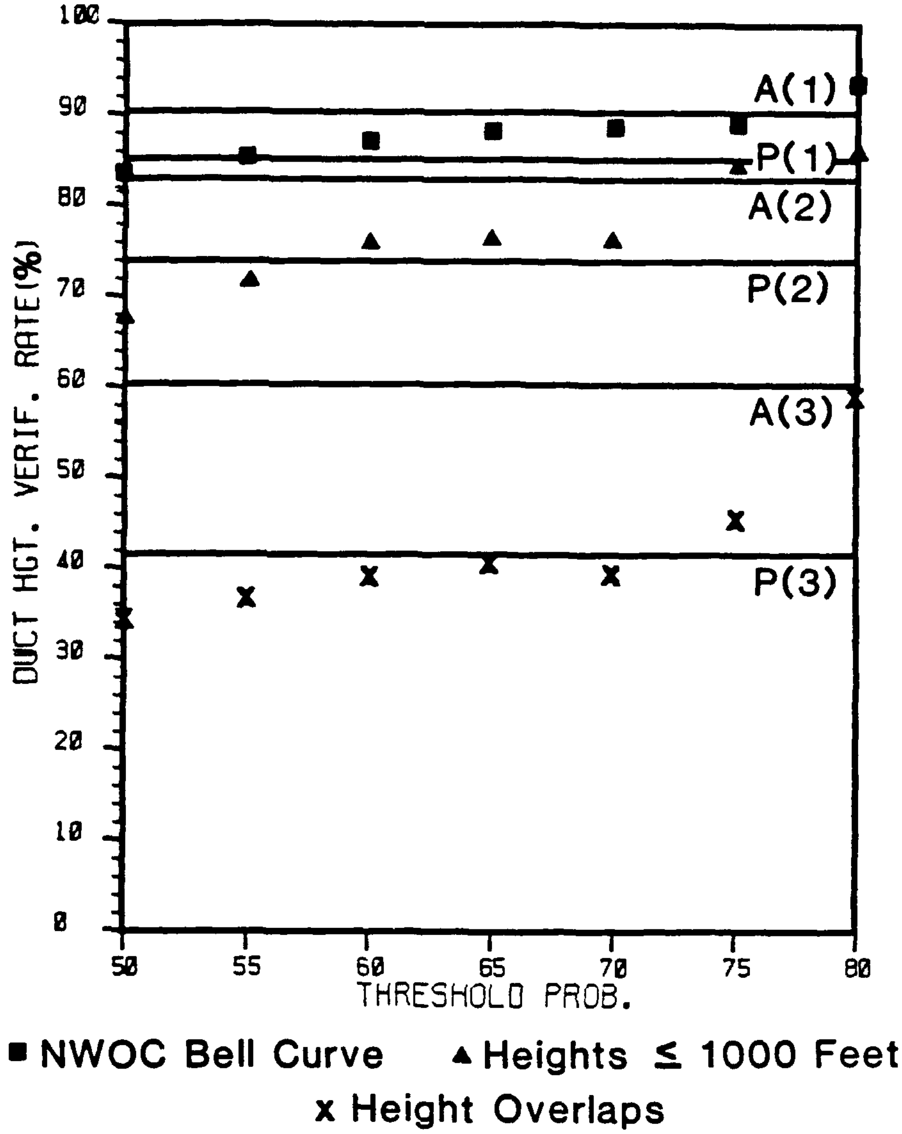
PREDICTOR		PREFIGURANCE	POSTAGREEMENT
HRD ANALYSIS		7/32 = .22	7/8 = .88
HRD 36 HR PROG		11/32 = .34	11/13 = .85
CLIM1	p* ≥ 30%	7/32 = .22	7/31 = .23
	p* ≥ 35%	7/32 = .22	7/29 = .24
	p* ≥ 40%	1/32 = .03	1/16 = .06
	p* ≥ 45%	1/32 = .03	1/11 = .09
	p* ≥ 50%	1/32 = .03	1/10 = .10

Figure 12. Surface-based ducting prefigurance and postagreement statistics for the HRD analysis and 36 hr prognosis, and for the HEPC climatology predictor CLIM1 at selected threshold probabilities, derived from all available data over the period April 1991 - March 1992.

3.3.2 Duct Height

Figure 13 presents duct height verification results for the HRD analysis and 36 hr prognosis, and the HEPC climatology (CLIM2), for the full April - March data set. Three separate verification rates are used: (1) the percentage based on the NWOC bell curve (Fig. 6); (2) the percentage of forecast ducts within 1000 ft of the observed; and (3) the percentage of forecast ducts which overlap the observed. In terms of a measure of duct height verification, the first rate (1) is the least demanding and the third (3), the most rigorous. The verification statistics of Figure 13 support this assertion, with verification rates for method (1) much larger than those for method (3). For all three verification rates, the HRD analysis is better than the HRD prognosis; differences between the two are most noticeable for the third verification scheme, which requires a forecast duct to overlap the observed. Verification results (all three methods) indicate that climatological (CLIM2) forecasts of duct height, at all threshold probabilities, compare very favorably with 36 hr forecasts from the HRD. As an example, consider $p^* = 60\%$; at this threshold probability, the number of duct forecast/duct observed data for the HRD prognosis and CLIM2 are nearly equal (154 and 156, respectively). Results indicate that, for two out of the three verification rates, CLIM2 duct height forecasts at this decision threshold ($p^* = 60\%$) were slightly better than those provided by the HRD 36 hr prognosis.

Duct height verification statistics for summer and winter periods (137 and 111 observed ducts, respectively) are depicted in Figures 14a-c. For the summer period, verification rates (1) and (2) for the HRD prognosis are quite comparable to those for the HRD analysis; on the other hand, the HRD analysis is noticeable better than the prognosis for the verification rate based on duct overlaps (3). During the winter season, all three verification rates for the HRD 36 hr prognosis are observed to be considerably lower than analogous rates for the HRD analysis. In general, CLIM2 duct height verification rates are noticeably better for the summer period than the winter. While verification rates at very high threshold probabilities ($p^* \geq 75\%$) indicate an opposite seasonal trend, results at these threshold probabilities are not statistically sound since they were determined from very small sample sizes. For the summer data set, CLIM2 height verification rates (over the entire range $p^* = 50\%$ to 80%) are all comparable to those determined from the HRD 36 hr prognosis. Wintertime CLIM2 height verification rates are not



PREDICTOR	DUCT HEIGHT VERIFICATION RATE				
	(1) Bell Curve	(2) Within 1000 ft		(3) Overlaps	
HRD ANALYSIS	90.3%	136/164	82.9%	99/164	60.4%
HRD 36 HR PROG	85.2%	114/154	74.0%	64/154	41.6%
CLIM2 $p^* \geq 50\%$	83.5%	135/198	68.2%	68/198	34.3%
$p^* \geq 55\%$	85.6%	131/182	72.0%	67/182	36.8%
$p^* \geq 60\%$	87.2%	119/156	76.3%	61/156	39.1%
$p^* \geq 65\%$	88.3%	89/116	76.7%	47/116	40.5%
$p^* \geq 70\%$	88.7%	72/94	76.6%	37/94	39.4%
$p^* \geq 75\%$	89.4%	28/33	84.8%	15/33	45.5%
$p^* \geq 80\%$	93.8%	19/22	86.4%	13/22	59.1%

Figure 13. Duct height verification rates for the HRD analysis and 36 hr prognosis, and the CLIM2 predictor, based on the full (April 1991- March 1992) data set. The three verification rates are: (1) the percentage based on the NWOC bell curve of Fig. 6, (2) the percentage of forecast ducts within 1000 ft of the observed and (3), the percentage of forecast ducts which overlap the observed.

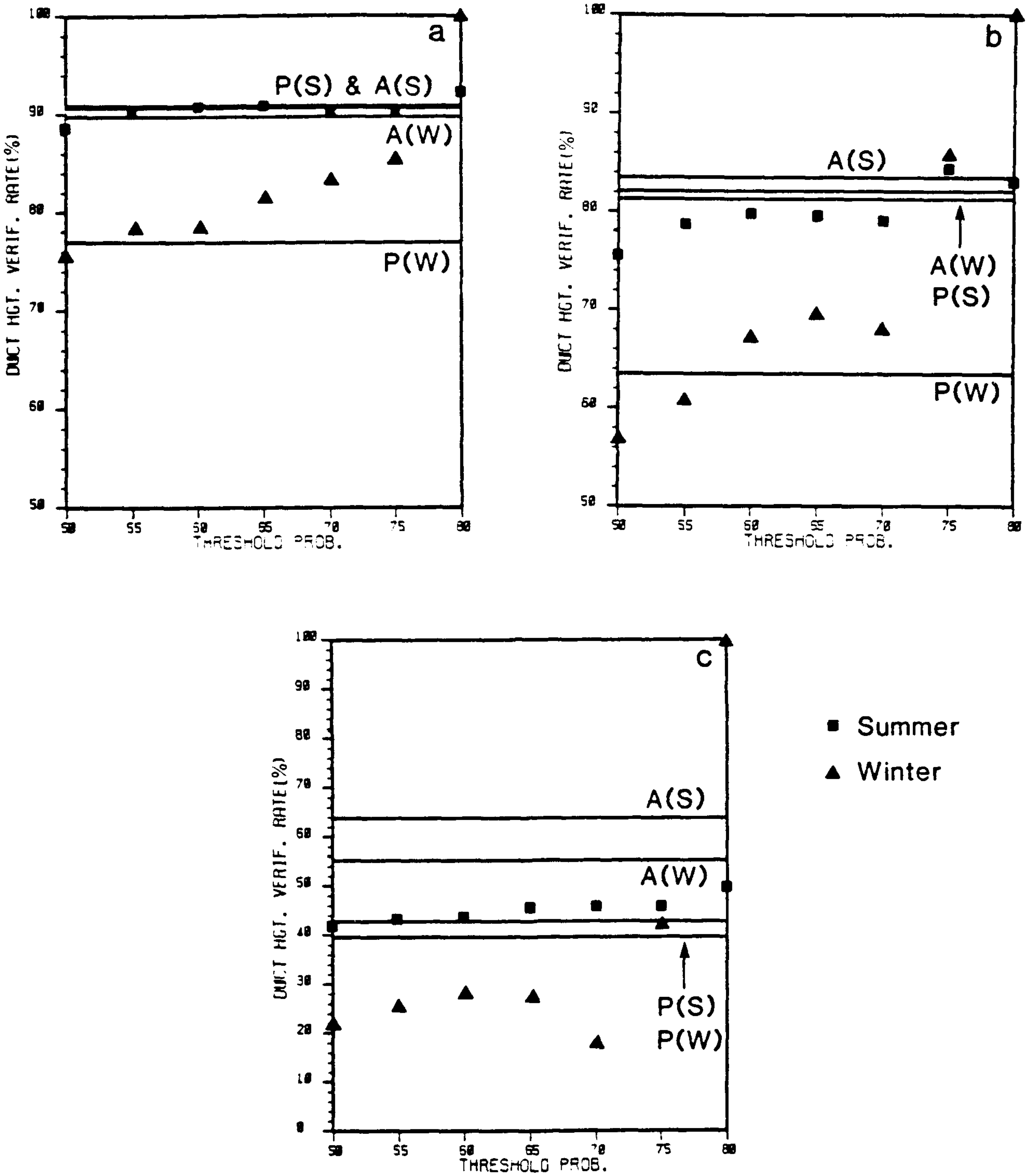


Figure 14. Duct height verification rates for the HRD analysis and 36 hr prognosis, and the CLIM2 predictor at selected threshold probabilities, based on seasonal (summer and winter) subsets: (a) the percentage based on the NWOC bell curve of Fig. 6, (b) the percentage of forecast ducts within 1000 ft of the observed and (c), the percentage of forecast ducts which overlap the observed.

as consistent as those for summer, exhibiting considerable variability as a function of p*. Moreover, wintertime comparisons between the HRD prognosis and CLIM2 are also not consistent for all three verification rates; depending on the verification scheme chosen, the HRD prognosis may be a better (or worse) predictor of duct height than climatology.

Due to a large disparity in the number of EPAC and CPAC duct observations (165 and 83, respectively) and the likelihood that sound statistical comparisons cou'l not be drawn between such nonuniform sample sizes, duct height verification statistics were not determined for geographical subsets. Duct height verification results for temporal (day and night) categories are only available for the HRD product since the HEPC climatological information required for duct height determination is not available in separate day and night categories. Table 1 gives the verification statistics for the HRD analysis and 36 hr prognosis based on day and night data sets. For all three verification rates, and for both temporal subsets, the HRD analysis is observed to be better than the prognosis. Additionally, all verification statistics indicate that the HRD product (both the analysis and the prognosis) is a better nighttime forecaster of duct height than during the day. As a case in point, the number of HRD 36 hr duct height forecasts which overlap the observed is only 18 of 70 for the "day only" data set, but 46 of 84 for the nighttime.

Table 1. Duct height verification rates for the HRD analysis and 36 hr prognosis, based on day and night categories.

PREDICTOR	DUCT HEIGHT VERIFICATION RATE		
	(1) Bell Curve	(2) Within 1000 ft	(3) Overlaps
HRD ANALYSIS			
DAY	88.0%	53/68 77.9%	38/68 55.9%
NIGHT	92.0%	83/96 86.5%	61/96 63.5%
HRD 36 HR PROG			
DAY	83.9%	47/70 67.1%	18/70 25.7%
NIGHT	86.3%	67/84 79.8%	46/84 54.8%

4. FURTHER CONSIDERATIONS

As formulated, the NWOC HRD product relies heavily on the ability to infer refractivity conditions from synoptic parameters and satellite imagery. A careful examination by this author of the HRD forecasting thumbrules (as given in NWOC, 1991) indicates that the inferences drawn between synoptic features and cloud patterns, and refractivity structure, are based on sound meteorological reasoning. Statistical results of the previous section indicate that, over a 12 month period and under many varying operational circumstances, the NWOC HRD exhibited considerable skill ($V \sim 0.30$) in refractivity assessment and forecasting within the North Pacific. The achievement of this skill level strongly suggests that the forecasting thumbrules and guidelines used in the production of the HRD chart are both firmly based and scientifically accurate.

The forecasting thumbrules developed by PMTC and utilized in the NWOC HRD product are largely based on statistical studies over the subtropical and lower middle latitudes of the eastern Pacific Ocean. Thus, it is reasonable to ask, how complete are these thumbrules, and can they be validly applied in other ocean areas besides the eastern North Pacific? In the previous section, the prefigurance and postagreement indices, and the Hanssen and Kuipers skill score, for the HRD 36 hr prognosis (full data set) were all found to be considerably less over the central Pacific than over the eastern Pacific. Additionally, the sudden drop in the prefigurance and percent correct indices for the months of February and March (see Fig. 7) was attributed to a considerable number of incorrect ducting forecasts in areas (North Pacific > 45°N, tropics < 20°N) not previously well sampled. Taken together, these results suggest that the NWOC HRD thumbrules may not be as accurate or as valid in regions outside their developmental base (viz., the subtropical and lower middle latitudes of the eastern Pacific). To realize maximum predictive value, the NWOC HRD thumbrules would have to be tailored (i.e., modified) for use in those areas of interest which are climatically distinct from the eastern Pacific (for example, the tropics, the polar regions, the Persian Gulf/Arabian Sea) . For marginal seas and gulfs, specific inference thumbrules could be developed to relate well-known, localized mesoscale phenomena to refractive conditions. In their present form, the NWOC

HRD forecasting thumbrules are likely to be fully valid for those worldwide regions climatically very similar to the eastern Pacific (viz., the subtropical and middle latitudes of the southeast Pacific, the northeast and southeast Atlantic, and the southeast Indian Ocean).

At the present time, work is underway at NRL Monterey to incorporate the various synoptic-satellite inference thumbrules for refractivity forecasting into an AI (artificial intelligence) expert system suitable for implementation on the TESS. The use of such an automated aid at NWOC would streamline the production of the operational HRD chart. In addition to probable time and manpower savings (the present manual production cycle requires several meteorological personnel and about a half hour), the use of automated inference rules would provide a more objective and consistent operational HRD product by removing biases of individual analysts and forecasters.

5. SUMMARY AND CONCLUSIONS

This technical note serves as an independent validation of the NWOC HRD product, in particular, of its capabilities in the assessment and short-range forecasting of lower tropospheric ducting. For verification, direct comparisons of refractive structure were made among HRD analyses and 36 hr prognoses, and available North Pacific shipboard upper air observations, over a 12 month period extending from April 1991 through March 1992. Various statistical indices (including the Hanssen and Kuipers skill score), derived from forecast/observed contingency tables, were used to assess the duct/no duct forecast capabilities of the HRD product. The accuracy of HRD duct height forecasts were evaluated using three separate verification schemes. In order to assess the degree of HRD forecasting skill and utility, direct comparisons were made between the HRD product and an EM propagation conditions climatology. For these comparisons, two distinct climatological predictors (CLIM1 and CLIM2) were utilized. CLIM1, defined as the percent occurrence of the dominant duct type (surface-based or elevated), was evaluated over a range of decision (threshold) probabilities from $p^* = 30\%$ to 50% ; CLIM2, which corresponds to the percent occurrence of any ducting (surface-based and elevated combined), was evaluated at threshold probabilities of 50% to 80%.

For the full April 1991 - March 1992 data set, individual performance statistics (i.e., prefigurance, postagreement, percent correct and false alarm rate) were quite similar for the HRD analysis and the 36 hr prognosis. The capability of the NWOC HRD product in correctly forecasting ducting and nonducting events over the North Pacific basin was about 2/3; HRD duct forecasts were reliable about 4 out of 5 times. The Hanssen and Kuipers skill score (the difference between the hit and false alarm rates) was near 0.30 for both the HRD analysis and 36 hr prognosis; this value is quite respectable, being considerably above the demarcation between "skill" and "no skill" (i.e., V = 0.0).

Except for the postagreement index, the full year performance statistics for the climatology predictors exhibited marked variability as a function of threshold probability; small (large) hit and false alarm rates were associated with high (low) values of p*. Based on the Hanssen and Kuipers discriminant, the "optimum" forecast skill for the climatological assessment of ducting occurred at threshold probabilities of p* = 30% for CLIM1 and p* = 60% for CLIM2 (V ~ 0.12 in both cases). Although both climatology predictors had better duct forecasting capabilities (i.e., higher PF indices) than the HRD product at lower threshold probabilities, and quite respectable duct forecast reliabilities (PA indices near 0.70), their false alarm rates were roughly twice as large. For the full 12 month data set, differences in skill scores between the HRD product (both the analysis and 36 hr prognosis) and the climatology predictors CLIM1 and CLIM2 were determined to be significant at a 95% level of confidence.

Results based on seasonal data sets indicate slightly better prefigurance and percent correct indices during summer (April - September), and slightly better postagreement and false alarm rate indices during winter (October - March), for all HRD and climatology predictors. Skill scores for climatology were generally better in winter than in summer; a quite respectable value V = 0.25 was attained during winter for CLIM1 at the threshold probability p* = 30%. In general, skill score differences between the HRD product and climatology were significant (at a 0.95 level of confidence) in summer, but not in winter.

Based on geographical divisions, results indicate that both HRD duct forecasting capability and reliability were considerable better in the eastern Pacific (EPAC) than in the central Pacific (CPAC). For the HRD 36 hr prognosis, the CPAC prefigurance index was only 0.51 and the Hanssen and Kuipers skill score only 0.20; analogous values for EPAC were PF = 0.68 and V = 0.30. At all threshold probabilities, the reliability of climatological forecasts of ducting (given by PA values) was appreciably better for EPAC than for the CPAC region. In general, skill score differences between the HRD product and climatology were not significant for the EPAC region, and were not significant between the HRD 36 hr prognosis and climatology in the central Pacific.

Results based on temporal (day and night) classifications indicate that, with the exception of the false alarm rate, statistical indices (including the Hanssen and Kuipers discriminant) were better during the nighttime for both the HRD product and climatology. Interestingly, all daytime statistical indices were virtually the same for the HRD analysis and 36 hr prognosis. Skill score differences between the HRD 36 hr prognosis and climatology (CLIM1) were not significant for the night data set; on the other hand, differences between these two predictors were significant (at the 95% confidence level) for the "day only" data set except at the CLIM1 threshold probability $p^* = 30\%$.

Full year results based on a relatively low number of surface duct observations suggest that the capability of the HRD product in correctly forecasting surface duct occurrences is limited; on the other hand, when issued, HRD forecasts of surface ducting appear to be very reliable. Surface ducting performance statistics for climatology (CLIM1) show prefigurance values comparable to those determined for the HRD analysis (viz., quite low) at lower threshold probabilities ($p^* = 30\%$ and 35%); on the other hand, postagreement values are much lower than those for the HRD product.

Based on three separate verification schemes, the HRD analysis was found to be a better predictor of duct height than the HRD 36 hr prognosis. Climatological assessments of duct height compare quite favorably with those from the HRD 36 hr prognosis; for example, over the intermediate CLIM2 threshold probabilities ($p^* = 60\%$ to 70%), two of the three duct height verification rates derived from climatology were slightly higher than those for the HRD prognosis. Seasonal verification rates indicate that 1), the HRD analysis was a considerably better predictor of duct height during the winter than the HRD 36 hr prognosis and 2), climatology was a considerably better duct height indicator in summer than in winter.

Overall statistical results from this study indicate that the NWOC HRD product is a useful forecasting tool. Its capability in forecasting the occurrence of ducting is significantly better than that using climatology. Specifically, the factor which most distinguishes the HRD product from climatology is its ability to forecast a high percentage of ducting events without a large number of false alarms.

The forecasting thumbrules used by the NWOC HRD are largely based on inferences drawn between synoptic and cloud patterns, and refractive structure, over the subtropical and lower middle latitudes of the eastern North Pacific. Any potential use of the NWOC HRD product in regions climatically distinct from the eastern Pacific (e.g., the tropics, the polar regions, the Persian Gulf/Arabian Sea) would likely require some modification of existing forecast thumbrules in order to assure maximum predictive value of the product. The automation of the present HRD production cycle using an expert system based on synoptic-satellite inference thumbrules would likely result in a more efficient and objective refractivity product.

REFERENCES

Hanssen, A.W. and W.J.A. Kuipers, 1965: On the Relationship Between the Frequency of Rain and Various Meteorological Parameters, Koninklijk Nederlands Meteorologisch Instituut, Meded. Verhand., 81, 2-15.

Helvey, R.A. and J. Rosenthal, 1983: Guide for Inferring Refractive Conditions from Synoptic Parameters, PMTC Technical Publication TP 5.

Lyons, S.W., 1985: Satellite-derived Refractive Duct Height Estimates, PMTC Geophysical Sciences Technical Note No. 98.

NAVWESTOCEANCEN, 1991: Horizontal Refractivity Depiction Project Development, NWOC Document (ltr) Ser 3140/01123, October 3, 1991.

Patterson, W.L., 1987: Historical Electromagnetic Propagation Condition Database Description, NOSC Technical Document 1149.

Rosenthal, J., S. Westerman and R. Helvey, 1985: Inferring Refractivity Conditions From Satellite Imagery, PMTC Geophysical Sciences Technical Note No. 96.

Rudloff, W., 1981: World Climates. Wissenschaftliche Verlagsgesellschaft mbH, Stuttgart, Germany.

Woodcock, F., 1976: The Evaluation of Yes/No Forecasts for Scientific and Administrative Purposes, Mon. Wea. Rev. 104 (10), 1209-1214.

DISTRIBUTION LIST

CHIEF OF NAVAL OPERATION
ATTN OP-096 OP 0961B
US NAVAL OBSERVATORY
WASHINGTON DC 20392-1800

SPAWARSYSCOM
ATTN PMW 165
WASHINGTON DC 20363-5100

NAVAL RESEARCH LAB 10
ATTN CODE 125L
JCSSC MS 39529-5004

NAVAL RESEARCH LAB
ATTN CODE 125P
JCSSC MS 39529-5004

COMNAVOCEANCOM
ATTN N312
JCSSC MS 39529-5000

COMFLENUMOCEANCEN
ATTN OPERATIONS OFFICER
MONTEREY CA 93943-5005

CINCLANTFLT
ATTN FLT METEOROLOGIST
NORFOLK VA 23511-5210

CINCPACFLT
ATTN CODE 02M
PEARL HARBOR HI 96860-7000

CINCUSNAVEUR
BOX N39
FPO AE 09510-0150

COMSECONDFLT
ATTN METEORO OFFICER
FPO AE 09501-6000

COMTHIRDFLT
ATTN FLT METEOROLOGIST
FPO AP 96601-6001

COMSEVENTHFLT
ATTN FLT METEOROLOGIST
FPO AP 96601-6003

COMXITHFLT
ATTN FLT METEOROLOGIST
FPO AE 09501-6002

J-3 ESD JOINT STAFF
ATTN OPS DIRECTORATE
WASHINGTON DC 20318-3000

COMNAVOCEANCOM
ATTN CODE N5
JCSSC MS 39529-5000

NAVWESTOCEANCEN
ATTN OPS OFFICER
BOX 113
PEARL HARBOR HI 96860

NAVEASTOCEANCEN
ATTN OPS OFFICER
MCADIE BLDG U-117
NORFOLK VA 23511-5399

NAVPOLAROCEANCEN
ATTN OPS OFFICER
4301 SUITLAND RD
WASHINGTON DC 20395-5180

US NAVOCEANCEN
ATTN OPS OFFICER
BOX 12 COMNAVMARIANAS
FPO AP 96630-2926

US NAVOCEANCOMCEN
ATTN OPS OFFICER
BOX 31
FPO AE 09540-3200

NAVPGSCOL
ATTN CODE MR
MONTEREY CA 93943-5000

NAVAIRSYSCOM
ATTN CODE 526W
WASHINGTON DC 20361-0001

NAVOCEANSYSCEN
ATTN J RICHTER CODE 54
SAN DIEGO CA 92152-5000

PACMISTESTCEN
ATTN GEOPHYS OFFICER
PT MUGU CA 93042

REPORT DOCUMENTATION PAGE

Form Approved
OBM No. 0704-0188

1. Agency Use Only *(Leave blank).*

2. Report Date. July 1992

3. Report Type and Dates Covered. Final

4. Title and Subtitle.
A Horizontal Refractivity Depiction Product: An Evaluation

5. Funding Numbers.
Program Element No. O&M,N
Project No. --
Task No. --
Accession No. DN252053
Work Unit No. O&M,N-2

6. Author(s).
Gerard N. Vogel

7. Performing Organization Name(s) and Address(es).
Naval Oceanographic and Atmospheric Research Laboratory
Atmospheric Directorate
Monterey, CA 93943-5006

8. Performing Organization Report Number.
NOARL Technical Note 284

9. Sponsoring/Monitoring Agency Name(s) and Address(es).
Commander, Naval Oceanography Command
Stennis Space Center, MS 39529-5000

10. Sponsoring/Monitoring Agency Report Number.
NOARL Technical Note 284

11. Supplementary Notes.

12a. Distribution/Availability Statement.
Approved for public release; distribution is unlimited.

12b. Distribution Code.

13. Abstract *(Maximum 200 words).*

The results from an independent validation of the Naval Western Oceanography Center (NWOC) Horizontal Refractivity Depiction (HRD) product, based on a 12 month North Pacific data set, are presented. The average capability of the HRD product (analysis and 36 hr prognosis) in correctly assessing observed refractive structure (either ducting or nonducting) was near 65%; HRD forecasts were reliable 79% of the time. Based on the Hanssen and Kuipers skill score (the difference between the hit and false alarm rates), the capabilities of the HRD product in the assessment and short-range forecasting of ducting were determined to be statistically better than those using an electromagnetic propagation conditions climatology. Any potential operational use of the NWOC HRD product in regions climatically distinct from the eastern North Pacific would likely require some modification of existing forecast thumbrules. The implementation of an automated synoptic-satellite inference rule base (as per an expert system) for the HRD, which would likely result in a more objective and consistent refractivity product, is recommended.

14. Subject Terms.
Refractivity
Ducting
Electromagnetic propagation

15. Number of Pages. 48

16. Price Code.

17. Security Classification of Report. UNCLASSIFIED

18. Security Classification of This Page. UNCLASSIFIED

19. Security Classification of Abstract. UNCLASSIFIED

20. Limitation of Abstract. Same as report

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102