ROYAL OBSERVATORY, HONG KONG

Technical Note (Local) No. 70

ROYAL OBSERVATORY'S

OBJECTIVE FORECAST VERIFICATION SCHEMES

by

S.W. Li

Published February 1997

Prepared by

Royal Observatory
134A Nathan Road
Kowloon
Hong Kong

551.509(512.317)

SUMMARY

This report describes the four objective schemes used by the Royal Observatory, during various periods from 1984 to present, to verify the accuracy of its weather forecasts. The description includes: the general methodology employed in objective forecast verification, elaboration of the four verification schemes, and modifications made to them. Verification results are presented, and compared where appropriate. The scores from objection verification, indicating the accuracy of Royal Observatory's weather forecasts, showed a sharp increase in 1985, then levelled off in the 1990s. An attempt is also made to see if there is any correlation between the verification results and those obtained in public opinion surveys undertaken by independent survey companies.

CONTENTS

CONTENTS (cont.)

FIGURES

FIGURES (cont.)

TABLES

1. INTRODUCTION

One of the main responsibilities of the Royal Observatory (RO) is to provide weather forecasts to the public in order to reduce loss of life and damage to property. The accuracy of weather forecasts is one of the key indicators of the Observatory's performance.

Starting from 1984, objective verification schemes were used to assess the accuracy of RO's public weather forecasts. This report describes the four verification schemes to date, the modifications made to them and the forecast accuracy statistics accumulated so far. Data were collected from relevant departmental files and internal reports.

2. OBJECTIVE VERIFICATION OF WEATHER FORECASTS

2.1 Objective forecast verification

In objective forecast verification, the forecasts and actual observations are coded according to a set of rules before comparison. The schemes were devised with a view to reflecting as closely as possible how the public would evaluate the accuracy of weather forecasts from their point of view.

Each forecast is broken into a number of elements for verification, namely:

(i) wind,
(ii) state of sky,
(iii) precipitation,
(iv) visibility, and
(v) temperature.

Wind direction, gustiness of winds, squalls, humidity, descriptive temperature changes (such as milder and cooler, etc.) and other weather phenomena (such as frost, hail, tornado and waterspout, etc.) are not verified.

The verification process is computerised and is conducted separately for different weather elements according to their respective "score tables" (explained below). Some of the weather elements, e.g. state of sky and temperature, are further broken down into their constituent components for more detailed comparison. As an example, please see Appendix I for verification procedures for the state of sky.

To take into account the significance of individual weather elements at different times of the year, different weightings, forming a "weighting table", are assigned to different weather elements for each month according to climatology and our assessment of the relative importance of the elements in the eyes of the public for that month.

After verifying the accuracy of each weather element, the final score for a forecast is obtained by summing the products of individual element scores and the corresponding weightings.

2.2 Accurate/acceptable forecasts

To assess the accuracy of weather forecasts as perceived by the public, a categorisation scheme is probably more useful than the monthly mean score since the public would be more concerned with the number of "accurate/acceptable" forecasts than the "averaged" performance of the weather forecasts over a month. Therefore, a subjective "accurate/acceptable" mark is designated as the point below which a forecast is considered not "accurate/acceptable" in the public's eyes.

Since the forecast verification started in 1984, different marks had been assigned as the threshold for "accurate/acceptable" forecasts in different times and in various documents. Percentage figures of forecasts with marks higher than or equal to 75 are available on file most of the time from January 1983 to June 1987. 75 is taken in this report as the "accurate/acceptable" mark for comparison of results for this period. Starting from July 1987, the "accurate/acceptable" mark was set at 80. To reflect the rising expectation of the public on the accuracy of RO's weather forecasts, a higher "accurate/acceptable" mark, 85, was adopted commencing January 1996. Percentage figures of forecasts with marks higher than 85 have been re-computed for the period from November 1994 to March 1996.

Percentages of "accurate/acceptable" forecasts have been used, in one form or another, as a key indicator of the performance of the RO in published documents like the Controlling Officer's Report in each year's Estimates prepared by the Financial Secretary and other discussion papers within the Government. Table 1 below lists the yearly percentages of "accurate/acceptable" forecasts given in these documents as from 1983.

Table 1. PERCENTAGE OF FORECASTS PREDICTING CORRECTLY THE MAJOR ASPECTS OF THE WEATHER ("ACCURATE/ACCEPTABLE" FORECASTS)

YEAR	Percentage of forecasts predicting correctly the major aspects of the weather ("accurate/acceptable" forecasts)
1983	76
1984	78
1985	84
1986	90
	---------- *
1987	77
1988	84
1989	86
1990	86
1991	89
1992	87
1993	87
1994	85
	---------- **
1995	87

Note :

* -- S4 replaced S1a and S2 in July 1987 with a new threshold of 80 for "accurate/acceptable" forecasts.

** -- The threshold for "accurate/acceptable" forecasts was increased from 80 to 85.

2.3 Persistence forecasts

A persistence forecast is a forecast coded up by the computer based on the actual weather conditions, assuming that "tomorrow's weather will be the same as today". It is then verified in the same way as a normal weather forecast. The difference between the score of a weather forecast and that of the corresponding persistence forecast, (F - P), is a measure of forecasters' skill in "catching" changes in weather.

3. EVOLUTION OF OBJECTIVE VERIFICATION SCHEMES

Altogether four schemes have been used in RO's history of forecast verification. Each scheme is described in detail in this and the next section.

The first verification scheme, Scheme 1 (S1), was devised and implemented in August 1984. Weather forecasts issued daily at 1915H were first verified as early as January 1983. Modifications to the scheme were introduced shortly after S1's implementation to correct minor discrepancies in verifying the state of sky and precipitation. Results of the revised scheme, S1a, were re-calculated from January 1983. In comparison with later schemes, S1 and S1a were rather rudimentary.

In February 1985, another scheme, Scheme 2 (S2), was drawn up based on a newly devised weighting table. In this scheme, score tables for winds, precipitation, temperatures and visibility were refined. The scheme was put into operation in parallel with S1a. Backlog data from January 1983 were also computed to allow comparison of the two schemes.

To measure forecasters' skill in predicting changes in the weather, a third scheme, Scheme 3 (S3), was temporarily implemented manually in CFO commencing April 1985. Scores based on S3 were available from January 1985 to March 1986.

The fourth scheme, S4, replacing S1a and S2, was formulated and put into operation in June 1987. All scheduled public weather forecasts were verified on a daily basis and scores are available from July 1987. A total of four minor modifications to S4 were made so far, in October 1988, November 1989, February 1993 and March 1996. The software of S4 was completely re-written in C language in 1994 when RO's main computer was replaced. The new software was put into operation in November 1994.

Figure 1 is a time chart showing the changes and modifications that have taken place since the introduction of objective forecast verification. Figure 2 shows the periods when data are available from various schemes since 1983.

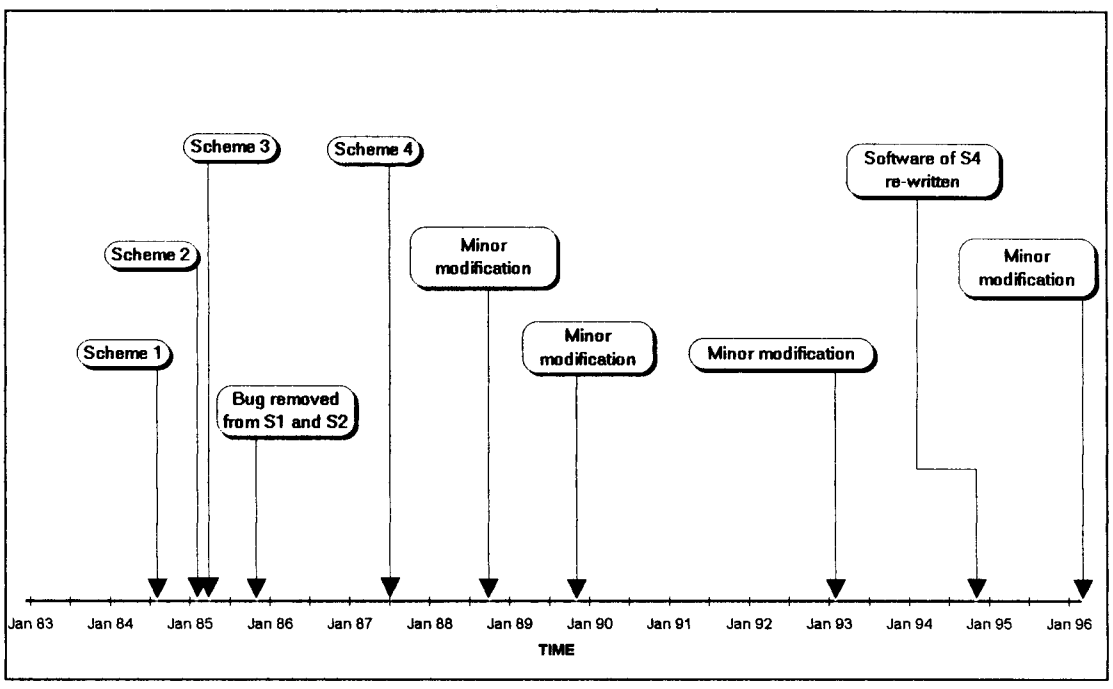

Figure 1. Time chart for objective forecast verification schemes.

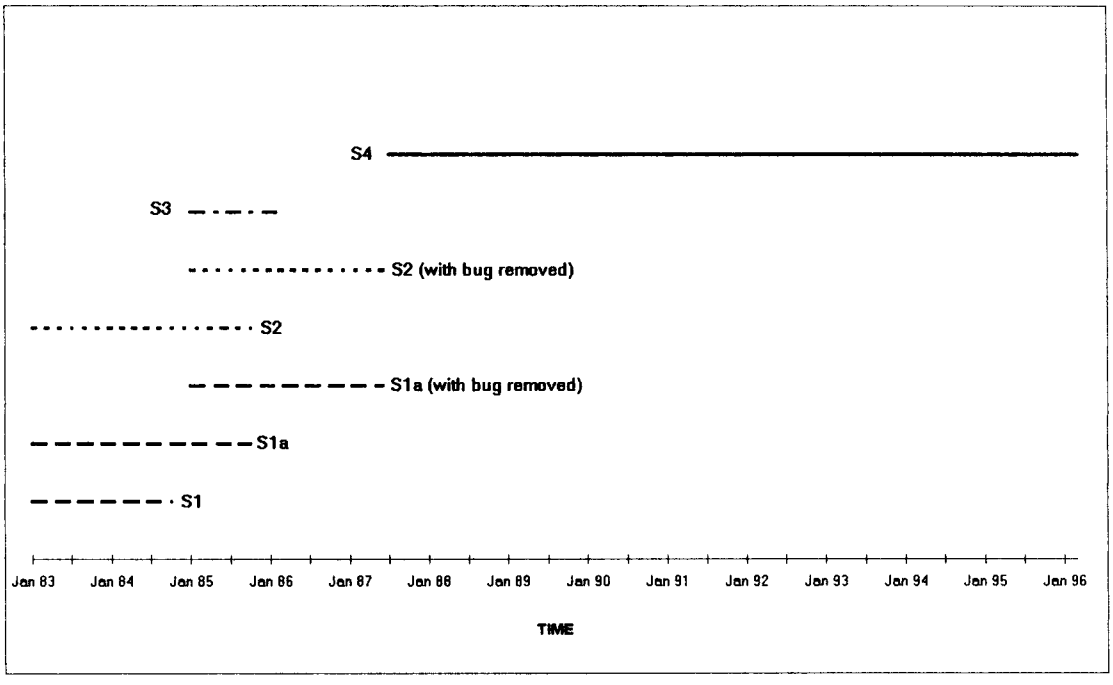

Figure 2. Data availability for the four forecast verification schemes.

4. COMPARISON BETWEEN DIFFERENT VERIFICATION SCHEMES AND MODIFICATIONS

4.1 Scheme 1 (S1, S1a)

In S1, weather elements were verified individually based on discrete skill score tables for each element. Categorisation of the score tables was simple and for most cases either 100 marks or 0 mark was awarded to a forecast element. The first modification (S1a) introduced more categories into the score tables to allow a finer comparison of the forecasts and the actual weather. Since S1a is an improved version of S1 and the monthly mean scores are also available from January 1983, S1a results, rather than S1, are used in this report.

4.2 Scheme 2 (S2)

S2 is quite similar to S1a except for the score and weighting tables. Score tables in S2 had finer categories than those of S1a. As an example, please see Appendix II(a) for the score tables of S1a and S2 for verification of visibility. According to a review of S1a and S2 in May 1986, the weighting tables for individual weather elements in S1a were stricter than those in S2. When comparing the weighting tables (Appendix II(b)), it is found that in S1a, relatively high weightings were given to elements pertinent to particular months, whereas in S2, the weightings tended to spread more evenly over all the elements throughout the year. The review considered that S1a's weighting tables were more realistic.

A programming error ("bug") was discovered in the computer software for S1a and S2 in November 1985. Results starting from January 1985 were re-calculated for both schemes with the corrected software. Monthly mean scores from January to October 1985 based on the software with and without the bug are compared in this report to see whether the bug has introduced significant error to the scores. The mean scores and standard deviations are listed in the following table:

	Mean	Standard Deviation
S1a with bug	70.7	4.86
S1a without bug	71.7	4.67
S2 with bug	83.5	3.28
S2 without bug	82.3	4.08

It can be seen that the differences between scores before and after the bug is removed are small, well less than the standard deviation of each mean score. Therefore, the error due to the bug is insignificant in so far as the monthly mean scores are concerned. In plotting graphs for comparison in this report, monthly mean scores computed with the old software are used up to December 1984, and those computed with the corrected software are used afterwards.

4.3 Scheme 3 (S3)

This scheme was specially designed to find out whether forecasters could correctly predict changes in the weather. The final score was the aggregate score of the number of right or wrong occasions in forecasting change for different weather elements.

The percentage for "Strike" forecasts and "False Alarm" forecasts were calculated according to the following formulae:

$$\text{Strikes (POD)} = \frac{\text{No. of occasions correctly forecasting 'change' in weather}}{\text{No. of occasions that 'change' in weather actually occurred}} \times 100\ \% \quad \text{------ (1)}$$

$$\text{False Alarm (FAR)} = \frac{\text{No. of 'change' forecasts that were wrong}}{\text{No. of 'change' forecasts issued}} \times 100\ \% \quad \text{------ (2)}$$

4.4 Scheme 4 (S4)

With experience gained on forecast verification and better computer facilities, quasi-continuous score tables, instead of discrete categories, were drawn up for S4 so as to eliminate "jumps" in the scores of forecasts which differed only slightly from each other. During the operation of S1a, S2 and S3, only the daily 1915H forecasts were verified. With the introduction of S4 since July 1987, scores for public weather forecasts issued at all scheduled times as well as those for 3-day forecasts have been computed.

S4 also takes into account whether a forecast can correctly predict changes in the weather for the specified forecast period. More lenient score tables are used for verification if the forecasts can correctly "catch" changes in the weather

either between the immediate past and the forecast period, or during the forecast period.

The first modification of S4 was introduced in October 1988 to refine the calculation of rainfall for verification. The second and the third one were necessary as a result of the termination of manual observations at Waglan Island in November 1989 and at Cheung Chau in February 1993 respectively. Verification of winds, weather and/or visibility were adjusted slightly to accommodate the changes. The fourth was made in March 1996 to introduce a new algorithm for verifying visibility when haze is forecast. These modifications are summarised in Appendix III.

The effect on scores introduced by the first and third modification were not indicated in the files. For the second modification, the effect was considered to be minimal. As for the fourth modification, monthly mean scores based on S4 with and without the modification have been computed for the period from January 1994 to August 1994. These scores are presented below and compared to see whether the fourth modification has introduced any significant change to the scores:

	Mean	Standard Deviation
S4 without the fourth modification	88.2	2.67
S4 with the fourth modification	88.1	3.76

It can be seen that the difference between scores with and without the fourth modification is small, well less than the standard deviation of the mean scores. Thus, the change due to the modification is insignificant in so far as the monthly mean scores are concerned.

5. FORECAST ACCURACY STATISTICS

5.1 Scheme 1a and 2 (S1a and S2)

The monthly mean scores from S1a and S2 are plotted together on Figure 3 and the 12-month running averages on Figure 4. As evident in Figure 4, scores are generally on a levelling trend in 1983 and most of 1984, until late 1984 when they started to rise. The increasing trend continued till July 1987 (after which S1a and S2 were no longer used).

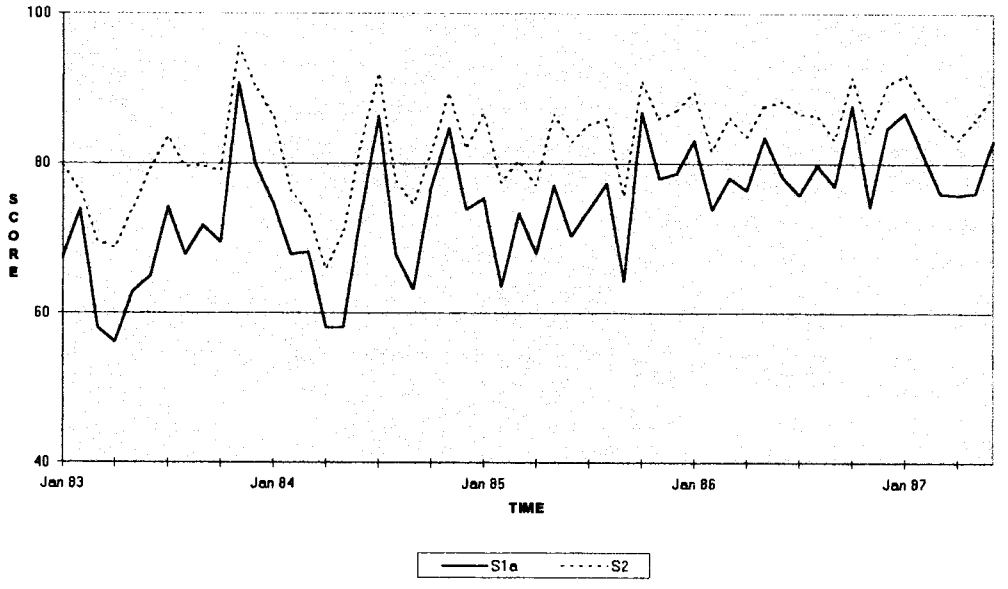

Figure 3. Monthly mean score based on S1a and S2.

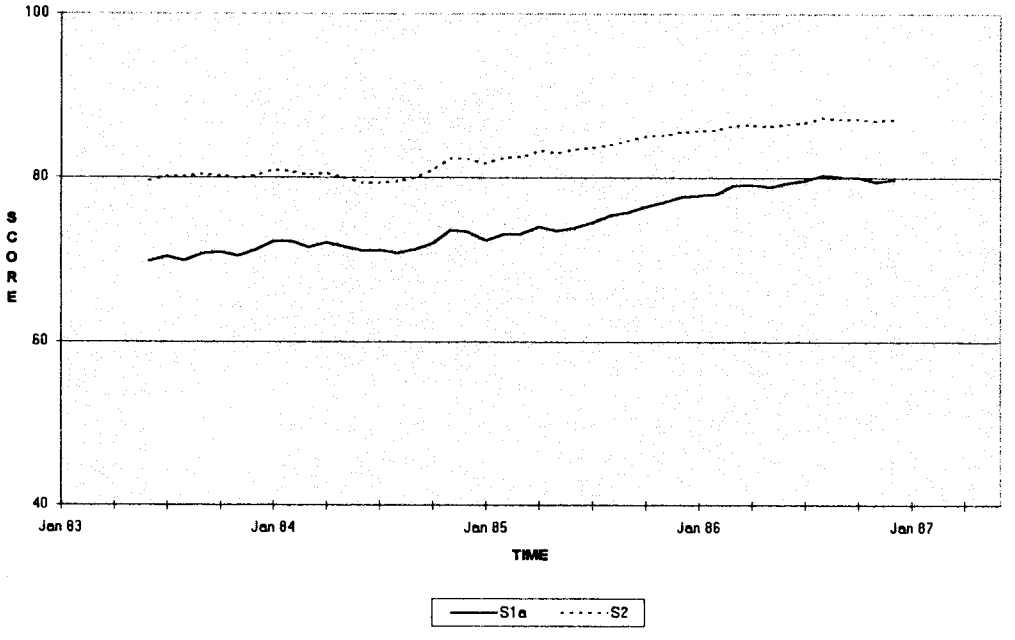

Figure 4. 12-month running average of monthly mean score based on S1a and S2.

The monthly mean percentages of "accurate/acceptable" forecasts (75 as the "accurate/acceptable" mark) based on S1a and S2 are compared in Figure 5. Data are not available for November and December 1984 for S1a and before January 1985 for S2. As the period with continuous data coverage is rather short, plots of 12-month running averages are not included here.

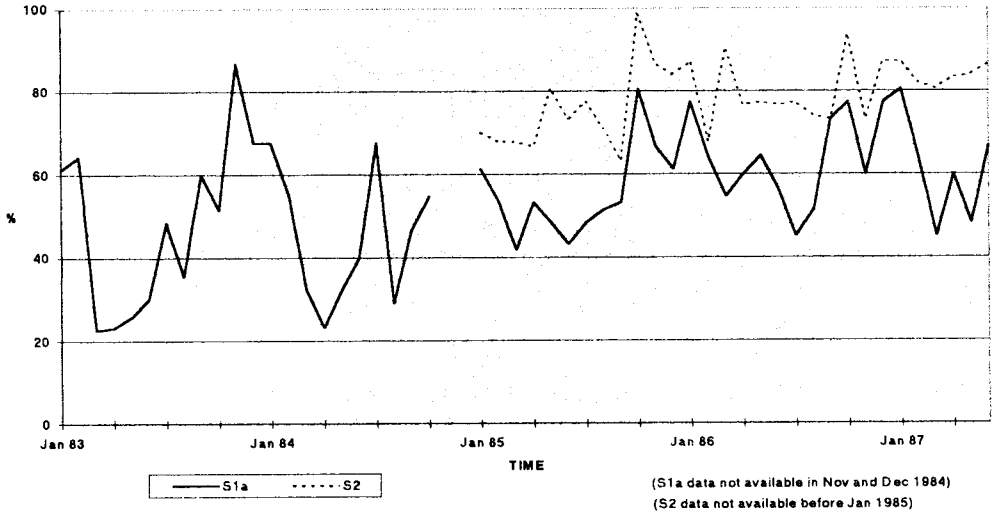

Figure 5. Monthly mean percentage of "accurate/acceptable" forecasts based on S1a and S2.

While scores and percentages of "accurate/acceptable" forecasts for both S1a and S2 basically follow the same rise and fall patterns, those derived from S2 are consistently higher than those from S1a. Thus, S2 is generally speaking more lenient.

Figure 6 and 7, showing time series of monthly mean and 12-month running average (F - P) respectively for S1a and S2, indicate that (F - P) values of S2 are generally less than those of S1a. In general, a rising trend is observed, particularly from the second quarter of 1985 onwards.

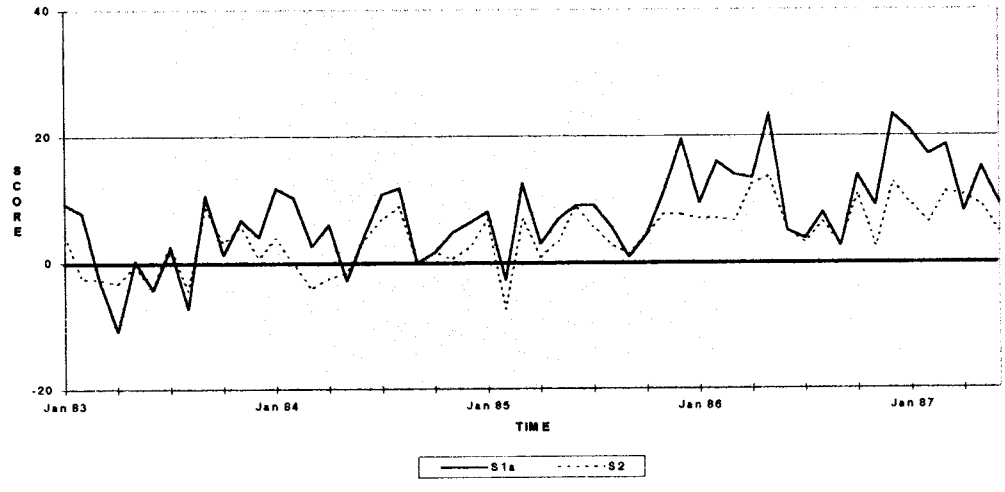

Figure 6. Monthly mean (F - P) based on S1a and S2.

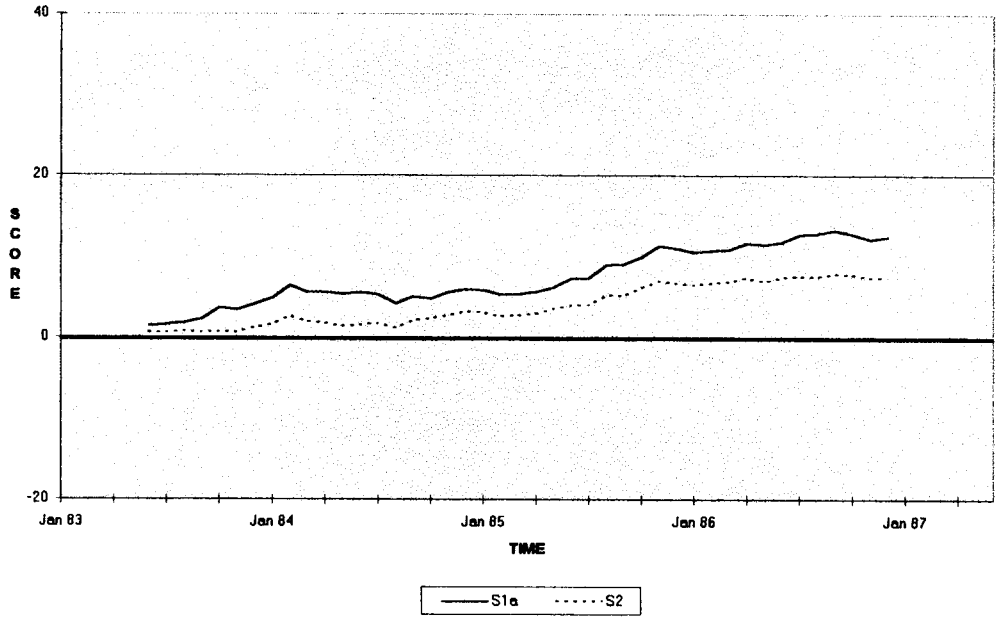

Figure 7. 12-month running average of monthly mean (F - P) based on S1a and S2.

5.2 Scheme 3 (S3)

Scores from S3 are available from January 1985 to March 1986 and these are plotted on Figure 8. As the data coverage is rather short, no trend can be observed.

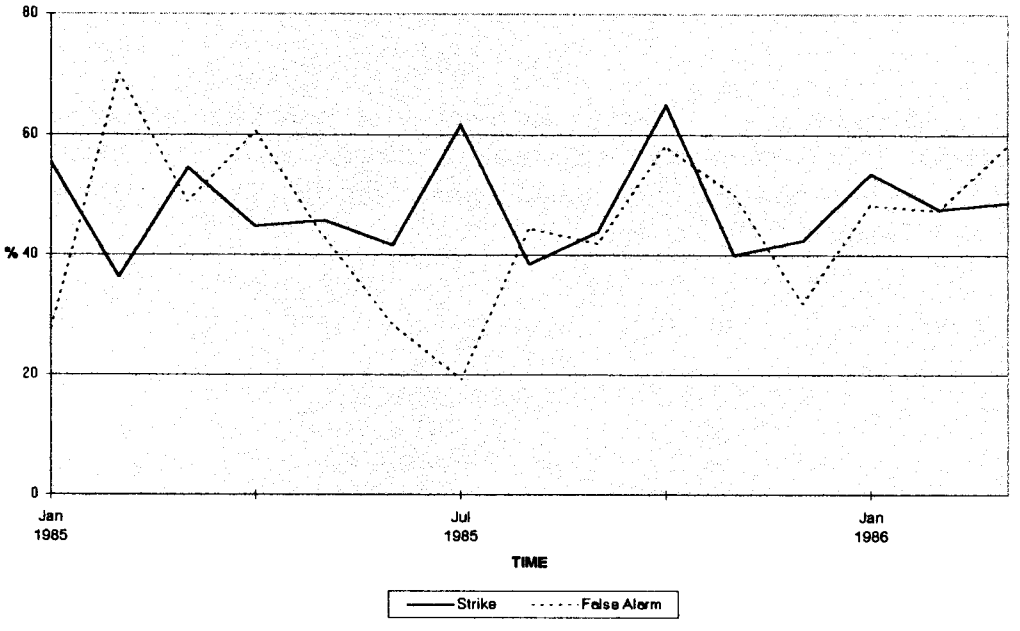

Figure 8. Monthly mean percentage for "Strikes" and "False Alarm" forecasts based on S3.

5.3 Scheme 4

The monthly mean score calculated from S4 for all scheduled public weather forecasts and the 12-month running average are plotted against time on Figure 9. Values of (F - P) are similarly plotted on Figure 10. After a notable rise in 1988, the monthly mean scores and (F - P) appear to be levelling off. Except for a couple of months, (F - P) values are all positive, indicating that the forecasters are able to catch "changes" in weather in most cases.

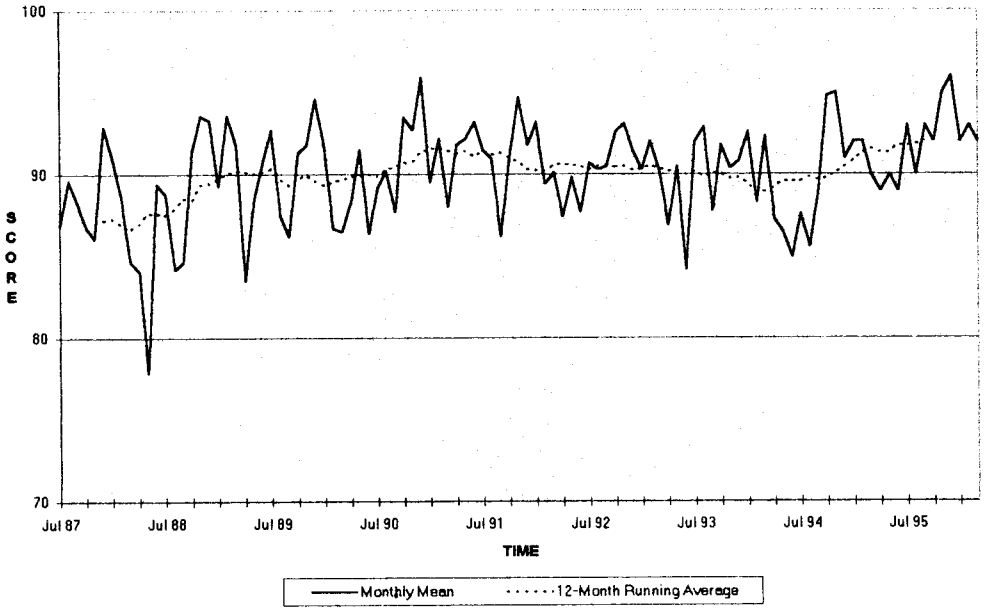

Figure 9. Monthly mean and 12-month running average score based on S4 (all scheduled forecasts).

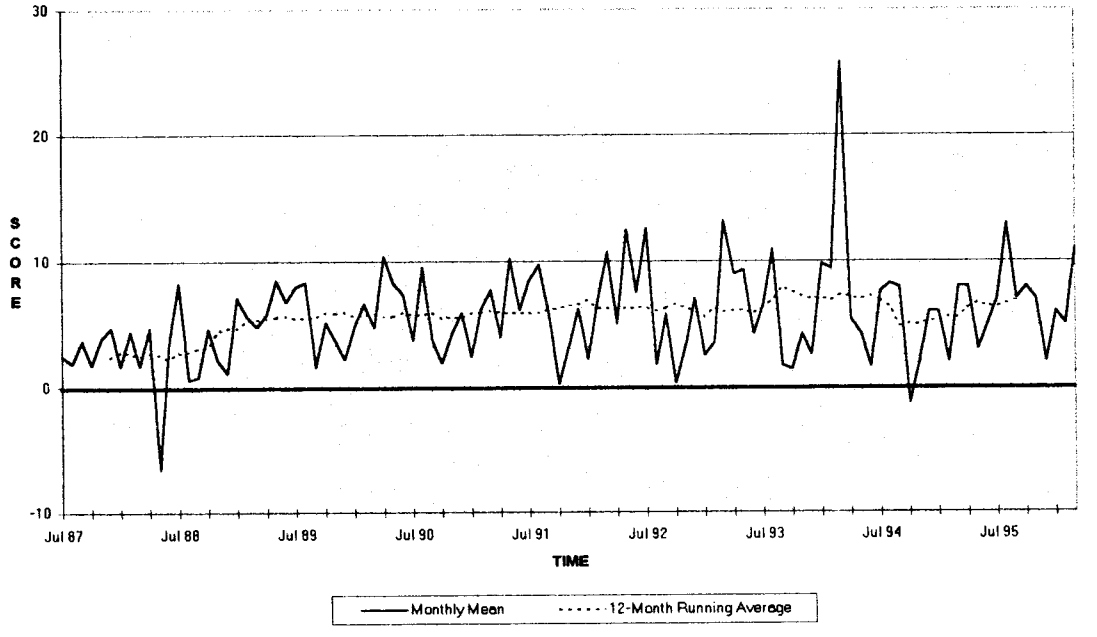

Figure 10. Monthly mean and 12-month running average (F - P) based on S4 (all scheduled forecasts).

A similar rise in 1988 can be observed in Figure 11, which shows time series of monthly mean percentage of "accurate/acceptable" forecasts (80 as the "accurate/acceptable" mark) and 12-month running average. The rise is then followed by a slow decreasing trend starting around mid 1990. A higher threshold, 85, was set for "accurate/acceptable" forecasts since November 1994. The time period after November 1994 is too short to reveal any trend.

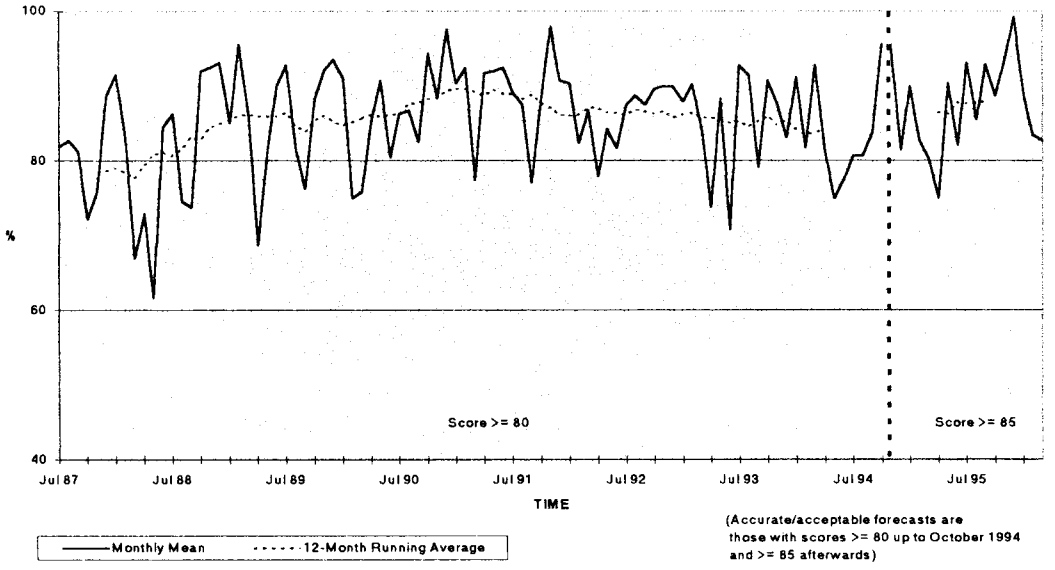

Figure 11. Monthly mean and 12-month running average percentage for "accurate/acceptable" forecasts based on S4 (all scheduled forecasts).

For 3-day forecasts, the monthly mean score and 12-month running average are plotted on Figure 12 and the corresponding (F - P) value on Figure 13. No definite trend can be seen in the years as a whole although (F - P) values show a significant drop around mid 1994.

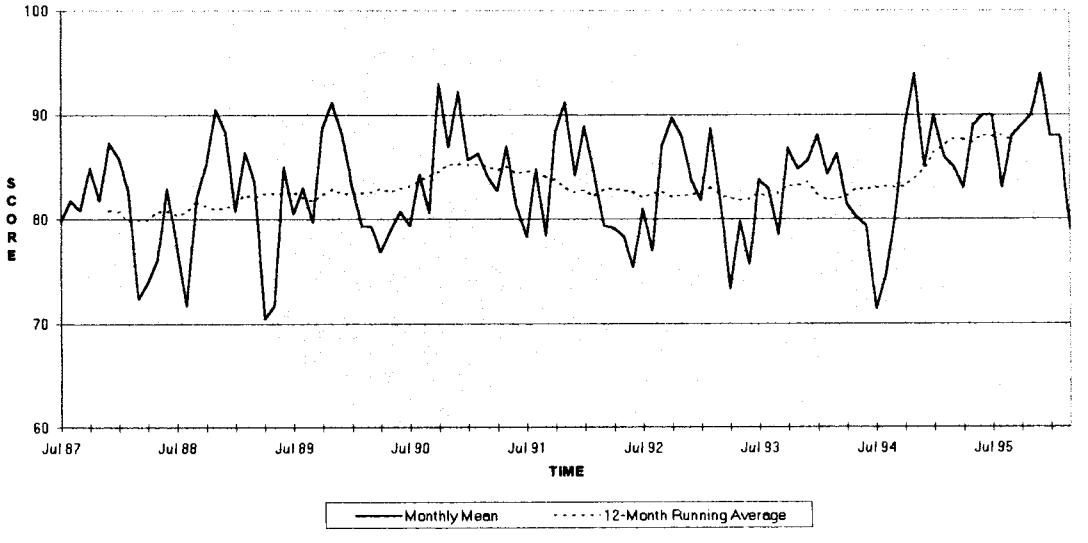

Figure 12. Monthly mean and 12-month running average score for 3-day forecasts based on S4.

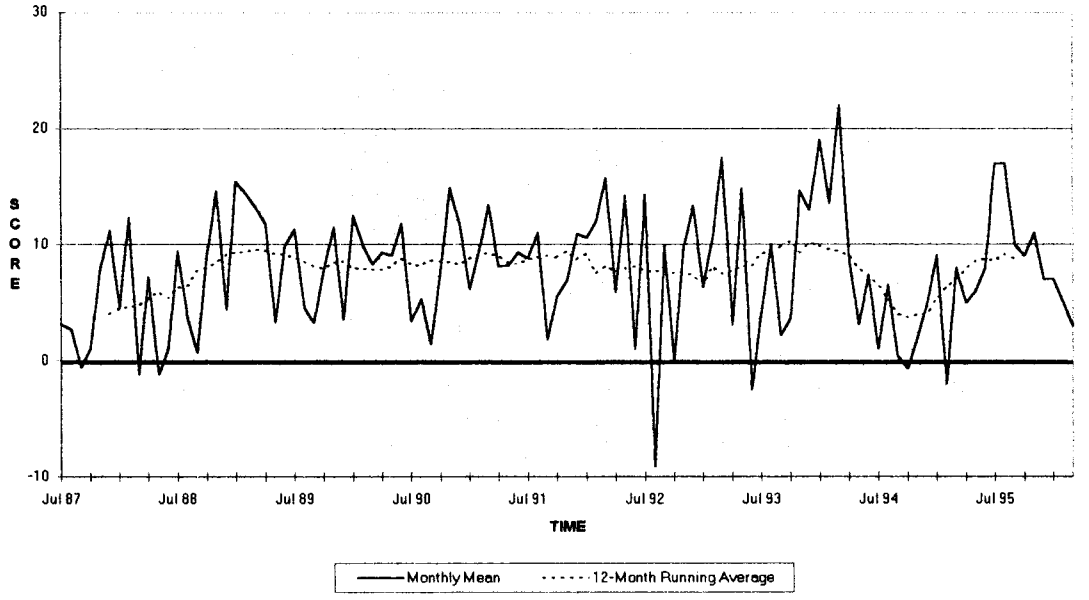

Figure 13. Monthly mean and 12-month running average (F - P) for 3-day forecasts based on S4.

Time series of 12-month running average of monthly mean scores of "Day-1", "Day-2" and "Day-3" forecasts are presented in Figure 14. Corresponding (F - P) values are plotted on Figure 15. Charts of monthly mean scores are not shown because the month-to-month variations are too large to show any trend. Consistent with common understanding, the average monthly scores for Day-1 forecasts are higher than those for Day-2 forecasts, which in turn are higher than those for Day-3 forecasts. On the other hand, Figure 15 shows that Day-2 forecasts outperform Day-1 forecasts in (F - P) values virtually all the time, and that Day-3 forecasts can sometimes beat persistence by as much as 10 marks.

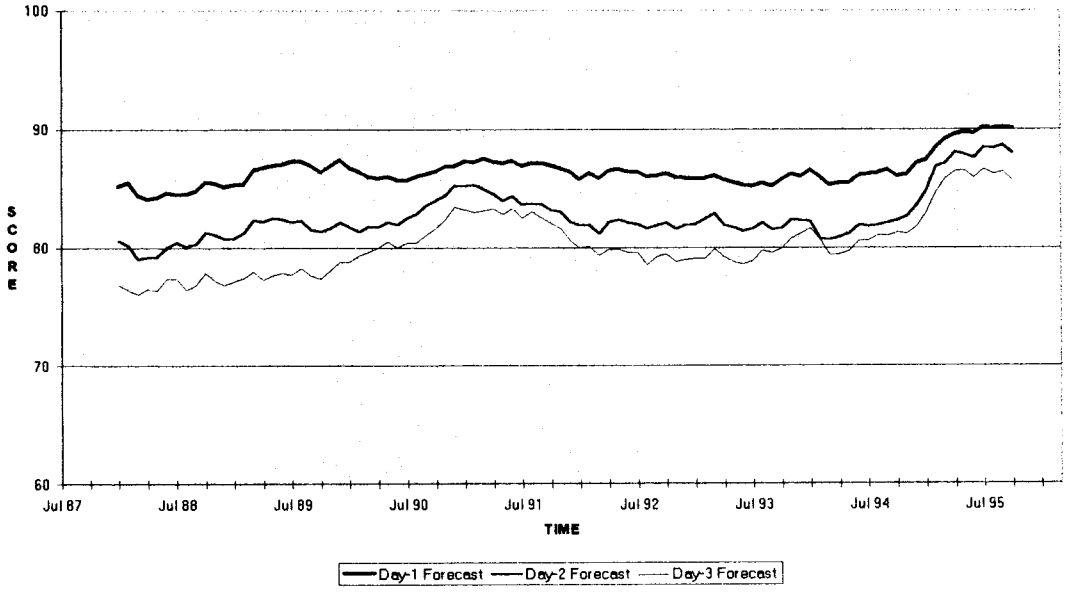

Figure 14. 12-month running average of monthly mean scores for "Day-1", "Day-2" and "Day-3" forecasts based on S4.

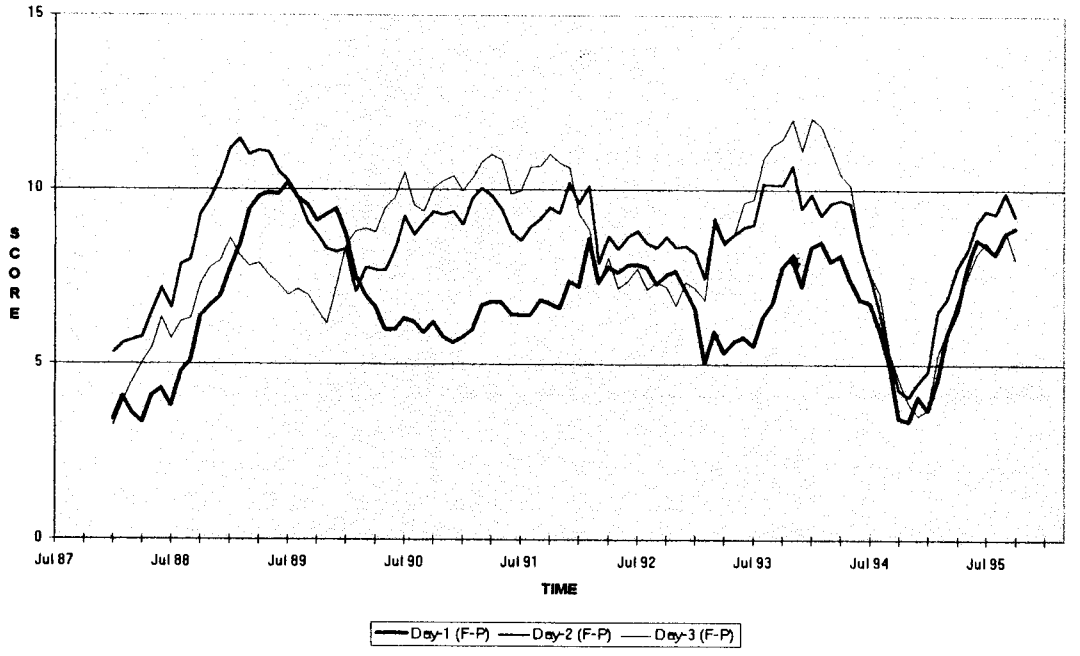

Figure 15. 12-month running average of (F - P) for "Day-1", "Day-2" and "Day-3" forecasts based on S4.

The number of months with (F - P) greater than or equal to a certain threshold, say 5, may also be an indicator of forecasters' skill in predicting changes in weather. Table 2 lists the number of months in a year with (F - P) greater than or equal to 5. After a rise from 1988 to 1989, a levelling off trend can be observed.

Table 2. NUMBER OF MONTHS WITH (F - P) GREATER THAN OR EQUAL TO 5

Year	Number of months with (F - P) $\geq$ 5	
	Public Weather Forecasts	3-day Forecasts
1988	1	5
1989	8	8
1990	6	10
1991	8	11
1992	8	9
1993	5	7
1994	8	7
1995	9	11

To depict the seasonal pattern, Figure 16 shows the average mean scores for each month, composited over the years from 1987 to 1996. The scores are generally lower in the summer months except for July. A similar pattern is also observed from the percentage figures of "accurate/acceptable" forecasts (Figure 17).

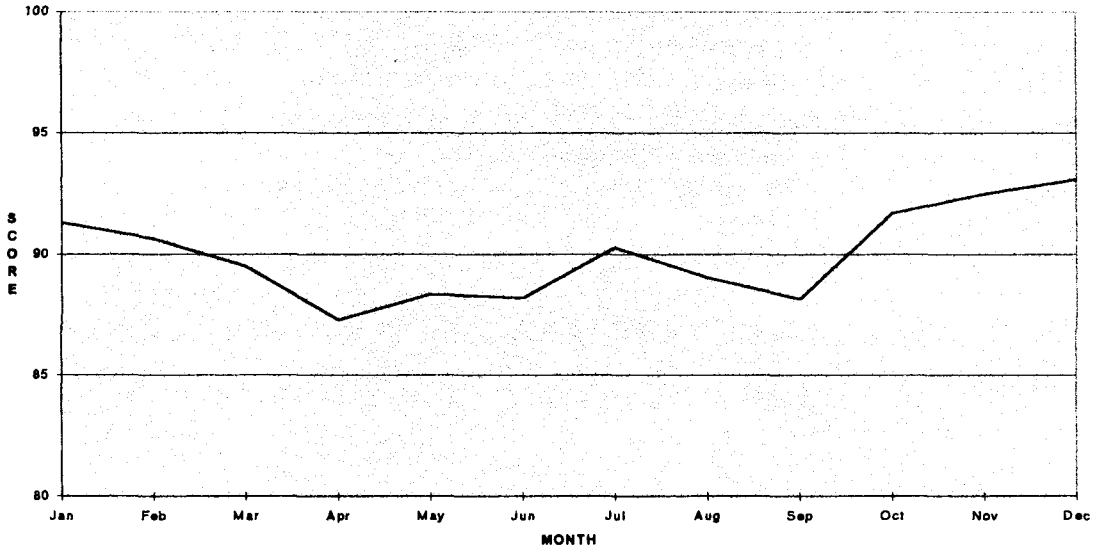

Figure 16. Composite average of monthly mean scores based on S4 for different months from July 1987 to March 1996.

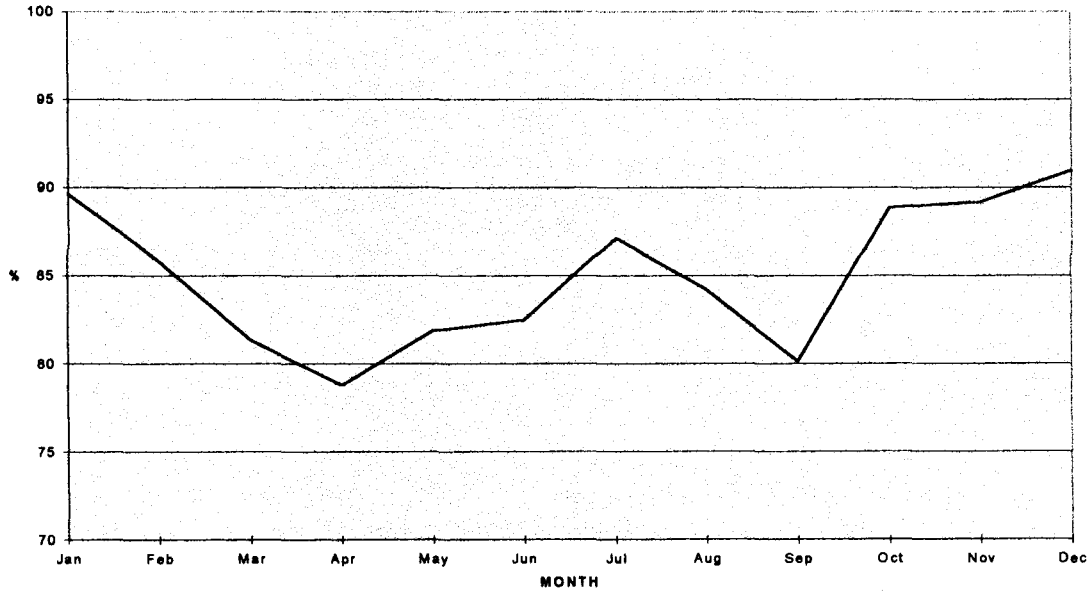

Figure 17. Composite average of monthly mean percentages of "accurate/acceptable" forecasts based on S4 for different months from July 1987 to October 1994 (a new threshold for "accurate/acceptable" forecasts is used commencing November 1994).

6. LONG TERM TREND

To show the trend in scores from the start of the objective forecast verification, monthly mean score and 12-month running average based on S1a and S4 are plotted on Figure 18. Those for (F - P) are shown on Figure 19. For the purpose of comparison, only the daily 1915H/1945H forecast scores based on S4 are used as only the daily 1915H forecasts were verified with S1a.

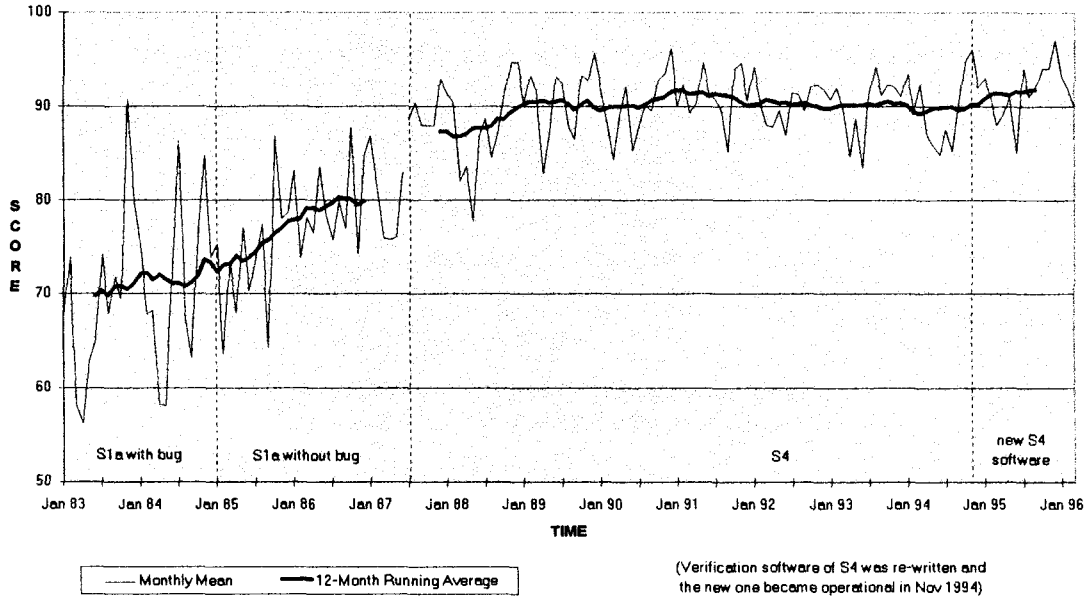

Figure 18. Monthly mean and 12-monthly running average score based on S1a and S4 (only 1915H/1945H forecasts).

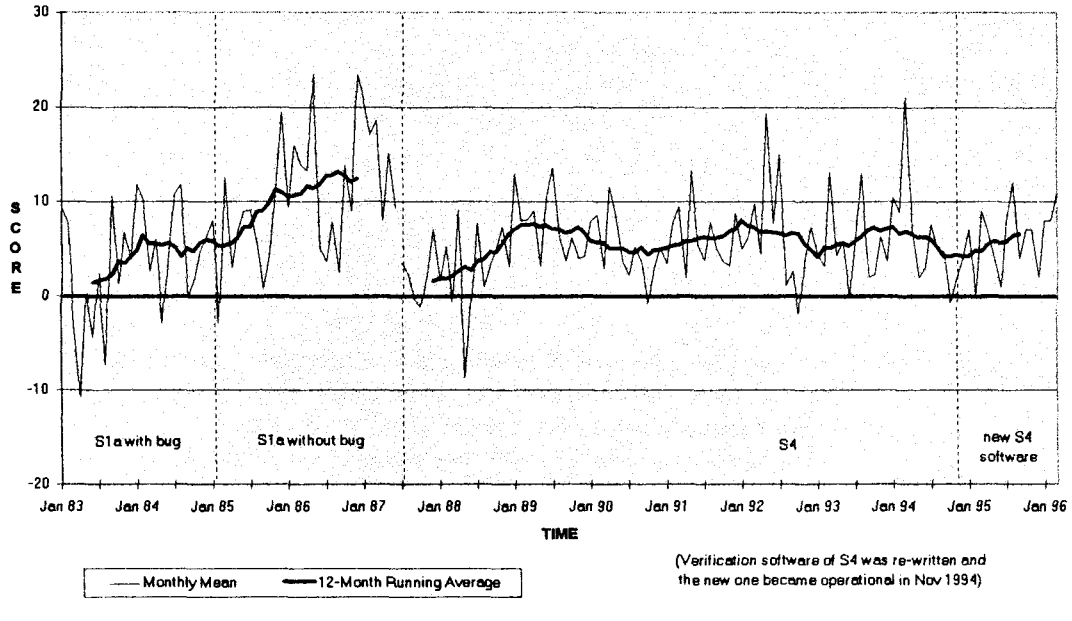

Figure 19. Monthly mean and 12-month running average (F - P) based on S1a and S4 (only 1915/1945H forecasts).

From the two figures, S1a and S4 appear to be significantly different from each other as evidenced by the "step" in mid-1987. Although the scores based on the two schemes cannot be compared directly, the trend of scores gives an indication of the long change in forecasting skill. For both the mean score and (F - P), the sharpest rise was registered in 1985 when numerical model products from the European Centre for Medium Range Weather Forecasts were first used in day-to-day forecasting. Afterwards the skill continued to improve steadily until 1989 when the scores reached a plateau. Then the improvement trend is less evident.

For ease of reference, Table 3 lists the yearly mean scores and the corresponding standard deviations calculated from the monthly mean scores for the period 1983 to1995.

Table 3. YEARLY MEAN SCORES AND CORRESPONDING STANDARD DEVIATIONS

Year*	Mean Score	Standard Deviation	Scheme
1983	70.0/79.6	9.28/7.64	S1a/S2
1984	71.4/79.4	8.89/7.71	S1a/S2
1985	73.9/82.8	6.58/4.15	S1a/S2
1986	79.4/86.6	4.41/3.01	S1a/S2
1988	87.6	4.63	S4
1989	90.1	3.24	S4
1990	90.0	3.08	S4
1991	91.1	2.28	S4
1992	90.5	1.87	S4
1993	90.0	2.50	S4
1994	88.9	3.28	S4
1995	91.8	2.26	S4

Note :

* -- 1987 is not included in the table because S1a and S2 were replaced by S4 in July 1987.

7. OBJECTIVE FORECAST VERIFICATION AND THE PUBLIC OPINION SURVEY

Commencing 1989, independent survey companies are commissioned to conduct public opinion surveys on RO's public weather forecasts. These surveys were carried out four times a year in January, April, July and October from 1989 to 1991. Due to budgetary consideration, starting from 1992 the number of surveys was reduced to two, one in April and another in October.

The key objective of the surveys is to gauge public opinion on the accuracy of weather forecasts issued by the RO. The questions used in the survey cover:

(i) incidence of reading, watching or listening to weather reports,

(ii) sources of weather information (e.g. radio, television, RO's Dial-a-Weather, etc.),

(iii) perceived accuracy of RO's weather forecasts (percent of forecasts that are accurate) over the past three to four months,

(iv) perceived accuracy of RO's weather forecasts over the past three months in terms of temperature, state of sky as well as typhoon forecast and warning, and

(v) overall assessment of the accuracy of weather forecasts nowadays as compared with that in the past few years.

The answer to (iii) above gives the average percentage of forecasts that the public considered accurate in the past three to four months. The figures represent the percentage of "accurate/acceptable" forecasts as perceived subjectively by the public (hereafter referred to as "survey result").

For the purpose of comparing with the survey results, average S4 scores of the previous four months of the month when public opinion surveys were conducted are calculated (hereafter referred to as "objective score"). If the survey is conducted in October, the average of monthly mean scores of July to October is computed. 4-month average percentages of "accurate/acceptable" forecasts based on S4 (hereafter referred to as "objective percentage") are similarly obtained. It should be noted that a new threshold for objective percentages was used since November 1994.

Time series of the survey result, objective score and objective percentage are plotted on Figure 20. The correlation coefficient (based on the least square method) between the survey result and the objective score is -0.007, while that between the survey result and the objective percentage is -0.161, indicating little or no correlation between the survey results and any of the two objective verification results.

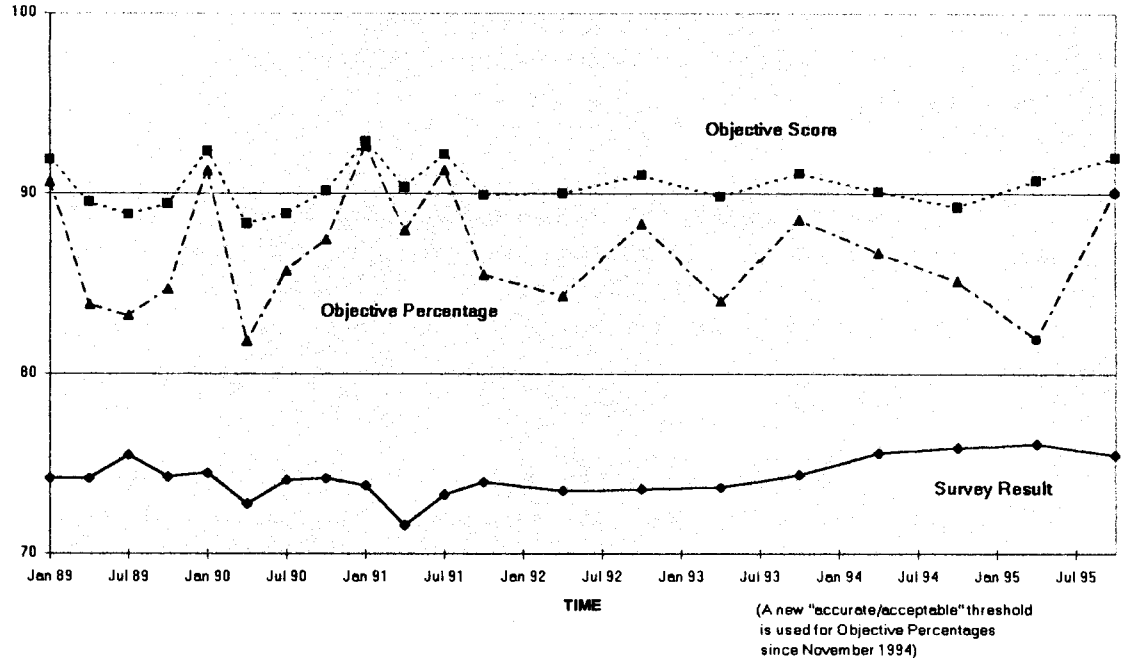

Figure 20. Objective score, objective percentage and survey result.

8. CONCLUSION

Objective verification schemes are used in the RO to verify weather forecasts since 1984. Weather forecasts and the actual weather are coded and then compared objectively to give a score for each of the forecasts. A forecast is considered "accurate/acceptable" if its score is equal to or higher than a pre-defined "accurate/acceptable" mark. These scores are also compared with the scores of persistence forecasts to give (F - P) values, a measure of forecaster's skill in predicting changes in weather.

Altogether there were four schemes in the history of RO's objective forecast verification. The first scheme was devised in 1984 and the second, a variation of the first, in 1985. The third, operated from January 1985 to March 1986, was specially designed to see whether forecasters could "catch" changes in the weather.

With experience in forecast verification and better computer facilities, the fourth scheme was devised with more sophisticated algorithms to better assess the accuracy of the weather forecasts. A total of four minor modifications to the scheme were made so far.

The accuracy of RO's public weather forecasts, reflected by the scores from objective verification, showed a sharp rise in 1985 when the products from European Centre for Medium Range Weather Forecasting were first used in the CFO for day-to-day weather forecasts. The scores then levelled off in the 1990s when the improvement trend became less evident.

Since 1989, the RO have employed independent survey companies to carry out public opinion surveys on the accuracy of RO's weather forecasts. Little or no correlation is observed between the results of the public opinion surveys and those of the objective verification.

ACKNOWLEDGMENT

The author would like to thank Dr. H.K. Lam, Mr. C.Y. Lam, Dr. B.Y. Lee and Mr. P.W. Ng for their valuable comments and contributions in the preparation of this report. The effort of Mr. S.T. Lai in reviewing the draft as well as the Development, Research and Administration Branch in re-computing some of the verification results are much appreciated.

APPENDIX I

VERIFICATION OF STATE OF SKY IN SCHEME 4

The state of sky is broken into two components, namely sunshine duration and cloud amount, for verification in Scheme 4 according to the following table:

Forecast Category	Sunshine duration (%) at King's Park within forecast period	Mean cloud amount (oktas) at RO within forecast period
1. Overcast	0	7.6 - 8
2. Cloudy	0 - 5	6.1 - 7.5
3. Bright	5.1 - 10	6.1 - 7.5
4. Mainly fine	10.1 - 50	0 - 6
5. Fine/sunny/clear	50.1 and above	0 - 6

A state of sky forecast will get 100 marks if the actual sunshine duration (%) and mean cloud amount (oktas) fall within the ranges of the corresponding category. Marks will be deducted according to a continuous curve if the actual sunshine duration and cloud amount are outside those ranges.

The final score is given by the arithmetic mean of the scores on sunshine duration and cloud amount during periods of non-zero available sunshine, else the score is given by cloud amount alone. For a forecast covering two periods, i.e. one with available sunshine and the other without, the final score is given by the weighted mean of the scores of the two periods according to their respective lengths.

APPENDIX II

SCORE TABLE FOR VISIBILITY AND WEIGHTING TABLE OF SCHEME 1a AND SCHEME 2

(a) SCORE TABLE FOR VERIFICATION OF VISIBILITY

Scheme 1a

Forecast	Actual visibility reports	
	Mist or Fog [1]	No mist or fog [2]
Mist or Fog	100	0
Not forecast	0	100

Scheme 2

Forecast	Actual visibility reports		
	Fog [3]	Mist [4]	No mist or fog [2]
Fog	100	80	0
Mist	80	100	50
Not forecast	0	50	100

Note: 1 -- Fog or mist was reported at any of Walgan, Cheung Chau, RO and Airport Meteorological Office.

2 -- No fog nor mist was reported at all four stations.

3 -- Fog was reported at any of the four stations.

4 -- Only mist, but no fog, was reported at any of the four stations.

APPENDIX II (cont.)

(b) WEIGHTING TABLES

Scheme 1a

Element	Month											
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Winds	15	10	5	5	10	20	20	20	20	20	20	15
State of sky	15	10	5	5	10	10	15	15	20	20	20	20
Precipitation	20	20	30	40	60	60	60	60	50	30	20	20
Temperature	50	50	30	20	20	10	5	5	10	30	40	45
Visibility	0	10	30	30	0	0	0	0	0	0	0	0

Scheme 2

Element	Month											
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Winds	20	10	10	10	20	20	20	25	25	25	20	20
State of sky	20	20	20	20	20	20	20	20	20	20	20	20
Precipitation	20	20	20	20	30	30	30	30	30	25	20	20
Temperature	30	30	20	20	20	20	20	20	20	20	30	30
Visibility	10	20	30	30	10	10	10	5	5	10	10	10

APPENDIX III

A BRIEF ACCOUNT OF THE FOUR MODIFICATIONS TO SCHEME 4

1. Modification I

This modification was proposed in September 1988. Only part of the proposal was implemented and the changes are summarised below:

(a) Printouts for scores of forecasts

The format of the printout was modified and additional information was included to facilitate interpretation of the data presented.

(b) Verification of rainfall forecast

Before the modification was made, if the forecast duration (t) of rainfall was different from 24 hours, the total rainfall amount recorded in that time interval was multiplied by the factor "24/t" to obtain a "normalised" daily total for verification.

It was noted that even in a typical rainy day, there would only be a few hours of rainfall. So, in forecasting rain for only a few hours of rainfall (e.g. "isolated showers in the morning"), the forecaster might have in mind similar rainfall amount as when forecasting rainfall for a whole day ("isolated showers tomorrow"). The factor to obtain the normalised daily total was changed to:

$$1 \quad \text{if} \quad T \geq 6$$

$$\frac{6}{T} \quad \text{if} \quad 6 > T \geq 1$$

where T is the number of hours that rainfall was recorded during the forecast period and should have a minimum value of 1.

2. Modification II

The modification was introduced as a result of the termination of manual observations at Waglan and the introduction of the Waglan automatic weather station (AWS). The modified scheme became operational on 1 November 1989. The changes made are summarised below:

(a) Verification of winds

3-hourly AWS observations of hourly mean winds and gusts were used to replace the manual observations.

(b) Verification of rain report

Rain is considered to be reported at Waglan only if

(i) Rainfall is measured by the AWS (i.e. 0.5 millimetre or more), or

(ii) Rainfall is recorded at RO, Hong Kong International Airport or Cheung Chau.

Condition (ii) is to cater for light rain situations which would reduce the visibility.

3. Modification III

The modification was a result of the automation of the weather station at Cheung Chau. Changes similar to those of Modification II for Waglan were made. The modified scheme became operational in February 1993.

4. Modification IV

The modification was made in March 1996. An additional category, namely haze, was included in verifying visibility forecasts. A forecast of haze will get 100 marks if the reported visibility is reduced to less than 5 000 metres by small dry particles and 0 mark if the reported visibility is ten kilometres or more. A linear scale between five and ten kilometres will award marks to the forecast depending on the reported visibility at the RO.