NOAA Technical Memorandum
NWS FCST 33

NWS VERIFICATION OF PRECIPITATION TYPE AND SNOW AMOUNT FORECASTS DURING THE AFOS ERA

National Weather Service
Silver Spring, MD
January, 1990

U.S. DEPARTMENT OF COMMERCE / National Oceanic and Atmospheic Administration / National Weather Service

NOAA TECHNICAL MEMORANDUMS

National Weather Service, Program Requirements and Development Division Series

The Program Requirements and Development Division of the Office of Meteorology is specifically responsible for preparation of program requirements and plans relating to forecasting services for the public, aviation, marine, agriculture, and fire weather interests. NOAA Technical Memorandums in the Program Requirements and Development Division series communicate scientific and technical information relating to field forecasting operations. The series includes information on present techniques, procedures, and performance data. Background information and detail on selected service operations are also given. The series provides a means for the personnel in the National Weather Service Headquarters and Regional Offices to report on forecasting methods of general interest and wide application.

NOAA Technical Memorandums in the Program Requirements and Development Division series facilitate rapid distribution of material which may be preliminary in nature and which may be published formally elsewhere at a later date. Publications 1 through 4 by this Division are in the former series, Weather Bureau Technical Notes (TN), Notes to Forecasters (FCST); publications 5 through 15 are in the former series, ESSA Technical Memorandums Weather Bureau Technical Memorandums (WBTM). Beginning with FCST 16, publications are now part of the series, NOAA Technical Memorandums, National Weather Service (NWS).

Publications listed below are available from the National Technical Information Service (NTIS), U.S. Department of Commerce, Sills Bldg., 5285 Port Royal Road, Springfield, VA 22161. Prices vary for paper copy. Microfiche copies available. Order by accession number, when given, in parentheses.

Weather Bureau Technical Notes

TN 8 FCST 1 On the Use of Probability Statements in Weather Forecasts. Charles F. Roberts, September 1965. (PB-174-647)

TN 13 FCST 2 Local Cloud and Precipitation Forecast Method (SLYH). Matthew H. Kulawiec, September 1965. (PB-168-610)

TN 16 FCST 3 Present and Future Operational Numerical Prediction Models. Charles F. Roberts, October 1965. (PB-169-126)

TN 23 FCST 4 Forecasting the Freezing Level Objectively as an Aid in Forecasting the Level of Icing. Jack B. Cox, December 1965. (PB-169-247)

ESSA Technical Memorandums

WBTM FCST 5 Performance of the 6-Layer Baroclinic (Primitive Equation) Model. Julius Badner, September 1966. (PB-173-426)

WBTM FCST 6 Forecasting Mountain Waves. Phillip A. Calabrese, September 1966. (PB-174-648)

WBTM FCST 7 A Method for Deriving Prediction of Soil Temperature from Medium-Range Weather Forecasts. Charles F. Roberts, June 1967. (PB-175-773)

WBTM FCST 8 Recent Trends in the Accuracy and Quality of Weather Bureau Forecasting Service. Charles F. Roberts and John M. Porter, November 1967. (PB-176-953)

WBTM FCST 9 Report on the Forecast Performance of Selected Weather Bureau Offices for 1966-1967. C. F. Roberts, J. M. Porter, and G. F. Cobb, December 1967. (PB-177-043)

WBTM FCST 10 Size of Tornado Warning Area When Issued on Basis of Radar Hook Echo. Alexander F. Sadowski May 1969. (PB-184-613)

WBTM FCST 11 Report on Weather Bureau Forecast Performance 1967-68 and Comparison with Previous Years. Charles F. Roberts, John M. Porter, and Geraldine F. Cobb, March 1969. (PB-184-366)

WBTM FCST 12 Severe Local Storm Occurrences 1955-1967. Staff, SELS Unit, NSSFC, Maurice E. Pautz, Editor, September 1969. (PB-187-61)

WBTM FCST 13 On the Problem of Developing Weather Forecasting Equations by Statistical Methods. Charles F. Roberts, October 1969. (PB-187-796)

WBTM FCST 14 Preliminary Results of an Empirical Study of Some Spectral Characteristics of Skill in Present Weather and Circulation Forecasts. Charles F. Roberts, November 1969. (PB-188-529)

WBTM FCST 15 Weather Bureau Forecast Verification Scores 1968-69 and Some Performance Trends from 1966. Robert G. Derouin and Geraldine F. Cobb, May 1970. (PB-192-949)

(Continued on inside back cover)

NOAA Technical Memorandum
NWS FCST 33

NWS VERIFICATION OF PRECIPITATION TYPE AND SNOW AMOUNT FORECASTS DURING THE AFOS ERA

Barry S. Goldsmith
Program Requirements and Development Division
Office of Meteorology

National Weather Service
Silver Spring, MD
January, 1990

UNITED STATES
DEPARTMENT OF COMMERCE
Robert A. Mosbacher
Secreteary

National Oceanic and
Atmospheric Administration
John A. Knauss
Under Secretary

National Weather Service
Albert W. Friday, Jr.
Assistant Administrator

TABLE OF CONTENTS

ABSTRACT

A study was conducted to analyze verification techniques and results of National Weather Service (NWS) subjective and objective forecasts of precipitation type and snow amount. The objective forecast system is based on the model output statistics approach developed by the Techniques Development Laboratory of the NWS. Results and conclusions are based on the data collected by the Automation of Field Operations and Services era verification (AEV) system, which was introduced in October, 1983. Six cool seasons (1983/84 - 1988/89) were included in the period of study.

Precipitation type forecasts are stratified into three categories: freezing, frozen, and liquid precipitation. Two sets of observations are used for verification purposes: a verifying time set, which is valid for a specific projection; and a window set, which is chosen from a compilation of precipitation type observations for one hour on either side of the verifying time. The number of cases in the window set exceed the number of cases in the verifying-time set. A difference in the definition of liquid precipitation between subjective and objective forecasts is discussed. Verification results of all 18-, 30-, and 42-hour projections for national and regional precipitation type forecasts are presented. Comparisons and contrasts of subjective and objective forecasts are based on various statistical scores such as the critical success index (CSI) and the Heidke skill score (skill).

For national and regional data, CSI scores were much higher for frozen precipitation than for freezing precipitation. Similar results were noted for subjective and objective forecasts. With the exception of the Southern Region, there were few differences in scores among regions, whose results reflected those of the nation.

Snow amount forecasts are broken into four categories: 0 (0-1 inch), 2 (2-3 inches), 4 (4-5 inches), and 6 (greater than or equal to 6 inches). Verification results are provided for the complete set of national and regional 24 hour forecasts. In addition to comparisons of subjective and objective forecasts, fluctuations in the scores resulting from both variation in snowstorm frequency and model reliability between seasons are discussed, as is subjective forecast improvement over objective forecasts for category 6.

Snow amount forecast results for category 0 were omitted from final analyses. In general, CSI scores were highest for category 6 and lowest for category 4. Subjective forecasts were better than objective forecasts for category 6, but both were similar for categories 2 and 4. For the most part, regional scores reflected national scores.

Appendices are provided which describe the AEV system, the MOS precipitation type and snow amount forecast systems, and the definitions of the appropriate verification scores used in the study.

1. INTRODUCTION

Since 1983, the National Weather Service (NWS) has monitored the quality of public and aviation weather forecasts through use of the Automation of Field Operations and Services (AFOS) system. Prior to that time the verification effort was dependent on paper forms. The AFOS-era verification (AEV) program (Ruth and Alex, 1987) collects aviation and public forecasts and collates them with surface and synoptic observations at each Weather Service Forecast Office (WSFO) to form verification matrices (see Appendix A for details). These matrices are automatically translated into a coded message at the WSFO and, after undergoing quality control, transmitted to the National Meteorological Center (NMC) where they are synthesized into a national verification archive (Dagostaro, 1985). Data for 94 offices in the contiguous United States - 47 WSFO's paired with one Weather Service Office (WSO) for which it has forecast responsibility - are included in the national AEV database. The data are stratified by 6-month warm and cool seasons. At the completion of each season, the archived data are accessed by the NWS Techniques Development Laboratory (TDL) and Office of Meteorology (OM), where verification statistics are computed. National and/or regional summaries are then prepared for further review.

Local forecasts (i.e., forecasts prepared by NWS meteorologists at WSFO's), Model Output Statistics (MOS) guidance forecasts (Glahn and Lowry, 1972), and corresponding observations are archived for public and aviation weather elements. Public elements include maximum and minimum temperature, probability of precipitation (PoP), cloud amount, 42-h significant wind speed (greater than or equal to 22 kt), and, for the cool season only, precipitation type and snow amount. Aviation elements include wind direction and speed, visibility, and ceiling height. In past documentation (e.g., Polger and Thompson, 1985), attention has been focused on temperature and PoP verification. Precipitation type and snow amount statistics, on the other hand, have been studied to a much lesser degree. This is partly because some of the individual cool seasons had insufficient data to provide a significant set of statistics. This paper will review verification techniques and results for both local and MOS forecasts of precipitation type and snow amount. Only 86 of the 94 AEV stations were used in this study since MOS forecasts of precipitation type and snow amount were not developed for certain warm weather sites. The results, which are presented from a national perspective, are based on data available since the implementation of the AEV software. In particular, the 1983/84 through 1988/89 cool seasons will be examined.

2. A STUDY OF THE PRECIPITATION TYPE FORECAST SYSTEM

2.1. Data Archival Procedures

Precipitation type forecasts and observations are recorded categorically. They are assigned a value of 1, 2, or 3, for freezing (Z), frozen (S), or liquid (R) precipitation, respectively. Values for the precipitation type forecasts are always specified, regardless of whether precipitation is mentioned in the forecast. If the forecaster expects mixed precipitation, then the worst-hazard category is recorded (National Weather Service, 1985). For instance, if a worded forecast reads "rain, mixed with freezing rain and sleet, likely today," the value entered must be Z. Meanwhile, the MOS forecast uses a threshold decision tree (National Weather Service, 1982a) which chooses its best category (see Appendix B for details).

Two sets of observations are archived in the verification process. One set is valid at the verifying time and is referred to as the "verifying-time" set in this study. The other set is chosen from a compilation of precipitation types for a 2-hour window about the verifying time. This set will be referred to as the "window set" in this study and includes all special observations. The verifying times are defined as 18-, 30-, and 42-hour projections from an initial time of either 0000 UTC or 1200 UTC. Although verifying times for local and MOS forecasts are the same, the release times differ. The MOS guidance is available to the forecaster approximately 3 hours after the initial time, and the local forecast is released approximately 9 hours after the initial time. For example, an 18-hour MOS precipitation type forecast generated from the 0000 UTC output valid at 1800 UTC corresponds to a 9-hour forecast issued by the local forecaster. Similar to the forecasts, the category of the verifying observation is determined by a hierarchy of hazard potential. The observed precipitation type which occurs in the category with the lowest value is always used as the verifying observation. For example, if Z and R are reported within the two hour window <u>or</u> a mixture of Z and R is noted at the verifying time, Z becomes the verifying observation.

It should be noted that not all local and MOS precipitation type forecasts are included in the verification archive. Specifically, only cases for which the local PoP forecast is greater than or equal to 30% <u>and</u> precipitation is observed are included. The 30% threshold was chosen because it is commonly used as the minimum value in forecasts that mention the possibility of measurable precipitation. The number of cases eliminated by this limitation accounted for 19% of all available verifying-time set data and 21% of all available window set data for the sample. The percent of cases eliminated differs because, for the window set, the hourly observations on either side of the verifying time plus any special observations within the window are considered. Therefore, for the window set, there is a greater probability of meeting the precipitation occurrence criterion. An observation of precipitation is recorded as an occurrence whether or not it is measurable.

2.2. <u>A Verification Problem</u>

A discrepancy shows up between the local and MOS categorical selection processes for potential mixed liquid/frozen events. In some cases, a local forecast of S may be countered by a MOS forecast of R. In MOS development (see Appendix B), mixed liquid/frozen precipitation type observations were included in the R category. Thus, not all MOS R forecasts were derived from liquid precipitation observations only; some may actually have been derived from mixed liquid/frozen precipitation observations. In theory, a MOS forecast of R and a local forecast of S could both have been correct for a case where mixed liquid and frozen precipitation occurred. Therefore, an observation of liquid/frozen precipitation that correctly verifies the local forecast of S will deem the MOS forecast of R as unsuccessful. Unfortunately, there was no practical way to discern the origin of a particular MOS R forecast from the data used in this study. Thus, no further analyses were undertaken for these obscure forecast cases.

2.3. Discussion of Results

A general analysis of local and MOS results common to both the window and verifying-time sets will be presented, followed by examination of the two sets separately. Tables 1a - 1d show consolidated contingency tables and computed scores (see Appendix D) for the local and MOS verifying-time sets and window sets based on national data. The consolidated tables consist of merged 18-, 30-, and 42-hour forecasts from the combined 0000 UTC and 1200 UTC cycles. Statistics for R will not be addressed since the calculated results were of little interest. Also, concern for cool season precipitation type lies primarily with the Z and S categories.

The most evident aspect of the data was revealed by differences in the POD, FAR, and CSI between Z and S, regardless of forecast type or forecast set. The reason for this difference is simple: Z is a relatively rare event that is difficult to forecast, as shown by reviewing any one of the four contingency tables. For Z, the number of incorrect forecasts and unforecast events each existed on the same order of magnitude as the number of correct forecasts ("hits"). Thus POD, FAR, and CSI values indicate a low level of accuracy. By contrast, S forecasting is facilitated by its high rate of occurrence. The number of hits was somewhat larger than the number of incorrect forecasts and unforecast events combined. Therefore, POD, FAR, and CSI values reflect a high level of accuracy. Fig. 1 shows the result.

When CSI was calculated for Z forecasts, the resulting value was almost always under 0.5. From Table 1c (local 2-hour window set), there were 568 hits, 969 incorrect forecasts, and 1586 unforecast events. This gave a CSI of 0.18. Additional analysis of the local and MOS data lends further insight into why the CSI scores were inherently low for each system. It follows from the definition of CSI (Appendix D) that a high POD (few unforecast events) combined with a low FAR (few incorrect forecasts) produces a high CSI; conversely, a low POD (many unforecast events) combined with a high FAR (many incorrect forecasts) produces a low CSI. The MOS selection process creates a relatively high number of freezing precipitation forecasts and, although this raises the POD, it consequently raises the FAR. Local forecasters, perhaps displaying some resiliency to the guidance, produce fewer numbers of Z forecasts. Although this lowers the FAR, it subsequently lowers the POD. With no noticeable differences in CSI between the two forecast systems, it can be concluded that, in general, lower values of local FAR and POD existed simultaneously with higher values of MOS FAR and POD.

Table 1. Contingency tables and computed scores of local and MOS precipitation type forecasts for the national verifying-time and window sets.

(a) LOCAL - VERIFYING TIME

OBSERVED	FORECAST Z	S	R	TOTAL
Z	355	498	390	1243
S	511	18132	1948	20591
R	369	1693	25900	27962
TOTAL	1235	20323	28238	49796

<u>COMPUTED SCORES</u>:

1. BIAS Z = 0.99, S = 0.99, R = 1.01
2. POD Z = 0.29, S = 0.88
3. FAR Z = 0.71, S = 0.11
4. CSI Z = 0.17, S = 0.80
5. SKILL= 0.788
6. PFC= 0.89
7. % IMP. OVER MOS (SKILL)= -0.4

(b) MOS - VERIFYING TIME

OBSERVED	FORECAST Z	S	R	TOTAL
Z	525	350	368	1243
S	906	18022	1663	20591
R	758	1381	25823	27962
TOTAL	2189	19753	27854	49796

<u>COMPUTED SCORES</u>:

1. BIAS Z = 1.76, S = 0.96, R = 1.00
2. POD Z = 0.42, S = 0.88
3. FAR Z = 0.76, S = 0.09
4. CSI Z = 0.18, S = 0.81
5. SKILL= 0.791
6. PFC= 0.89

(c) LOCAL - WINDOW

OBSERVED	FORECAST Z	S	R	TOTAL
Z	568	964	622	2154
S	576	22545	2971	26092
R	393	2012	37011	39416
TOTAL	1537	25521	40604	67662

<u>COMPUTED SCORES</u>:

1. BIAS Z = 0.71, S = 0.98, R = 1.03
2. POD Z = 0.26, S = 0.86
3. FAR Z = 0.63, S = 0.12
4. CSI Z = 0.18, S = 0.78
5. SKILL= 0.799
6. PFC= 0.89
7. % IMP. OVER MOS (SKILL)= -0.9

(d) MOS - WINDOW

OBSERVED	FORECAST Z	S	R	TOTAL
Z	816	712	626	2154
S	1039	22505	2548	26092
R	856	1618	36942	39416
TOTAL	2711	24835	40116	67662

<u>COMPUTED SCORES</u>:

1. BIAS Z = 1.26, S = 0.95, R = 1.02
2. POD Z = 0.38, S = 0.90
3. FAR Z = 0.70, S = 0.09
4. CSI Z = 0.20, S = 0.79
5. SKILL= 0.786
6. PFC= 0.89

For S, a more common event in the cool season, the number of hits was one to two orders of magnitude <u>greater</u> than the <u>combination</u> of incorrect forecasts and unforecast events for the window set. As shown in Table 1c, 22545 hits were recorded, compared to 2976 incorrect forecasts and 3547 unforecast events. Thus, the hits value became the dominant term. Therefore, the CSI computed was 0.78. Unlike the Z forecasts, specific analyses of the local and MOS systems for S showed no significant differences in their respective POD and FAR values.

Differences between local and MOS Z scores were readily apparent for both verifying-time and window set data. Total Z forecasts issued by MOS outnumbered local Z forecasts nearly 2 to 1. This accounted for an increase in the MOS POD, FAR, and Bias (Tables 1b and 1d) over the local POD, FAR, and Bias (Tables 1a and 1c). Local forecasters appeared to be conservative when specifying Z. Normally, freezing precipitation is mentioned with other precipitation types in a worded forecast. However, from the apparent dearth of Z cases in the sample, it could be argued that worded forecasts of mixed precipitation were seldom used. The objective MOS system produced Z forecasts much more liberally. This is explained by MOS probability forecasts equalling or exceeding the critical Z thresholds at a relatively high rate. Important differences in S scores were not readily apparent from the data.

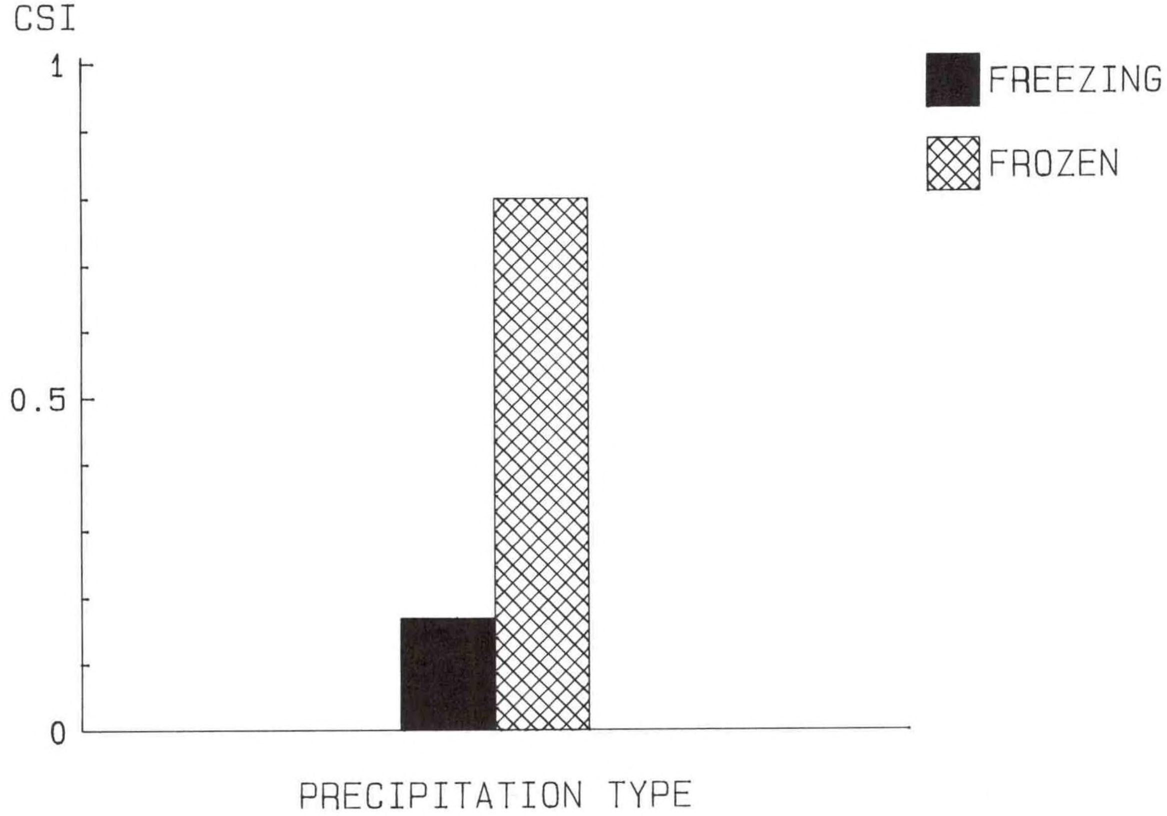

Figure 1. CSI for freezing (Z) and frozen (S) local precipitation type forecasts for the consolidated 6-year AEV data set (1983/84 - 1988/89). The values shown are based on the verifying-time set.

Tables 1a and 1c show similar skill scores between local and MOS forecasts. For both the verifiying-time and the window sets, MOS skill held a slight edge over local skill, although there was little discernable difference between them. Regardless of the comparisons, both forecast systems displayed a high degree of skill on an absolute scale.

A comparison of the window and verifying-time sets gives more interesting results. Most notably, the total number of verified forecasts was much larger for the window set (Tables 1c and 1d) than for the verifying-time set (Tables 1a and 1b). More potential verifying observations in the 2-hour window allowed for an increased likelihood of meeting the observed precipitation criterion (regardless of category); hence the increase in number. This difference was magnified for the Z category (again due to relative rareness of the event). For local and MOS forecasts, bias was reduced considerably in the window set. Changes in POD and FAR were more subtle. Decreases were noted in both forecast systems from the verifying-time set to the window set. For POD, the decrease was due to the near doubling of the number of observations relative to the smaller increase in the number of hits; for FAR, the decrease was due to the smaller increase in the number of forecasts relative to the increase in the number of hits. The larger decrease in FAR between verifying-time and window sets resulted in increased CSI for the latter.

Given the aforementioned results it would appear that scoring precipitation type forecasts by a window produces better results. However, the window can be as much a burden as it is a boon. This is simply because for a more common element the probability of error increases. For example, a forecast of S would not verify if the precipitation changed over to any form of Z at any time during the window. Perhaps this is why window set CSI scores were slightly <u>worse</u> than the associated verifying-time set scores. Similarly, it is likely that lower window scores existed for the R category (especially since there are two more hazardous categories to verify), although statistics were unavailable for this case.

Tables 2a - 2h show contingency tables of local (a - d) and MOS (e - h) forecasts, broken down by region, for the 18-hour verifying time-set. The Southern and Western regions had the fewest total events. Freezing and frozen events are relatively infrequent in the Southern Region as compared with the Eastern and Central Regions. Compared to the other regions, the Western Region has fewer AEV stations and fewer cases of <u>any</u> precipitation. Further inspection revealed that the Central Region was the only region to have more S than R cases and the Western Region had virtually no Z cases (an average of 2.8 per season per station). Fig. 2 shows the CSI of local and MOS Z forecasts for the Eastern, Southern, and Central Regions, combined cycles and seasons, for the verifying-time set. Western Region data are not shown since occurrences of Z were rare. Relatively high scores for the Southern Region may be attributed to the meteorological conditions which produce freezing precipitation. During the cool season, it is not unusual to observe subfreezing low-level air masses (primarily within the boundary layer) overspreading areas of the Southern Region while mid-level temperatures remain

Table 2. Contingency tables and computed scores of local and MOS precipitation type forecasts, broken down by region, for the verifying-time set.

(a) EASTERN LOCAL

OBSERVED	FORECAST Z	FORECAST S	FORECAST R	TOTAL
Z	49	52	50	151
S	55	2399	206	2660
R	49	196	3284	3529
TOTAL	153	2647	3540	6340

COMPUTED SCORES:
1. BIAS Z = 1.01, S = 1.00, R = 1.00
2. POD Z = 0.32, S = 0.90
3. FAR Z = 0.68, S = 0.09
4. CSI Z = 0.19, S = 0.82
5. SKILL= 0.813
6. PFC= 0.90
7. % IMP. OVER MOS (SKILL)= 1.0

(e) EASTERN MOS

OBSERVED	FORECAST Z	FORECAST S	FORECAST R	TOTAL
Z	50	47	54	151
S	91	2364	205	2660
R	71	170	3288	3529
TOTAL	212	2581	3547	6340

COMPUTED SCORES:
1. BIAS Z = 1.40, S = 0.97, R = 1.01
2. POD Z = 0.33, S = 0.89
3. FAR Z = 0.81, S = 0.08
4. CSI Z = 0.16, S = 0.82
5. SKILL= 0.805
6. PFC= 0.89

(b) SOUTHERN LOCAL

OBSERVED	FORECAST Z	FORECAST S	FORECAST R	TOTAL
Z	31	18	25	74
S	24	194	79	297
R	26	23	2545	2594
TOTAL	81	235	2649	2965

COMPUTED SCORES:
1. BIAS Z = 1.09, S = 0.79, R = 1.02
2. POD Z = 0.42, S = 0.65
3. FAR Z = 0.62, S = 0.17
4. CSI Z = 0.25, S = 0.57
5. SKILL= 0.686
6. PFC= 0.93
7. % IMP. OVER MOS (SKILL)= -4.2

(f) SOUTHERN MOS

OBSERVED	FORECAST Z	FORECAST S	FORECAST R	TOTAL
Z	38	17	19	74
S	36	209	52	297
R	25	35	2534	2594
TOTAL	99	261	2605	2965

COMPUTED SCORES:
1. BIAS Z = 1.34, S = 0.88, R = 1.00
2. POD Z = 0.51, S = 0.70
3. FAR Z = 0.62, S = 0.20
4. CSI Z = 0.28, S = 0.60
5. SKILL= 0.716
6. PFC= 0.94

(c) CENTRAL LOCAL

OBSERVED	FORECAST Z	FORECAST S	FORECAST R	TOTAL
Z	77	68	34	179
S	88	2816	235	3139
R	45	140	2108	2293
TOTAL	210	3024	2377	5611

COMPUTED SCORES:
1. BIAS Z = 1.17, S = 0.96, R = 1.04
2. POD Z = 0.43, S = 0.90
3. FAR Z = 0.63, S = 0.07
4. CSI Z = 0.25, S = 0.84
5. SKILL= 0.772
6. PFC= 0.89
7. % IMP. OVER MOS (SKILL)= -5.3

(g) CENTRAL MOS

OBSERVED	FORECAST Z	FORECAST S	FORECAST R	TOTAL
Z	85	61	33	179
S	90	2878	171	3139
R	60	130	2103	2293
TOTAL	235	3069	2307	5611

COMPUTED SCORES:
1. BIAS Z = 1.31, S = 0.98, R = 1.01
2. POD Z = 0.47, S = 0.92
3. FAR Z = 0.64, S = 0.06
4. CSI Z = 0.26, S = 0.86
5. SKILL= 0.815
6. PFC= 0.90

(d) WESTERN LOCAL

OBSERVED	FORECAST Z	FORECAST S	FORECAST R	TOTAL
Z	2	11	4	17
S	9	969	112	1090
R	8	61	1197	1266
TOTAL	19	1041	1313	2373

COMPUTED SCORES:
1. BIAS Z = 1.12, S = 0.96, R = 1.04
2. POD Z = 0.12, S = 0.89
3. FAR Z = 0.89, S = 0.07
4. CSI Z = 0.06, S = 0.83
5. SKILL= 0.828
6. PFC= 0.91
7. % IMP. OVER MOS (SKILL)= -1.8

(h) WESTERN MOS

OBSERVED	FORECAST Z	FORECAST S	FORECAST R	TOTAL
Z	5	11	1	17
S	8	981	101	1090
R	3	62	1201	1266
TOTAL	16	1054	1303	2373

COMPUTED SCORES:
1. BIAS Z = 0.94, S = 0.97, R = 1.20
2. POD Z = 0.29, S = 0.90
3. FAR Z = 0.69, S = 0.07
4. CSI Z = 0.18, S = 0.84
5. SKILL= 0.844
6. PFC= 0.92

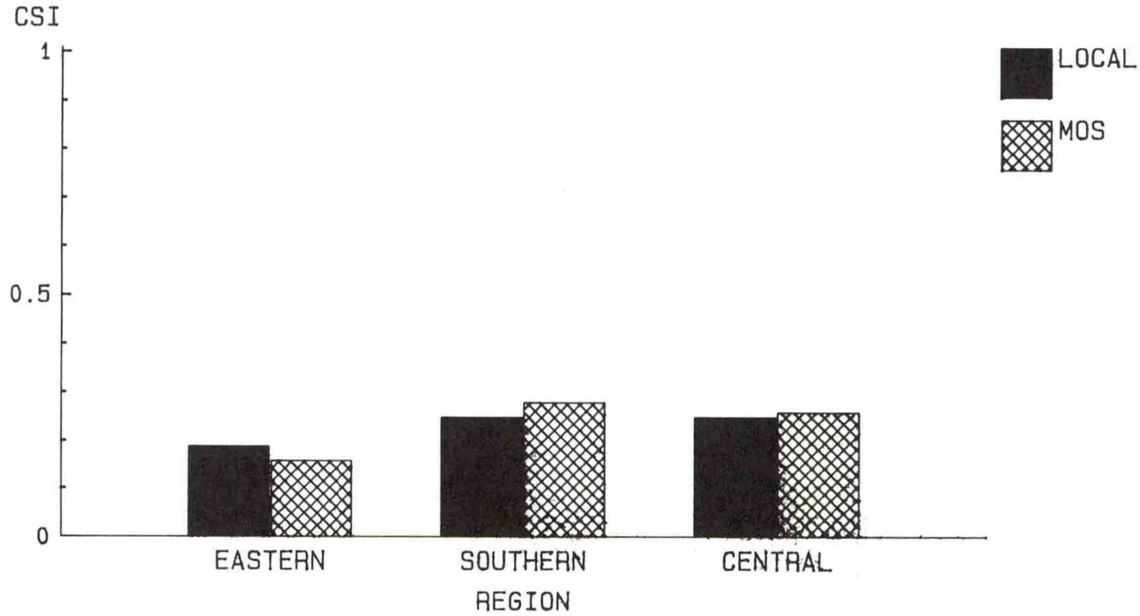

Figure 2. CSI for freezing (Z) local and MOS precipitation type forecasts, broken down by region, for the 6-year sample of verifying-time set data. Western Region data are not shown since the CSI is highly inadequate due to insufficient data.

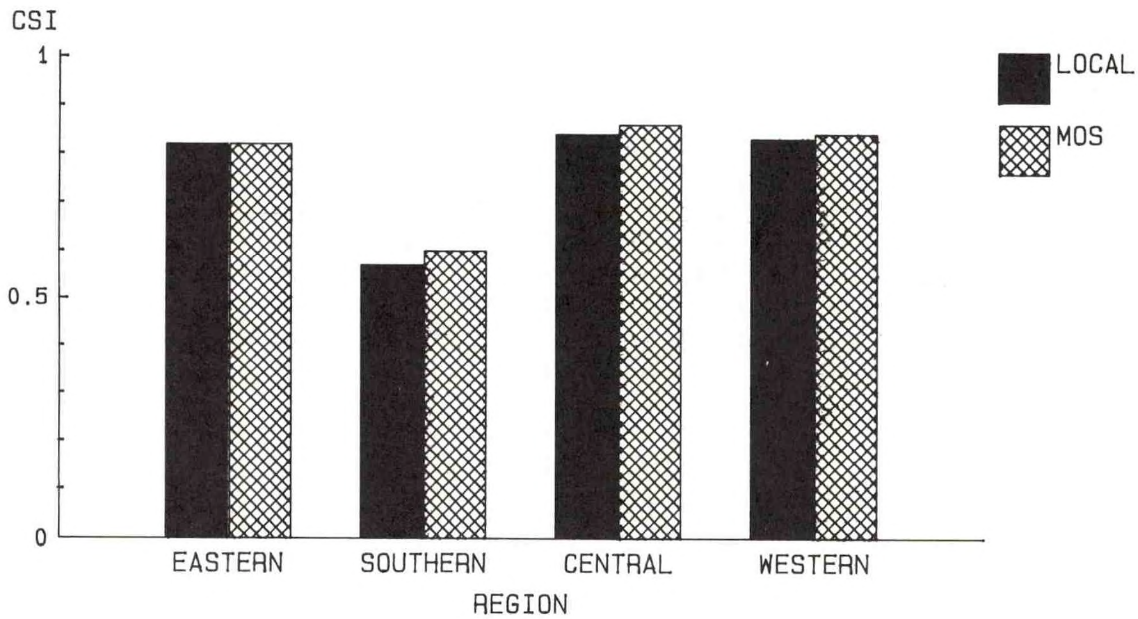

Figure 3. Same as Figure 2 except for frozen (S) precipitation type forecasts.

slightly above 0°C. Thus, occurrences of Z relative to S are much more frequent in the Southern Region than in the other regions.

Fig. 3 shows CSI scores for S for all regions including the Western Region. The Central Region was the only region to have more S events than any other precipitation type; the highest local and MOS CSI values were contained in the region, as well. In direct contrast to Fig. 2, the Southern Region had the lowest CSI. The degree of difficulty in forecasting S for the Southern Region is much higher compared to the other regions and is related to the argument presented above. For the window set, a similar configuration of scores was evident for both the Z and S categories.

With the exception of the Eastern Region, local forecasters showed slightly less skill than MOS in forecasting precipitation type. The skill differential was accentuated in the Southern and Central Regions, where MOS POD and CSI scores for the Z and S categories were noticeably better than the local scores. The increase in the POD and CSI was directly attributable to the increased number of hits, a variable common to POD, CSI, and skill scores.

The largeness of the database can effectively mask the scores for relatively rare situations. In these cases, the forecast success or failure hinges on whether the verifying-time or window verification system is applied. One example shown here describes how local forecasters can improve on MOS for the window set, given a specific synoptic setting. Later release should allow local forecasters more time to carefully analyze the necessary elements and parameters, and therefore more opportunity to pinpoint the precipitation type, especially when Z versus S demarcations are involved. A simple example shows how this, when combined with knowledge of model flaws, affects the window set scores. To illustrate the idea of continuous temporal thinking, consider the following synoptic situation. Suppose that strong cold air damming began east of the Appalachian mountains while a vigorous short wave approached from the west. Precipitation began at the station in the afternoon and continued through the following morning. The 1200 UTC MOS guidance forecast S at 12 hours (0000 UTC) but, because MOS fails to handle damming cases effectively, forecast R at 0600 UTC (18 hours). A local forecaster, having not only an advantageous release time lag but also a better sense of the evolving damming situation, forecasts Z at 0600 UTC. Assuming no special observations occurred within the window, what verified is illustrated as follows:

	0500 UTC	0600 UTC	0700 UTC
OBSERVATION	S	S	Z
MOS	--	R	--
LOCAL	--	Z	--

Although both local and MOS failed to hit at the verifying hour (0600 UTC), the local forecaster's insight produced a hit in the 2-hour window (0500 - 0700 UTC).

--THIS PAGE INTENTIONALLY LEFT BLANK--

3. A STUDY OF THE SNOW AMOUNT FORECAST SYSTEM

3.1. Data Archival Procedures

Local and MOS snow amount forecasts are verified categorically. Local forecasts are recorded in whole inch amounts. To keep consistency between forecast systems, the local forecasts are grouped into categories used by MOS before they are processed. MOS forecasts are immediately stored by category. The values of 0, 2, 4, and 6, are assigned to the 0-1, 2-3, 4-5, and greater than or equal to 6 inch categories, respectively. The MOS snow amount forecast system is summarized in Appendix C.

The fact that all snow amount forecasts are verified (i.e., unconditional upon snow occurrence) creates a discontinuity in the categorical database. This means that the majority of verifiable hits will occur in category 0, even including cases of cloudless days. The stratification of local whole inch snow amount forecasts into MOS categories creates an additional verification problem which tends to benefit both forecast systems at the lowest and highest categories. For example, at the lowest category, a forecast for clear skies (category 0) when 1 inch of snow accumulated would verify correctly since all observations of less than 2 inches are included in the sample. At the highest category, a forecast of any whole inch amount greater than or equal to 6 inches would verify correctly if the observed value is greater than or equal to 6 inches, regardless of the magnitude of the error between the forecast and the observation.

3.2. Discussion of Results

Tables 3a and 3b show the consolidated contingency tables and computed scores for the local and MOS snow amount forecasts, respectively. As previously mentioned, category 0 data prevailed in the sample, representing over 98% of the available information. Local and MOS scores were not only identical for this category but were of little interest since measurable snowfall was not observed at any given station for the majority of cool season days. Thus, there will be no further examination of category 0 data. Many interesting conclusions can be drawn from the remaining categorical data,

Table 3. Contingency tables and computed scores for (a) local and (b) MOS snow amount forecasts.

(a) LOCAL

OBSERVED	FORECAST 0	2	4	6	TOTAL
0	142406	1361	214	55	144036
2	958	491	128	42	1619
4	157	134	84	28	403
6	70	60	40	66	236
TOTAL	143591	2046	466	191	146294

COMPUTED SCORES:
1. BIAS (0) = 1.00, (2) = 1.26, (4) = 1.16, (6) = 0.81
2. POD (0) = 0.99, (2) = 0.30, (4) = 0.21, (6) = 0.28
3. FAR (0) = 0.01, (2) = 0.76, (4) = 0.82, (6) = 0.65
4. CSI (0) = 0.98, (2) = 0.15, (4) = 0.11, (6) = 0.18
5. SKILL = 0.337
6. PFC = 0.98
7. % IMP. OVER MOS (SKILL) = 14.0

(b) MOS

OBSERVED	FORECAST 0	2	4	6	TOTAL
0	142503	1182	286	65	144036
2	1047	371	160	41	1619
4	178	113	78	34	403
6	88	44	56	48	236
TOTAL	143816	1710	580	188	146294

COMPUTED SCORES:
1. BIAS (0) = 1.00, (2) = 1.06, (4) = 1.44, (6) = 0.80
2. POD (0) = 0.99, (2) = 0.23, (4) = 0.19, (6) = 0.20
3. FAR (0) = 0.01, (2) = 0.78, (4) = 0.87, (6) = 0.74
4. CSI (0) = 0.98, (2) = 0.13, (4) = 0.09, (6) = 0.13
5. SKILL = 0.296
6. PFC = 0.98

however. Those dealing with local versus MOS comparisons will be discussed, and year to year trends will be acknowledged.

Fig. 4 shows CSI results for each category. Local forecasts improved on MOS guidance for all categories. The greatest improvements were in category 6. These improvements were probably related to inherent weaknesses in the MOS system. Categorical critical probability thresholds (see Appendix C) are tailored to the developmental sample data. Additionally, the small amount of developmental sample data for category 6 can create statistical instability when applied to independent data. Finally, snow amounts of greater than or equal to 6 inches are frequently associated with mesoscale situations which MOS is unable to resolve. Because of these weaknesses, a season with the potential for an anomalous number of heavy snow events could produce an anomalous number of MOS category 6 forecasts, many of which would fail to verify.

As Fig. 4 shows, category 4 CSI values were noticeably lower than those for category 6. For local forecasts, a possible reason for this particular disparity may be that storms capable of producing 4-inch snowfalls are often nebulous (i.e., ill-defined) or in their infancy (and thus somewhat unpredictable) at initial forecast time. This, combined with the small range of category 4, may have reduced the CSI.

Table 3 shows that for all categories (including 0), local forecasters exhibited much better skill at snow amount forecasting than MOS guidance, improving on MOS by 14%. It should be noted that the calculation of skill gives equal weight to all categories. For example, although over 95% of the category 0 forecasts verify correctly and make up 98% of the total sample, the effect of a much lower percentage of correctly verifying forecasts in the remaining categories is essentially equal to that of category 0. A likely reason for this improvement once again stems from the limited MOS database and the inability of MOS to resolve mesoscale snow events. It appears local forecasters utilize time and available guidance products wisely to enhance their skill.

Fig. 5 gives a chronological display of national CSI scores for local and MOS category 6 forecasts. Notable fluctuations from year to year were evident. Visual inspection shows a relatively sharp increase in scores from 1985/86 to 1986/87 and a sharp decline in scores from 1987/88 to 1988/89. Changes in both synoptic regimes and the degree of accuracy of numerical models for forecasting snow events played a role in these variations. While relatively few well-organized snowstorms prevailed nationwide in 1985/86, many more were evident in 1986/87. In 1988/89, a predominantly zonal flow existed in the middle troposphere; this regime likely aided in decreasing the predictability of snow-producing perturbations for forecasters and numerical models alike. Additionally, the number of category 6 events decreased sharply nationwide from 1987/88 to 1988/89 by nearly 50% (47 cases to 28 cases). Figs. 6 and 7 show chronological CSI scores for category 2 and category 4. As shown, variations in these categories were minimal compared to those for category 6. Local high marks were evident for category 2 in 1985/86 and 1987/88.

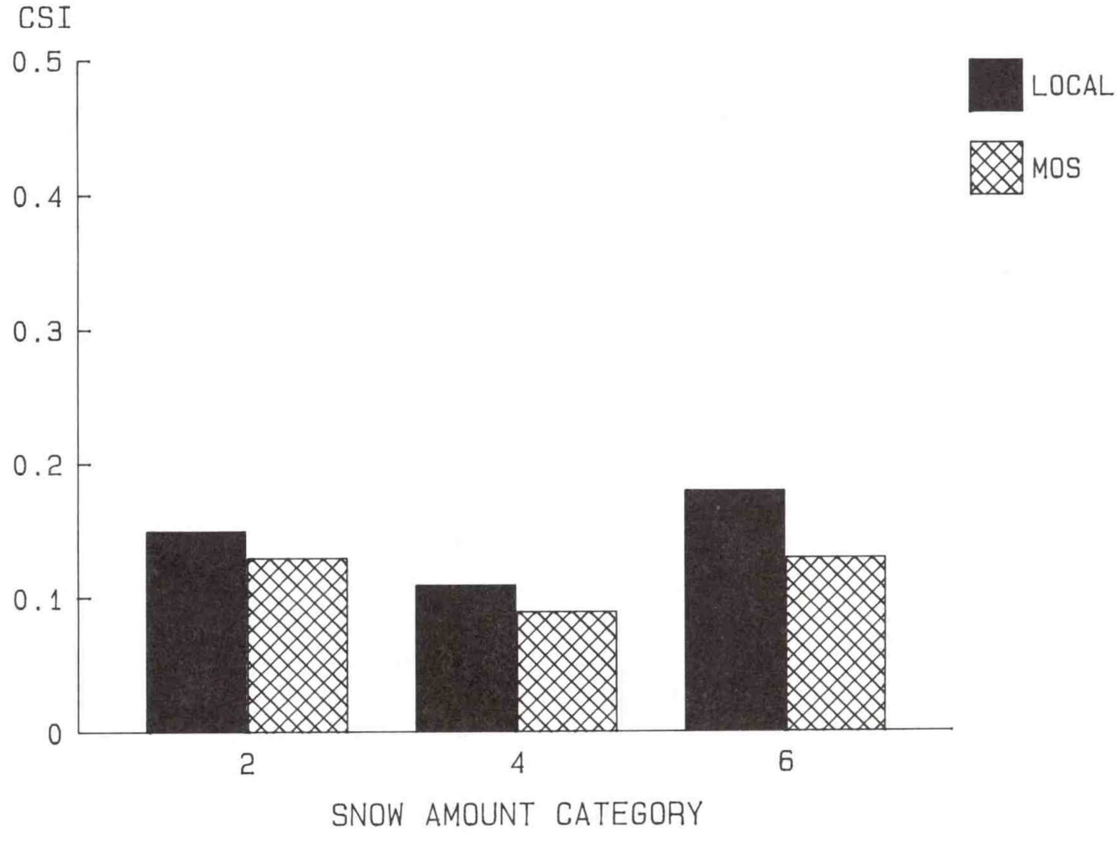

Figure 4. CSI for local and MOS snow amount forecasts, broken down into three verifying categories, for the 6-year sample. Categories 2(2-3 inches), 4(4-5 inches), and 6(>6 inches) are shown.

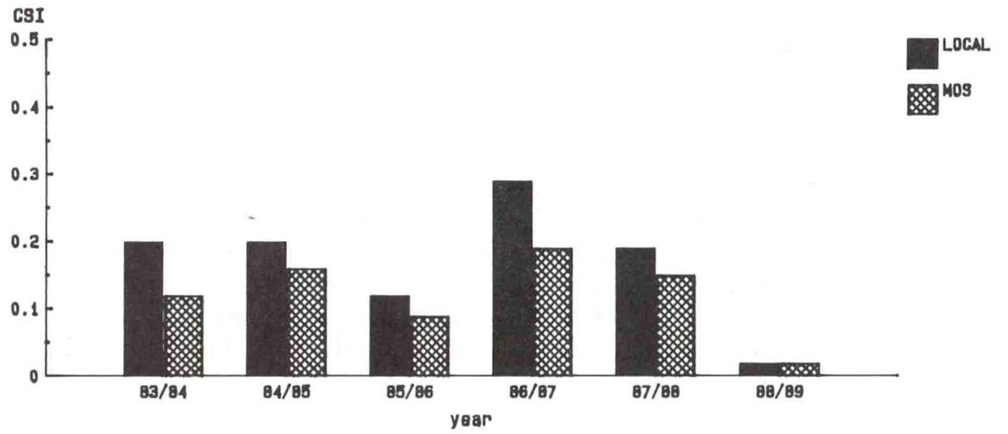

Figure 5. Six year trend of CSI for local and MOS snow amount forecasts of category 6 only.

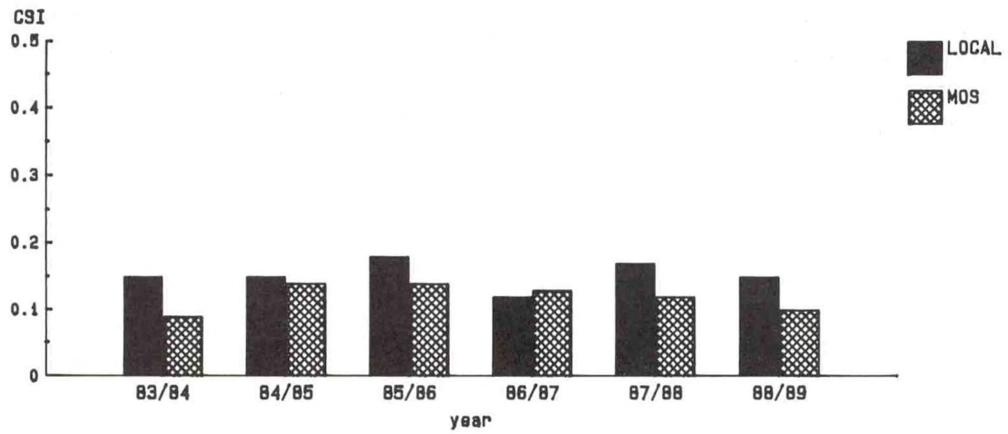

Figure 6. Same as Figure 5 except for forecasts of category 2.

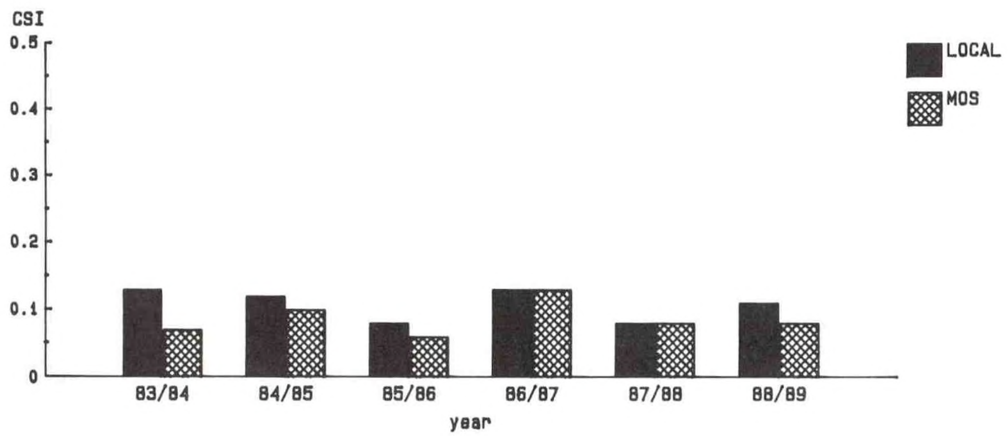

Figure 7. Same as Figure 6 except for forecasts of category 4.

Tables 4a - h show regional consolidated contingency tables for local (a - d) and MOS(e - h) forecasts for the 6-year sample. As with the national samples, category 0 scores were of little interest. Generally, bias was highest for categories 2 and 4, especially in the Eastern and Southern Regions. Fig. 8 shows the CSI for regional local and MOS data. Local forecasts generally improved upon MOS forecasts for all categories. Values were highest in the Eastern and Central Regions for both forecast systems, especially for category 6. The consistently low Southern and Western Region scores for all categories can be attributed to relatively few events in the sample (resulting in possible noise in the data) and the difficulty in forecasting and pinpointing snow amounts in the high terrain which covers much of the area.

As shown in Table 4, local forecasters improved over MOS in skill for all regions, with the greatest improvements occurring in the Central and Western Regions. Local improvement over MOS in the Eastern Region approached zero. Western improvements were best simply because MOS guidance is unable to effectively handle the orographic and mesoscale nature of snow events there. Also, the paucity of developmental data coming from Western Region stations contributed to less accurate guidance forecasts. Conversely, the lowest improvements shown in the Eastern Region were due in part to more synoptic-scale snow events, which were handled with more success by MOS.

Table 4. Contingency tables and computed scores of local and MOS snow amount forecasts, broken down by region.

(a) EASTERN LOCAL

OBSERVED	FORECAST 0	2	4	6	TOTAL
0	35915	477	80	28	36500
2	280	162	51	17	510
4	50	48	34	10	142
6	28	23	20	34	105
TOTAL	36273	710	185	89	37257

COMPUTED SCORES:
1. BIAS (0) = 0.99, (2) = 1.39, (4) = 1.30, (6) = 0.85
2. POD (0) = 0.98, (2) = 0.32, (4) = 0.24, (6) = 0.32
3. FAR (0) = 0.01, (2) = 0.77, (4) = 0.82, (6) = 0.62
4. CSI (0) = 0.97, (2) = 0.15, (4) = 0.12, (6) = 0.21
5. SKILL = 0.350
6. PFC = 0.97
7. % IMP. OVER MOS (SKILL)= 2.7

(e) EASTERN MOS

OBSERVED	FORECAST 0	2	4	6	TOTAL
0	35933	423	110	34	36500
2	281	141	74	14	510
4	47	37	42	16	142
6	34	16	28	27	105
TOTAL	36295	617	254	91	37257

COMPUTED SCORES:
1. BIAS (0) = 0.99, (2) = 1.21, (4) = 1.79, (6) = 0.87
2. POD (0) = 0.98, (2) = 0.28, (4) = 0.30, (6) = 0.26
3. FAR (0) = 0.01, (2) = 0.77, (4) = 0.84, (6) = 0.70
4. CSI (0) = 0.97, (2) = 0.14, (4) = 0.12, (6) = 0.16
5. SKILL = 0.341
6. PFC = 0.97

(b) SOUTHERN LOCAL

OBSERVED	FORECAST 0	2	4	6	TOTAL
0	41228	131	26	10	41395
2	51	27	5	2	85
4	9	4	3	2	18
6	11	2	5	3	21
TOTAL	41299	164	39	17	41519

COMPUTED SCORES:
1. BIAS (0) = 1.00, (2) = 1.93, (4) = 2.17, (6) = 0.81
2. POD (0) = 1.00, (2) = 0.32, (4) = 0.17, (6) = 0.14
3. FAR (0) = 0.00, (2) = 0.84, (4) = 0.92, (6) = 0.82
4. CSI (0) = 1.00, (2) = 0.12, (4) = 0.06, (6) = 0.09
5. SKILL = 0.248
6. PFC = 0.99
7. % IMP. OVER MOS (SKILL)= 16.4

(f) SOUTHERN MOS

OBSERVED	FORECAST 0	2	4	6	TOTAL
0	41296	74	226	3	41395
2	62	13	7	3	85
4	11	3	2	2	18
6	13	4	1	3	21
TOTAL	41382	94	32	11	41519

COMPUTED SCORES:
1. BIAS (0) = 1.00, (2) = 1.11, (4) = 1.80, (6) = 0.52
2. POD (0) = 1.00, (2) = 0.15, (4) = 0.11, (6) = 0.14
3. FAR (0) = 0.00, (2) = 0.86, (4) = 0.94, (6) = 0.73
4. CSI (0) = 1.00, (2) = 0.08, (4) = 0.04, (6) = 0.10
5. SKILL = 0.213
6. PFC = 1.00

(c) CENTRAL LOCAL

OBSERVED	FORECAST 0	2	4	6	TOTAL
0	47174	534	95	16	47819
2	474	238	60	21	793
4	66	58	40	16	180
6	22	27	15	26	90
TOTAL	47736	857	210	79	48882

COMPUTED SCORES:
1. BIAS (0) = 1.00, (2) = 1.08, (4) = 1.17, (6) = 0.88
2. POD (0) = 0.99, (2) = 0.30, (4) = 0.22, (6) = 0.29
3. FAR (0) = 0.01, (2) = 0.72, (4) = 0.71, (6) = 0.73
4. CSI (0) = 0.98, (2) = 0.17, (4) = 0.11, (6) = 0.18
5. SKILL = 0.353
6. PFC = 0.97
7. % IMP. OVER MOS (SKILL)= 21.9

(g) CENTRAL MOS

OBSERVED	FORECAST 0	2	4	6	TOTAL
0	47148	528	119	24	47819
2	527	180	65	21	793
4	82	55	28	15	180
6	26	21	26	17	90
TOTAL	47783	784	238	77	48882

COMPUTED SCORES:
1. BIAS (0) = 1.00, (2) = 0.99, (4) = 1.32, (6) = 0.86
2. POD (0) = 0.99, (2) = 0.23, (4) = 0.16, (6) = 0.19
3. FAR (0) = 0.01, (2) = 0.77, (4) = 0.88, (6) = 0.78
4. CSI (0) = 0.97, (2) = 0.13, (4) = 0.07, (6) = 0.11
5. SKILL = 0.290
6. PFC = 0.97

(d) WESTERN LOCAL

OBSERVED	FORECAST 0	2	4	6	TOTAL
0	18089	219	13	1	118322
2	153	64	12	2	231
4	32	24	7	0	63
6	9	8	0	3	20
TOTAL	18283	315	32	6	18636

COMPUTED SCORES:
1. BIAS (0) = 1.00, (2) = 1.36, (4) = 0.51, (6) = 0.30
2. POD (0) = 0.99, (2) = 0.28, (4) = 0.11, (6) = 0.15
3. FAR (0) = 0.01, (2) = 0.80, (4) = 0.78, (6) = 0.50
4. CSI (0) = 0.98, (2) = 0.13, (4) = 0.08, (6) = 0.13
5. SKILL = 0.280
6. PFC = 0.98
7. % IMP. OVER MOS (SKILL)= 36.4

(h) WESTERN MOS

OBSERVED	FORECAST 0	2	4	6	TOTAL
0	18126	157	35	4	18322
2	177	37	14	3	231
4	38	18	6	1	63
6	15	3	1	1	20
TOTAL	18356	215	56	9	18636

COMPUTED SCORES:
1. BIAS (0) = 1.00, (2) = 0.93, (4) = 0.89, (6) = 0.45
2. POD (0) = 0.99, (2) = 0.16, (4) = 0.10, (6) = 0.05
3. FAR (0) = 0.01, (2) = 0.83, (4) = 0.89, (6) = 0.89
4. CSI (0) = 0.98, (2) = 0.09, (4) = 0.05, (6) = 0.04
5. SKILL = 0.205
6. PFC = 0.97

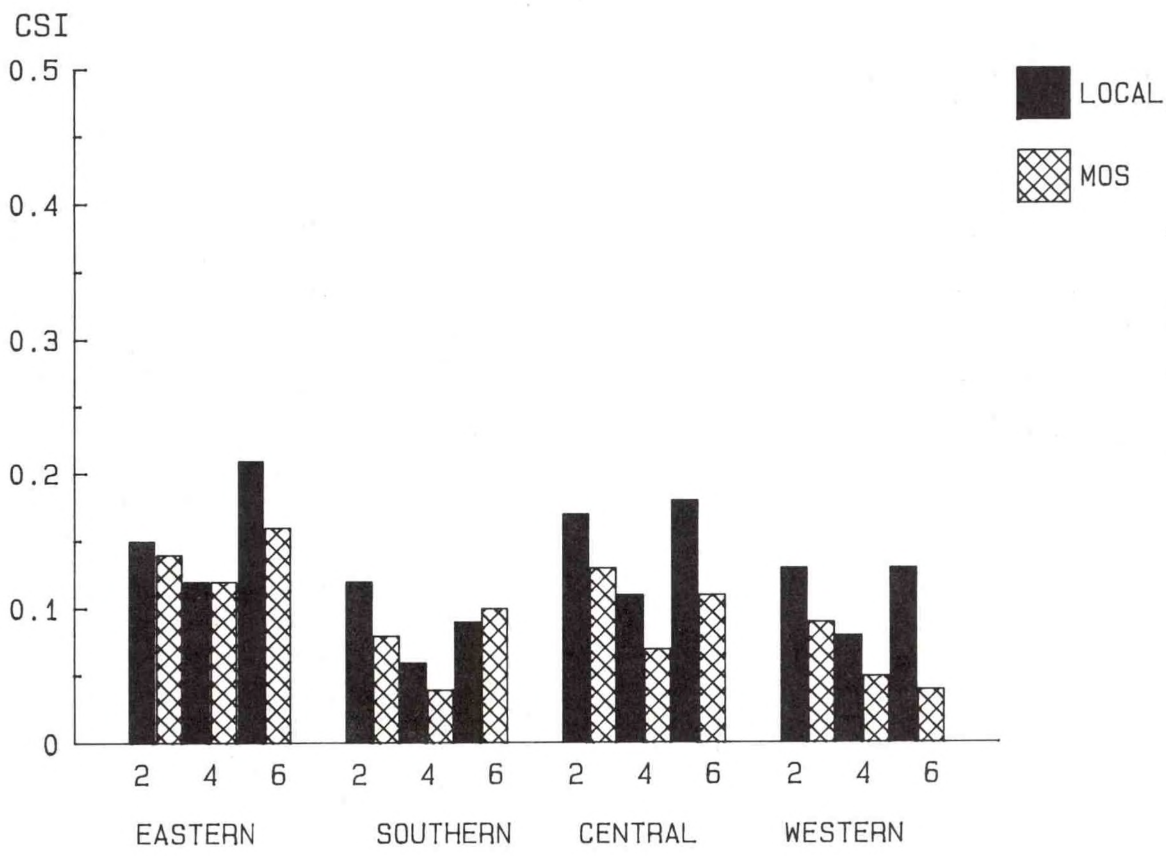

Figure 8. Regional breakdown of CSI for local and MOS snow amount forecasts, by category, for the 6-year sample.

--THIS PAGE INTENTIONALLY LEFT BLANK--

4. SUMMARY AND CONCLUSIONS

This paper discussed the AEV-based precipitation type and snow amount verification results. Precipitation type forecasts for the consolidated national sets of verifying-time and 2-hour window data were analyzed. The verification of regional consolidated 18-hour verifying-time set data was also presented. In general, scores for freezing precipitation (Z) were worse than those for frozen precipitation (S). The two categories were then compared for the 2-hour window and verifying-time sets. Slight improvements to local forecasts of Z verified with the 2-hour window were attributed in part to the increased number of potential verifying observations and forecasters tendency to think in terms of a temporal range. For S, a slight decrease in the scores was evident, perhaps since a changeover to Z is more likely during the 2-hour window. A general trend was not present for either freezing versus frozen/liquid or frozen versus freezing/liquid during the five year sample, and a degree of skill was evident for all forecasts of precipitation type. For regional data, the highest CSI values for Z existed in the NWS Southern and Central Regions. Shallow cold-air precipitation events are more common in these regions than in any other. Western region scores for Z were deemed insignificant due to a lack of forecasts and events. CSI scores for S were nearly equal among all regions, except Central Region scores were a shade higher than the others since S occurred there most frequently. MOS skill scores were higher than local skill scores for all regions except the Eastern Region.

Snow amount forecasts from the consolidated national and regional databases were analyzed. The forecasts were verified on every day of the cool season. This creates a situation where most of the verifying forecasts and observations fall into category 0 (0-1 inches). National CSI scores showed local improvement over MOS for categories 2 (2-3 inches), 4 (4-5 inches), and 6 (> 6 inches). Local forecasts had much greater skill than MOS forecasts overall, indicating the ability of the forecasters to better evaluate the potential for snowstorms. Further conclusions of note were drawn primarily from category 4 and category 6 data. Local improvements over MOS guidance were greatest for category 6. The MOS bias for category 4 was relatively high, a result of an algorithm flaw in the MOS best category decision making process. CSI values were higher for category 6 than category 4 for both systems. This may have been due to the nature of potential snowstorms, as those that create category 6 data tend to be more well organized than those that create category 4 data at forecast initialization time. Year-to-year fluctuations were evident in category 6 during the period of study. These fluctuations were likely due to two factors: (1) differences in the general synoptic regime which produced varying numbers of well-organized snowstorms, and (2) variations in effective handling of snow events by numerical models. Fluctuations were minimal for category 2 and category 4 forecasts. Analysis of regional forecasts showed that local CSI scores improved upon MOS scores for most every instance. In general, the Eastern and Central Regions had the highest scores; the Southern and Western Regions the lowest. Most likely, the Western Region scores were impacted by the difficulty in forecasting snow accumulations due to orography and predominant mesoscale features.

Also, there were relatively few cases in the sample. Greatest local improvement over MOS in skill was noted in this region as well, since the guidance is inherently inadequate. In contrast, for the Eastern Region, the apparent ability of MOS to forecast synoptic-scale storms reduced the gap between the two forecast systems.

5. REFERENCES

Bocchieri, J. R., 1983: Automated guidance for forecasting snow amount. Monthly Weather Review , 111, 2099-2109.

__________, and G. J. Maglaras, 1983: An improved system for forecasting precipitation type. Monthly Weather Review, 111, 405-419.

Dagostaro, V. J., 1985: The National AFOS-era verification data processing system. TDL Office Note 85-9, National Weather Service, NOAA, U.S. Department of Commerce, 45 pp.

Glahn, H. R., and D. A. Lowry, 1972: The use of model output statistics (MOS) in objective weather forecasting. Journal of Applied Meteorology, 11, 1203-1211.

National Weather Service, 1985: Public/aviation forecast verification. NWS Operations Manual Chapter C-73, NOAA, U.S. Department of Commerce, 20 pp.

______, 1983: The FOUS-12 bulletin. NWS Technical Procedures Bulletin No. 325, NOAA, U.S. Department of Commerce, 11 pp.

______, 1982a: The National Verification Plan. NOAA, U.S. Department of Commerce, 81 pp.

______, 1982b: Operational probability of precipitation type forecasts based on model output statistics. NWS Technical Procedures Bulletin No. 319, NOAA, U.S. Department of Commerce, 14 pp.

______, 1982c: The use of model output statistics for predicting snow amount. NWS Technical Procedures Bulletin No. 318, NOAA, U.S. Department of Commerce, 14 pp.

Polger, P. D. and R. M. Thompson, 1986: New insights from the NWS public forecast verification. Preprints, 11th Conference on Weather Forecasting and Analysis, American Meteorological Society, 5 pp.

Ruth, D. P. and C. A. Alex, 1987: AFOS-era verification. NOAA-TDL Computer Program NWS TDL CP 87-2, 50 pp.

APPENDIX A: THE AFOS-ERA VERIFICATION (AEV) PROGRAM

All AEV data are initially collated at each WSFO in two separate formats: the Public Verification Matrix (PVM) and the Aviation Verification Matrix. The PVM consists of a combination of the manually-entered forecast (MEF) matrix, locally issued coded city forecasts (CCF's), corresponding MOS guidance, surface synoptic reports (SSM's), and surface airways observations (SAO's). Before the verification program is executed the forecaster builds the MEF matrix (Fig. A.1). This step is necessary to collect forecast values not available through automated procedures. The precipitation type and snow amount forecasts are included in the MEF matrix. Next, the verification program that accesses the AFOS database and retrieves the necessary data for storage in the PVM is executed. An example of the PVM is shown in Fig. A.2.

Data collected from the AEV system do not always match that specified by the National Verification Plan (NVP) (National Weather Service, 1982a). After the plan was published, local data stratifications were set to match MOS stratifications for the sake of consistency. For precipitation type forecasts, the verification methods now used vary greatly from those defined in the NVP. Precipitation type forecasts were to be verified in three 12-hour continuous <u>periods</u> rather than at finite times. The AEV system uses distinct verifying points in time and 2-hour windows about these points, all of which are valid at the <u>center</u> of a 12-hour period. Additionally, the verifying observation was to be the occurrence of freezing precipitation within the period or, failing this criterion, any <u>measurable</u> non-freezing precipitation. The AEV system uses the occurrence of <u>any</u> type of precipitation, whether or not it is measurable.

For snow amount, a discontinuity occurs in category comparisons. Originally, six snow amount categories were designated in the NVP. However, the introduction of MOS forecasts forced all local forecasts to be verified within the constraints of the four MOS categories (0-1, 2-3, 4-5, and >6 inches). Although this created continuity, it allowed a more lenient system of verification, especially for the lowest category (less than 2 inches). The NVP included a stratification of the 0-2 inch group: 0 and trace; and measurable snow less than or equal to 1 inch.

EXPMCPMEF						
WOUSOO KEX1 999999						
FORECAST CYCLE (MMDDHH)		110112		FORECASTER 01		
STATION XYZ	12H	18H	24H	30H	36H	42H
PRECIPITATION TYPE		1		2		2
SNOW AMOUNT			00			
CLOUD AMOUNT	3	3	4			
SIGNIFICANT WINDS (Y/N)						Y

Figure A.1 Sample Manually Entered Forecast (MEF) matrix a forecaster submits to the AEV verification program. More detail is described in Ruth and Alex (1987).

XYZ		1101	1200	FORECASTER 01	1101	0000	FORECASTER 02
ELEMENT	PROJ	MOS	LOCAL	OBSERVED	MOS	LOCAL	OBSERVED
TEMP M/M	12-24	26	30	30	42	45	44
DEG F	24-36	43	37	37	28	28	30
24H/12H	36-48	24	20	18	48	43	37
	48-60	39	37	34	26	23	18
12H POP	12-24	05	10	0	02	0	0
PERCENT	24-36	20	40	12	10	20	0
	36-48	20	20	0	30	50	12
POPT	18($\pm$1)	3	1	000 000	3	3	000 000
CTGY	30($\pm$1)	3	2	000 002	3	2	000 000
Z/F/L	42($\pm$1)	2	2	000 000	3	3	000 002
SNOW AMT	12-24	0	0	0	0	0	0
WINDS	42($\pm$3)	3515	Y	3318 3224	0514	N	0310 0118
SIG/DEG-KTS							
CLOUD AMT	12	2	3	3	2	1	1
CTGY	18	3	3	4	2	2	2
	24	3	4	4	2	3	3

Figure A.2 Sample Public Verification Matrix (PVM) that is sent for processing by the AEV program. Note that the data in the MEF (Figure A.1, forecaster 01) is used as input to the PVM. Further detail is shown in Ruth and Alex (1987).

APPENDIX B: THE MOS PRECIPITATION TYPE SYSTEM

The MOS precipitation type system (Bocchieri and Maglaras, 1983) is based on the condition that precipitation occurs. That is, only precipitation cases were included in the developmental sample (National Weather Service, 1982b). In MOS development, the definition of the predictand was as follows: for category 1 (Z), <u>any</u> observation of freezing precipitation, occurring singularly <u>or</u> with other precipitation types; for category 2 (S), an observation of frozen precipitation occurring singularly <u>only</u>; for category 3 (R), an observation of liquid <u>or</u> mixed liquid and frozen precipitation. Predictors included many low-level model output temperatures and winds, as well as transformed thermal fields.

The MOS best category selection process works as follows: the conditional probability of precipitation type for Z is checked against the preset threshold value. If the value is equalled or exceeded, Z is chosen <u>without</u> looking at the probabilities (and associated thresholds) for the other categories. If the threshold is not equalled or exceeded, S is checked. If the test fails again, R is automatically chosen. Threshold values are a function of relative frequency of the respective precipitation type for the entire sample. Lower thresholds were generally associated with lower relative frequencies.

Fig. B.1 shows the section of a FOUS-12 bulletin (National Weather Service, 1983) containing precipitation type forecasts. Precipitation type forecasts are available at projections of 6 to 60 hours in 6-hour increments. The AEV program only uses the 18-, 30-, and 42-hour projections since these projections generally fall at the midpoint of a local 12-hour forecast period.

	NMCFPCXYZ		FOUS12		KWBC	011450		
PROJ(H)	06	12	18	24	30	36	42	48
POPT	0003/3	0210/3	0516/3	0227/3	0439/3	0156/2	0273/2	0359/2

Figure B.1. Sample of FOUS-12 parameters for MOS probability of precipitation type (PoPT) forecasts for station XYZ, 12Z cycle. Only the 18-, 30-, and 42-hour projections are verified. More detailed information on the FOUS-12 bulletin may be obtained in National Weather Service (1983).

APPENDIX C: THE MOS SNOW AMOUNT SYSTEM

MOS snow amount forecasts exist in two forms: conditional and unconditional. Conditional forecasts imply that precipitation must occur and that it must be in the form of snow and/or sleet. The developmental predictand data base, therefore, only includes measurable snow and/or sleet cases. Predictor data chosen through linear screening regression include variables that only relate to precipitation <u>amount</u>, not precipitation <u>type</u>. Unconditional forecasts do not imply precipitation occurrence in the frozen form. An unconditional forecast is produced by multiplying the conditional forecast by factors of precipitation occurrence (i.e., PoP) and precipitation type. For the AEV sample, unconditional forecasts are stored.

The forecasts are immediately stored by category. The values of 0, 2, 4, and 6, are assigned to the categories of 0-1, 2-3, 4-5, and greater than or equal to 6 inches of snow, respectively. For the MOS forecasts, the unconditional probability of snow amount (Bocchieri, 1983) is estimated by (using category 2 as an example)

$$PoSA(2)=PoSA(S)(2) \times PoP(12) \times PoF$$

where PoSA(S)(2) is the conditional probability forecast of greater than or equal to 2 inches of snow given by the MOS equation; PoP(12) is the value of the PoP for the 12-24 hour period; and PoF is the mean value in the 12-24 hour time frame of probability of frozen precipitation (e.g. the category 2 forecast from the PoPT system). Mean PoF is defined as

$$PoF = \frac{PoF(12) + 2(PoF(18)) + PoF(24)}{4}$$

Note that the value from the middle of the period is weighted twice that of the beginning and ending values.

The threshold decision tree (National Weather Service, 1982c) for selecting the best category is identical to the precipitation type tree, with the exception of number of categories (three instead of two). Like precipitation type, thresholds are directly related to the relative frequency of occurrence of snow, by category, in a given MOS region. For snow amount, the rarer the event, the lower the mean probability of the event occurring, and hence the lower the threshold.

Due to errors in the program that determines snow amount thresholds, the decision tree was not strictly adhered to for the sample data. Rather than setting the best category to zero (0-1) if the 2 inch threshold was not exceeded, the process continued, comparing the unconditional probability of ≥ 4 inches with its associated threshold. It is possible that the sharp decline in thresholds between the 2 and the 4 inch category may have also contributed to the high bias shown in the data. The algorithm has since been corrected to be consistent with the original plan.

Fig. C.1 shows the section of the FOUS-12 bulletin containing the snow amount forecasts. An explanation is provided below the figure. Snow amount forecasts are available only at the 24-hour projection and are verified, along with local forecasts, in the 12-24 hour time frame.

	NMCFPCXYZ FOUS12 KWBC 011450	
PROJ		24
POSA	5625/4712/3201/2	

Figure C.1. Sample of FOUS-12 parameters of MOS snow amount forecasts for station XYZ, 12Z cycle. Further detail on the FOUS-12 bulletin may be found in National Weather Service (1983).

APPENDIX D: DEFINITIONS OF THE VERIFICATION SCORES

With the National Verification Plan (NVP) as a guideline, various statistical scores were generated from the raw AEV data for the local and MOS forecasts. Table D.1 shows the matrices from which the pertinent scores are developed. These scores are summarized as follows:

1. <u>BIAS(n)</u>. Defined as

$$X_{Tn}/X_{nT};$$

the ratio of the total number of forecasts to the total number of observations for category n. The optimum bias of any sample is 1.0.

2. <u>POD(n)</u> (Probability of Detection, category n). Defined as

$$X_{nn}/X_{nT};$$

the ratio of correctly forecast events to the total number of events for category n. Values range from 0 - 1, with 1 being the optimum.

3. <u>FAR(n)</u> (False Alarm Ratio, category n). Defined as

$$\frac{X_{Tn} - X_{nn}}{X_{Tn}};$$

the ratio of unsuccessful forecasts to the total number of forecasts for category n. Values range from 0 - 1, with 0 being the optimum.

4. <u>CSI(n)</u> (Critical Success Index, category n). Defined as

$$\frac{X_{nn}}{X_{nn} + (X_{Tn}-X_{nn}) + (X_{nT}-X_{nn})};$$

the ratio of the number of correct forecasts to correct forecasts plus unsuccessful forecasts plus unforecast events for category n. Generically, CSI means the fraction of the time an event was correctly forecast when there was indeed a threat. Values range from 0 - 1, with 1 being the optimum.

5. Heidke Skill Score (SKILL). Defined as

$$\frac{\Sigma X_{nn} - (\Sigma(X_{nT}X_{Tn})/X_{TT})}{X_{TT} - (\Sigma(X_{nT}X_{Tn})/X_{TT})} ;$$

the number of correct forecasts minus the expected value from the contingency table divided by the total number of forecasts minus the expected value from the contingency table. Values range from 0 - 1, with 1 being the optimum.

6. PFC (Percentage of Forecasts Correct). Defined as

$$\Sigma X_{nn}/X_{TT};$$

the ratio of the sum of the diagonals (correct forecasts) of each category n to the total number of forecasts.

7. Local SKILL Improvement over MOS. Defined as

$$\frac{SKILL_n(LOCAL) - SKILL_n(MOS)}{SKILL_n(MOS)} \times 100\%$$

measure of local improvement, or failure to improve, relative to MOS.

(Continued from inside front cover)

NOAA Technical Memorandums

NWS FCST 16	Weather Bureau April 1969 to March 1970 Verification Report With Special Emphasis on Performance Scores within Echelons. Robert G. Derouin and Geraldine F. Cobb, April 1971. (COM-71-00555)
NWS FCST 17	National Weather Service May 1970 to April 1971 Public Forecast Verification Summary. Robert G. Derouin and Geraldine F. Cobb, March 1972. (COM-72-10484)
NWS FCST 18	Long-Term Verification Trends of Forecasts by the National Weather Service. Duane S. Cooley and Robert G. Derouin, May 1972. (COM-72-11114)
NWS FCST 19	National Weather Service May 1971 to April 1972 Public Forecast Verification Summary. Alexander F. Sadowski and Geraldine F. Cobb, July 1973. (COM-73-11-55 7/AS)
NWS FCST 20	National Weather Service Heavy Snow Forecast Verification 1962 to 1972. Alexander F. Sadowski and Geraldine F. Cobb, January 1974. (COM-74-10518)
NWS FCST 21	National Weather Service April 1972 to March 1973 Public Forecast Verification Summary. Alexander F. Sadowski and Geraldine F. Cobb, June 1974. (COM-74-1 1467/AS)
NWS FCST 22	Photochemical (Oxidant) Air Pollution Summary Information. Stephen W. Harned and Thomas Laufer, December 1977. (PB-283868/AS)
NWS FCST 23	Low-Level Wind Shear: A Critical Review. Julius Badner, April 1979, 72 pp. (PB-300715)
NWS FCST 24	Probability Forecasting -- Reasons, Procedures, Problems. Lawrence A. Hughes, January 1980, 89 pp. (PB80-164353)
NWS FCST 25	National Weather Service Public Forecast Verification Summary--April 1973 to March 1978. Duane S. Cooley, Frederick S. Zbar, Dean F. Dubofsky, and A. Kristine Campbell, March 1981, 136 pp. (PB81-231714)
NWS FCST 26	National Weather Service 1980 Watch/Warning Verification: Flash Flood, Winter Storm and High Wind. A. Kristine Campbell, August 1981, 36 pp. (PB82-148719)
NWS FCST 27	National Weather Service 1981 Watch/Warning Verification: Flash Flood, Winter Storm and High Wind. A. Kristine Campbell, July 1982. (PB83-118018)
NWS FCST 28	National Weather Service Public Forecast Verification Summary, April 1978 to March 1982. Paul D. Polger, April 1983. (PB83-232173)
NWS FCST 29	Public Response to Hurricane Probability Forecasts. Jay Baker, January 1984. (PB84-158658)
NWS FCST 30	1982 and 1983 Watch/Warning Verification: Flash Flood, Winter Storm and High Wind A. Kristine Campbell, January 1985. (PB85-20-1899)
NWS FCST 31	A 20-Year Summary of National Weather Service Verification Results For Temperature and Precipitation. Gary M. Carter and Paul D. Polger, August 1986, 50pp. (PB88-235353)
NWS FCST 32	NWS Watch/Warning Verification Flash Flood, Winter Storm, and High Wind January 1980 - June 1986. Robert E. Laplante, July, 1987, 36 pp. (PB87-223053)

NOAA SCIENTIFIC AND TECHNICAL PUBLICATIONS

The National Oceanic and Atmospheric Administration was established as part of the Department of Commerce on October 3, 1970. The mission responsibilities of NOAA are to assess the socioeconomic impact of natural and technological changes in the environment and to monitor and predict the state of the solid Earth, the oceans and their living resources, the atmosphere, and the space environment of the Earth.

The major components of NOAA regularly produce various types of scientific and technical information in the following kinds of publications:

PROFESSIONAL PAPERS-Important definitive research results, major techniques, and special investigations.

CONTRACT AND GRANT REPORTS-Reports prepared by contractors or grantees under NOAA sponsorship.

ATLAS-Presentation of analyzed data generally in the form of maps showing distribution of rainfall, chemical and physical conditions of oceans and atmosphere, distribution of fishes and marine mammals, ionospheric conditions, etc.

TECHNICAL SERVICE PUBLICATIONS-Reports containing data, observations, instructions, etc. A partial listing includes data serials; prediction and outlook periodicals; technical manuals, training papers, planning reports, and information serials; and miscellaneous technical publications.

TECHNICAL REPORTS-Journal quality with extensive details, mathematical developments, or data listings.

TECHNICAL MEMORANDUMS-Reports of preliminary, partial, or negative research or technology results, interim instructions, and the like.