PACIFIC REGION TECHNICAL NOTES

81 - 007

March 6, 1981

LONG TERM S_1 SCORES AND THEIR INTERPRETATION

Peter Haering, Chief Meteorologist
Pacific Weather Centre

INTRODUCTION

The computer modelling community both in the U.S. and Canada has for years used the S_1 score as the primary support material to indicate the gradual improvement of computer produced guidance. The contention being that a decrease in the S_1 score indicates an improvement in the computer produced guidance.

At a recent W04 OIC's course held at PWC; one OIC asked the question "can you tell me if over the past 2 or 3 years forecasts have improved, because "feed-back" from users seem to suggest that they have not improved or may in fact be getting worse?" At the time I had no convincing reply one way or the other, and suggested that I did not believe that they had improved to any great extent. The question neverthe less was disturbing, since according to S_1 modelling statistics, the past3 years showed steady improvement in the S_1 score. Furthermore, statistics available suggest that subjective guidance based on computer guidance follows the trends established by the computer guidance.

A recent numerical weather prediction training workshop given by the Training Branch at PWC in mid February again stimulated discussion on S_1 scores and their interpretation.

DISCUSSION

1) Attachment 1 gives some back-ground material on the S_1 score as presented at NWP workshop.
2) Attachment 2 shows a graph of 48 hour S_1 scores from 1971 - 1980, for various Canadian, and U.S. models as well as the CMC subjective verification.
3) Attachments 3 is a copy of a recently received U.S. NWS Western Region Technical attachment.

.../2

The enclosed documentation demonstrates gradually lower S_1 scores towards 1980. However attachment 2 also shows a marked yearly fluctuation in the scores (i.e. the scores are higher in summer and lower in winter). The "out-of-phase" relationship in the 1970-1976 time span for the Canadian model has never been explained. (It is suggested that the high "tuning" of the Canadian Baroclinic Model for summer conditions may possibly provide an answer to that question). In any case S1 scores are not only sensitive to prognostic performance (Eg) but also to the circulation of the atmosphere (G_L). If G_L is large the S_1 score will be smaller. Thus, in winter when gradients are large the S_1 scores are lower than for summer months. So far one has always assumed that the lowering of the S_1 score is attributable only to improved prognostic performance and <u>not</u> to a change in circulation. Is this assumption valid? Certainly over the past 3 years at Vancouver the winters have been milder than normal and the summers cooler than normal. Ski operators on the north shore mountains of Vancouver have experienced a lack of snow during the past 3 years curtailing their operations drastically. The question should be asked "Is the claimed improvement of S_1 scores in the past 3 years due to a change in circulation rather than prognostic performance?" If this were true perhaps a plot of the root mean square errors (RMSE) would show little improvement over the time period in question. With this in mind available RMSE values were plotted using the CMC statistics for the Pacific Region (the same statistics that were used to compute the S_1 scores). Attachment 4 shows the results for 1976 - 1980. The gap of data in the 1978 period is probably due to our office move. This graph suggests no improvement in RMSE values for the summer months and at most questionable improvement in the winter months. Furthermore, the Chicago and Salt Lake City data really do not show much improvement after 1968.

Has the operational forecaster been fed propaganda regarding the performance of computer prognoses which is not justified? The recent NWP training workshop which presented verification statistics on performance of computer mean - sea level prognoses seems to say yes to that question.

ATTACHMENT I

S1 SCORES

(NWP TRAINING NOTES) AES TRAINING BRANCH

S_1 Score

Teweles and Wobus proposed the S_1 score for prog verification. Since then it has been routinely used by NMC and CMC. The S_1 score is a function of the pressure difference between pairs of points selected within the main portion of the forecast area. This score is defined as:

$$S_1 = 100 \frac{\sum_{i=1}^{N} |E_g|}{\sum_{i=1}^{N} |G_L|}$$

where: E_g error in the forecast pressure difference (gradient) between selected stations or geographical locations. Note that the absolute value is to be taken after the error in gradient has been calculated.

G_L observed or forecast pressure difference (gradient) whichever is larger.

Characteristics of the S_1 score are:

1) the full range of S_1 is from 0 to 200. A good prog will have a good (i.e., low) score, but a bad S_1 score (e.g. $S_1 > 60$ or 65) does not necessarily imply a bad prog. If the gradients are strong, the S_1 score will be very sensitive to the exact placing of the gradients. Hence, a better S_1 score will often be achieved by not forecasting a development at all, than by misplacing it.

2) the S_1 score has a seasonal trend, i.e., larger scores in the summer than in the winter, because pressure gradients are smaller in summer. This characteristic is due to the fact that smaller gradients give a smaller denominator in the above formula leading then to a larger S_1 score.

3) for large, intense systems, better (i.e., lower) S_1 scores are obtainable than for weak, poorly defined pressure patterns. This again is due to the small denominator when weak systems are considered.

4) the S_1 score is more sensitive to poor forecasts on an individual system when a smaller area is being verified. With a larger area the error is averaged in with better forecasts on neighbouring systems.

5) there has been no standard established as to what constitutes a significant improvement in S_1 scores, i.e., how much better is 65 over 75. However, a long term trend to lowering scores must imply increased forecasting ability.

6) modifications to the S_1 score include adding a handicap and/or a running mean of seven S_1 scores. See Teweles' article if interested.

ATTACHMENT 2

48HR CMC S1 SCORES

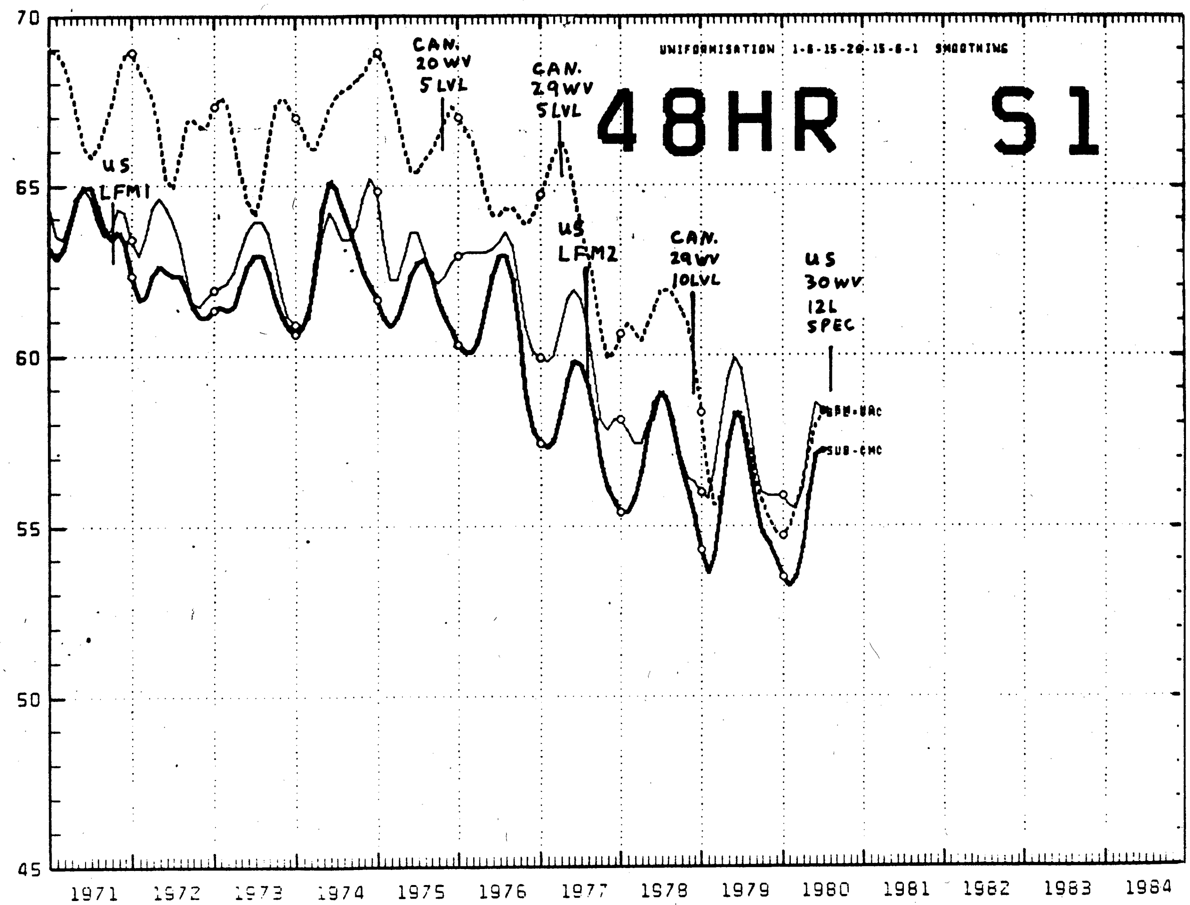

FIGURE 2

STD/CMC

ATTACHMENT 3

FROM U.S. NATIONAL WX. SERVICE

WESTERN REGION

WESTERN REGION TECHNICAL ATTACHMENT
No. 81-5
February 10, 1981

FORECAST VERIFICATION GRAPHS

Each year SSD updates three long-term verification graphs for use in showing changes of skill over the last 20-30 years. The current graphs given below include data through December 31, 1980.

The NMC 36-hour surface prognostic chart graph, is a plot of the S_1 score [1] of the 6- and 7-layer PE and spectral prognoses, Figure 1. (The spectral progs began August 12, 1980.) The S_1 score can be considered as the mean vector error of the forecast geostrophic wind. The lower the score, the better the prog. The skill percentage indicated on the left margin is an arbitrary evaluation assuming the average S_1 score of 24-hour persistence forecasts ($S_1 = 0.70$) as no skill, and the average S_1 score difference between manual isobaric analyses of the same plotted surface data ($S_1 = 0.20$) as 100% skill. The curve from 1948-1974 represents verification scores of the manual, and (after 1958) man-machine mix 30-hour prognoses. NMC stopped preparing this prognosis in 1974. The curve from 1967 to 1980 is for 36-hour PE and recently spectral machine prognoses. The verification area covers roughly the United States, southern Canada, and adjacent Pacific and Atlantic ocean areas. Note the marked improvement after 1958. This was the year automatic data processing made NWP guidance available to NMC forecasters before they issued their progs. The PE model became operational in 1966, the big improvement from 1977 to 1978 is attributed to NMC's decreasing the mesh length of the PE model.

The Chicago graph, Figure 2, is a smoothed plot of the percentage of correct temperature and precipitation forecasts for the 3 forecast periods issued twice a day. The verification procedure has remained unchanged since 1942; therefore, the trends rather than absolute values can be considered representative of the improvement of NWS forecasts in general. (We are indebted to Mr. Ray Waldman, MIC, CHI WSFO, for making this data available to us each year.) The marked improvement in early 1950s is considered to be a reflection of the benefits of the interaction between forecasters and professors when the forecast office was moved in 1951 to be colocated with Dr. Sverre Petterssen's University of Chicago Department of Meteorology. Improvement over the last 20 years must be attributed in part to improving NWP guidance.

The Salt Lake City maximum temperature graph, Figure 3, is a smoothed plot of the number of misses by 10 degrees or greater of tomorrow's maximum temperature. Newspaper data were used to develop the values from 1949 to 1960. We tried to include precipitation forecasts, but they were not stated precisely enough in the paper to verify. Here again, the improvement since 1960 should be attributed in part to improved NWP guidance.

These are the only consistent long-term verifications that we know of to show the remarkable forecast improvements that have been made since the advent of operational NWP in 1954.

Reference:

[1] "Verification of Prognostic Charts" by Teweles and Wobus, Bulletin of AMS, Volume 35 (1954), pp. 455-463.

NMC 30-Hour Surface Prognostic Charts 1948–1974

NMC 36-Hour PE Surface Prognostic Charts 1967–1980

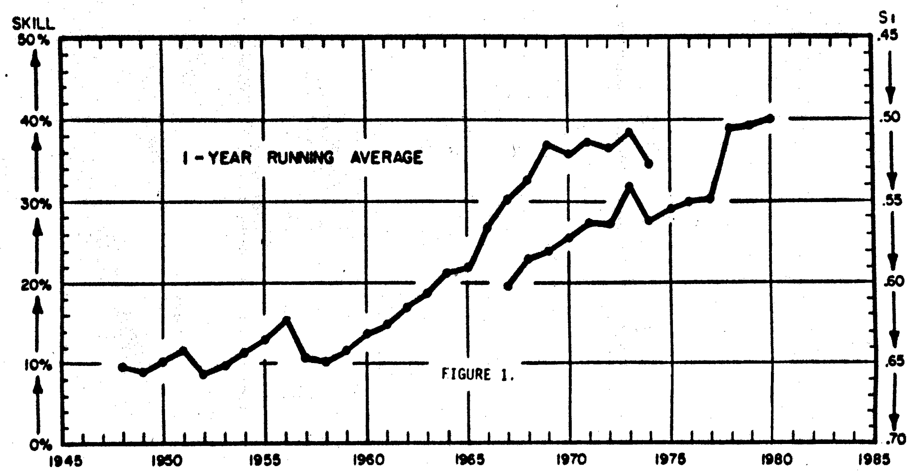

FIGURE 1.

Percentage Of Correct Weather & Temperature Forecasts Chicago, Illinois 1942–1980

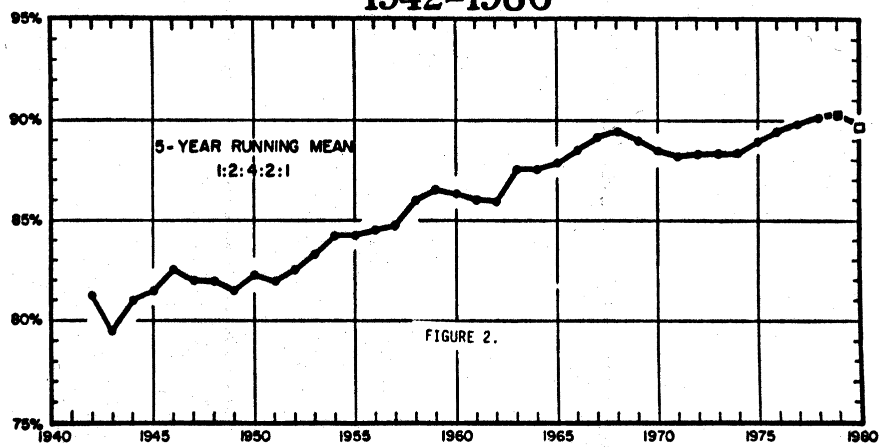

FIGURE 2.

Temperature Forecast Errors ≥ 10 Degrees Salt Lake City, Utah 1948–1980

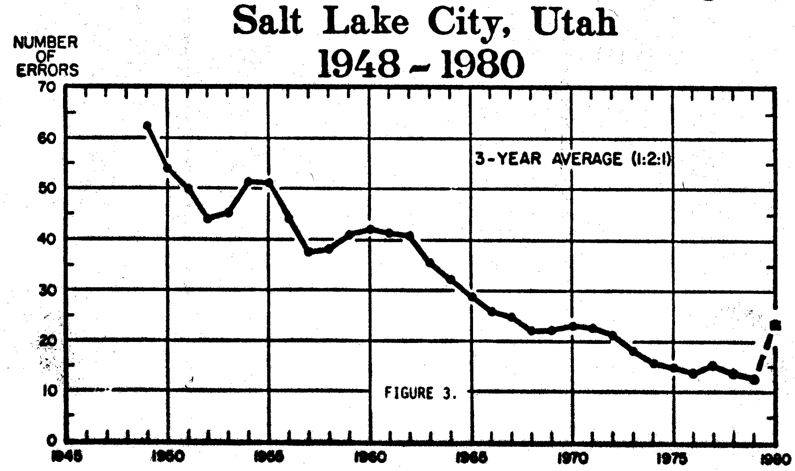

FIGURE 3.

ATTACHMENT 4

PLOT OF CMC RMSE SCORES FOR PACIFIC REGION