Technical Report 208

ESTIMATING CONDITIONAL PROBABILITY AND PERSISTENCE

By
John T. McCabe
Colonel, USAF

PUBLISHED BY
AIR WEATHER SERVICE (MAC)
UNITED STATES AIR FORCE
JUNE 1968

CHANGE

October 1968 Technical Report 208

ESTIMATING CONDITIONAL PROBABILITY AND PERSISTENCE

Technical Report 208, June 1968, is changed as follows:

1. Make the following pen and ink change:

Page	Action to be Taken
2	Change Equation (2) to read $r_{xy}^2 = 1 - \left[2\sigma_a^2\sigma_b^2 / \left(\sigma_a^2 + \sigma_b^2\right)\right]$

2. After necessary action, file this change in front of page i, PREFACE.

JOHN T. McCABE
Colonel, USAF
USAF ETAC
Washington, D. C. 20333

DISTRIBUTION	
Hq AWS.	50
Wings	5
(except 7 Wea Wg)	17
USAF ETAC	25
Squadrons	2
Detachments	1
Special	

PREFACE

The Air Weather Service supports an ever-changing military community in a greater and greater portion of the natural environment. Its services are tailored to customer needs and many of its efforts truly pioneer aerospace science applications in new fields and in parts of the world long neglected by weathermen. As a result, AWS forecasters, as individuals, often find themselves with only a minimum of experience in forecasting for their local areas and depend heavily on information, methods, and techniques developed by their predecessors. They also have at their disposal, to varying degrees, a history of past weather in such forms as map files, observation records, and climatological summaries. One of the most useful of the climatic aids to local forecasting is the set of conditional probability tables which describes how, in the past, the weather behaved subsequent to certain initial conditions. With these tables, a forecaster is able to isolate those past cases similar to his own present weather state and examine certain aspects of what previously had followed.

There are many types of conditional probability tables which are usually computer-generated from historical data. The data base needed to prepare such tables is of the order of at least ten years of hourly observations. Conditional probability statistics generated from a lesser data base may suffer from small sample size, especially for event categories that occur rather infrequently.

This report describes a statistical method for generating estimates of conditional (and persistence) probability information from unconditional probability statistics. The unconditional statistics are less affected by short or incomplete periods of records, and often can themselves be reliably estimated. The method can be used manually; however, it has been programmed for the IBM 7044 and is being used to support the AWS mission at the USAF Environmental Technical Applications Center (ETAC) where basic input data for most locations are generally available or can be derived.

AWS units that may have a need for the conditional (or persistence) probability programs or their output products are invited to contact ETAC through appropriate channels, normally their squadron and wing aerospace sciences or technical services office.

Technical Report 208 June 1968

JOHN T. McCABE
Colonel, USAF
Hq USAF ETAC
Washington, D. C. 20333
14 June 1968

DISTRIBUTION:

Hq AWS	50
Wings	5
except 7 Wea Wg	17
Hq USAF ETAC	25
Squadrons	2
Forecasting Detachments	1
Special	

An earlier edition of this report was published in the "Proceedings of the First Statistical Meteorological Conference," hosted by Travelers Research Center, Inc., at Hartford, Connecticut, May 27-29, 1968.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

ESTIMATING CONDITIONAL PROBABILITY AND PERSISTENCE

SECTION A — ESTIMATING CONDITIONAL PROBABILITY

As used here, conditional probability is defined as the probability that a particular event will occur at a given lag time after the occurrence of some "initial condition." The initial conditions and subsequent events may be the same or they may differ.

Most climatological summaries of conditional probability cover a complete range of combinations of specified event categories. For example, if total cloud cover is classified according to four categories; (a) clear, (b) scattered, (c) broken, and (d) overcast; then conditional probabilities would generally be given for each of the initial conditions (a, b, c, d) paired with the subsequent occurrence of each of the four conditions. Because of the marked diurnal and seasonal variations in frequency of occurrence of meteorological events, conditional probability tables are usually prepared for initial conditions as a function of time of day and month or season. Within Air Weather Service most conditional probability work has involved categories of combined ceiling and visibility. However, some tables have been prepared for ceiling categories alone, visibility alone, total cloud cover, and precipitation. Extensions and refinements of the basic conditional probability idea have been made by considering "trends" prior to the initial time, the time of onset of the initial conditions, and various combinations of additional parameters (e.g., wind speed/direction, dew-point depression, and the presence or absence of precipitation) at the initial time. Additional information on conditional-persistence summaries is available in 4th Weather Wing Technical Papers 66-1 and 67-1 [1] [2].

Invariably, one of the primary requirements for preparation of conditional probability statistics is a large data base. Consider, for example, the four-category classification prepared as a function of hour of the day and month. Over a ten-year period, a 30-day month would provide 300 sequences of events having initial conditions at a given hour of the day. The four initial categories would have an average of 75 occurrences each, and each of these, when paired with the four subsequent categories, would have an average of less than 20 occurrences over the ten-year period. Of course, the rare or relatively infrequent event would have considerably fewer occurrences on which to base a pattern of behavior. Also, in this type of summary, observations are paired (by the nature of the conditional process); each observation being paired with all others over the range of lag times being considered. When an observation is missing, all of its paired combinations are also lost. Therefore, missing observations compound the limitations imposed by the data sample size.

In an effort to circumvent the restrictions imposed by data requirements, climatologists have for some time sought to establish statistical models of various weather parameters of interest. For example, the direction-speed frequency distributions of winds have been modeled after the circular and elliptical distributions [3] [4] and Gringorten of the Air Force Cambridge Research Laboratories has modeled the duration of certain meteorological events after a simple Markov process [5]. In climatology, an acceptable model usually provides a means of determining a great deal of information about a parameter (or combination of parameters) from only the few statistics needed to describe the model. Often these statistics can be derived from relatively small data samples or estimated from mapped or graphed values.

The method described below parallels Gringorten's work in some respects but differs in that it is concerned primarily with conditional probability; here, persistence (or duration) is a secondary consideration. This method also takes into account the diurnal variability of the parameter.

The Elliptical Distributions

The notation (N, 0, 1) is used to identify the standard normal distribution having zero mean and unit standard deviation. Two (N, 0, 1) variables (x, y) may be graphed orthogonally and, if they are uncorrelated, their joint distribution is customarily referred to as the circular normal distribution because the contours of constant probability density are circles. If the (N, 0, 1) variables are correlated, their joint distribution is elliptical with axes 45° to the x and y axes. If the variables are positively correlated, the major axis will lie between the like-sign x and y axes; if negatively correlated, between the unlike-sign x and y axes. Figure 1 is a graphical presentation of the joint x, y distribution for an uncorrelated (circular) case and a positive correlated (elliptical) case. The remainder of this report considers only cases of positive correlation ($r_{xy} \leq 0$).

The correlation coefficient, given by

$$(1) \qquad r_{xy} = \sqrt{1 - S_y^2/\sigma_y^2}$$

where S_y is the standard error of estimate and σ_y, the standard deviation of y, is related to the elliptical parameters by:

$$(2) \qquad r_{xy}^2 = 1 - \left[2\sigma_a^2\sigma_b^2 \Big/ \left(\sigma_a^2 + \sigma_b^2\right)\right]$$

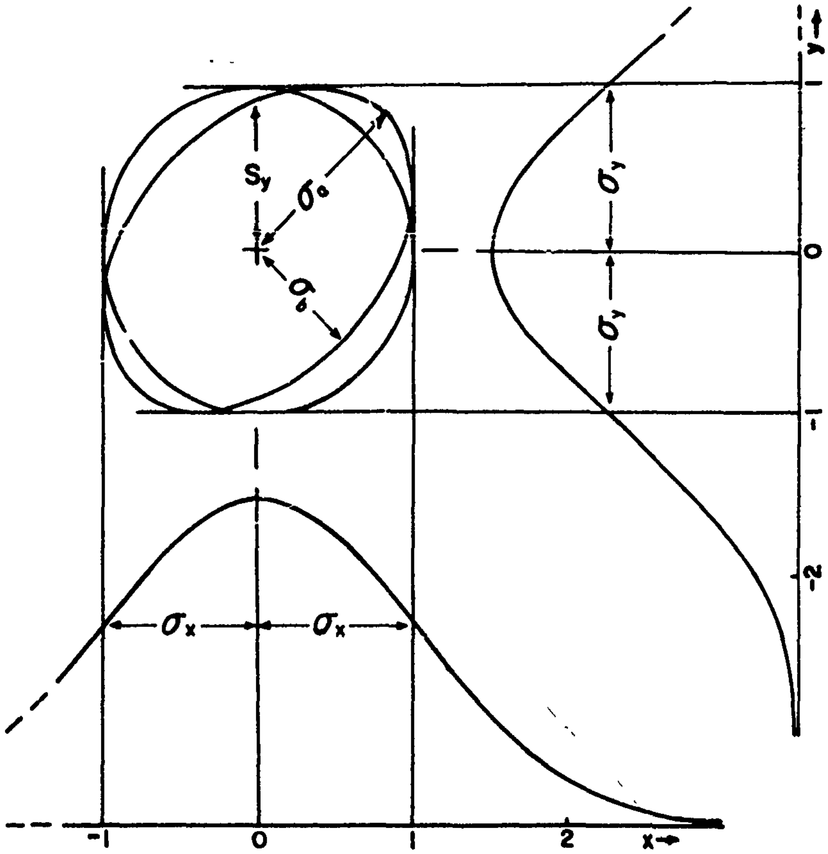

Figure 1. Example of Joint (N, 0, 1) Distributions; Uncorrelated (Circular) and Correlated (Elliptical) Cases.

where σ_a and σ_b are the standard deviations along the major and minor axes, respectively (see Figure 1).

Because x and y are (N, 0, 1) variables,

$$(3) \qquad \left(\sigma_a^2 + \sigma_b^2\right)/2 = \sigma_y^2 = \sigma_x^2 = 1$$

and

$$(4) \qquad r_{xy}^2 = 1 - \sigma_a^2 \sigma_b^2$$

Thus, for the positive correlation case with $\sigma_a \geq \sigma_b$,

$$(5) \qquad r_{xy} = \sigma_a^2 - 1 = 1 - \sigma_b^2$$

Conditional probabilities of the two (N, 0, 1) variables of known correlation can be determined from tables of the elliptical normal distribution [6]. For any interval of x and y, the conditional probability of $Y_1 \leq y \leq Y_2$ given $X_1 \leq x \leq X_2$ is equal to the joint probability that $Y_1 \leq y \leq Y_2$ and $X_1 \leq x \leq X_2$ (obtainable from the known elliptical distribution), divided by the probability that $X_1 \leq x \leq X_2$ (obtainable from the standard normal distribution). This is shown graphically in Figure 2 where

$$(6) \qquad P(y|x) = \frac{P(x,y)}{P(x)} = \frac{\text{Probability area E}}{\text{Sum of probability areas B, E and H}}$$

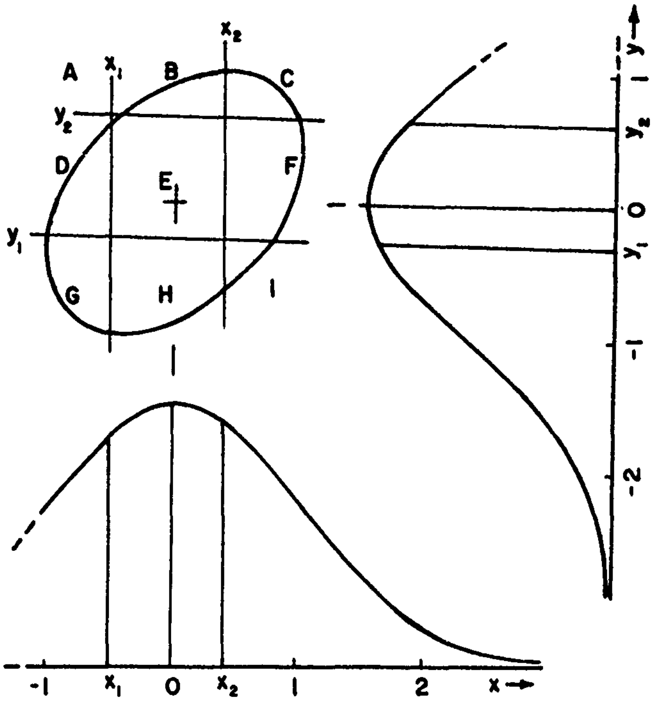

Figure 2. Elliptical Distribution for $r_{xy} = 0.4$.

Elliptical to Circular Transformation

For the purpose of determining conditional probabilities of the above types, i.e., where X_1, X_2, Y_1, Y_2 are constants, one can relate the elliptical distribution to an equivalent projection of the circular distribution rotated an angular amount (θ) about a diameter. The equivalent circular distribution will differ from the elliptical distribution in the orientation of the x and y axes which, having been orthogonal in the elliptical distributions, now form an angle of $90° + 2\alpha$ to one another; where, in degrees,

$$(7) \qquad \alpha = \mathrm{Tan}^{-1}\,(1/\mathrm{Cos}\ \theta) - 45^\circ$$

But

$$(8) \qquad \mathrm{Cos}\ \theta = \sigma_b/\sigma_a$$

and using Equation (5), it develops that

$$(9) \qquad \alpha = \mathrm{Tan}^{-1}\,\sqrt{(1 + r)/(1 - r)} - 45^\circ$$

Figure 2 depicts the standard elliptical normal distribution for $r_{xy} = 0.4$ and Figure 3, its equivalent circular normal distribution with transformed x, y axes ($\alpha \approx 12^\circ$).

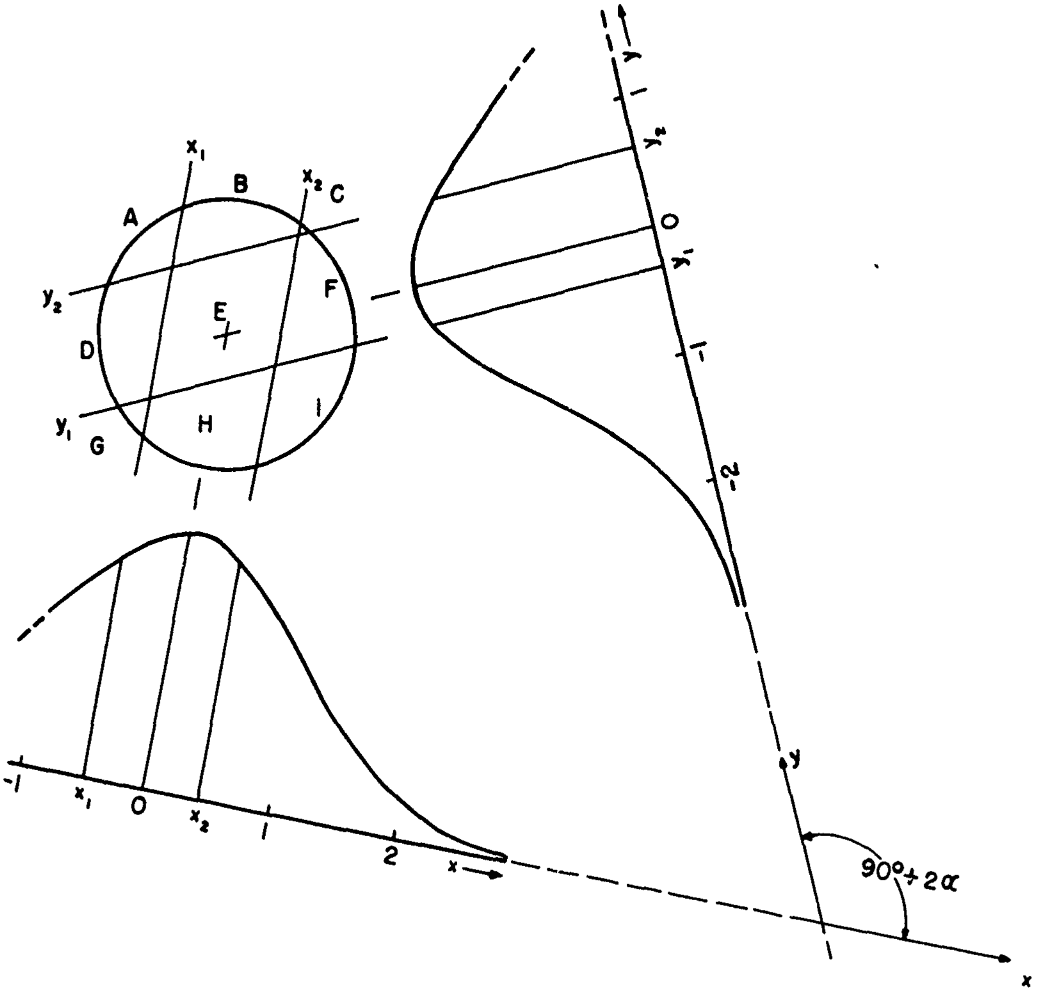

Figure 3. Circular Normal Distribution with x, y Axes Transformed ($r_{xy} = 0.4$).

Use of the Mil Diagram

Conditional probabilities can be determined graphically from the circular-transformed distribution with the aid of a circular normal frequency diagram or plot. Figure 4 is such a diagram. It is constructed so as to provide frequency (probability) in mils (1 mil = .001) equal to the number of mil-areas contained in any portion of the distribution. Thus, using the transformed x, y coordinates to define the "joint-probability" portion of the distribution, one obtains the "y given x" conditional probability, $P(y|x)$, by dividing this joint probability, $P(x,y)$, by the unconditional probability, $P(x)$.

Correlation Estimates

In the last ten years, the USAF Environmental Technical Applications Center (ETAC) and its predecessor, the USAF Climatic Center, have prepared conditional probability summaries for hundreds of Air Force locations. Summaries have generally been made for a variety of ceiling-visibility categories prepared from hourly observations with periods of record ten years or greater. Plots of conditional probability versus unconditional probability at initial time and at lag time, and as a function of the length of lag period have been made from these summaries. The plots reveal a pattern indicating a regular decrease of correlation with increasing lag time. Correlation values have been computed from these plots using a technique which is the reverse of the method described above. That is, elliptical distributions were determined which best conformed to the conditional versus unconditional probabilities, and the correlation coefficients determined from the elliptical parameters.

Figure 5 shows the average relationship between correlation and lag period computed from several conditional summaries. Also shown are the correlation curves for an assumed Markov process with 0.94 and 0.95 one-hour correlations. With a Markov process, the correlation for an n-hour lag period equals the one-hour correlation raised to the n^{th} power.

Automated Method

Conditional probabilities, as described above, can be computer-calculated by referring to stored elliptical distribution tables, once the elliptical parameters and joint-probability boundaries are determined. Instead, a program has been written which directly parallels the previously described manual method. In this computer technique, each mil area of the mil diagram is identified as an i,j point near the center of the mil area. The 1000 mil areas are thus represented by 1000 points of known i,j location. Equations are determined for the lines which bound the joint probability area of interest, and each of the 1000 mil points is tested to determine if it falls within

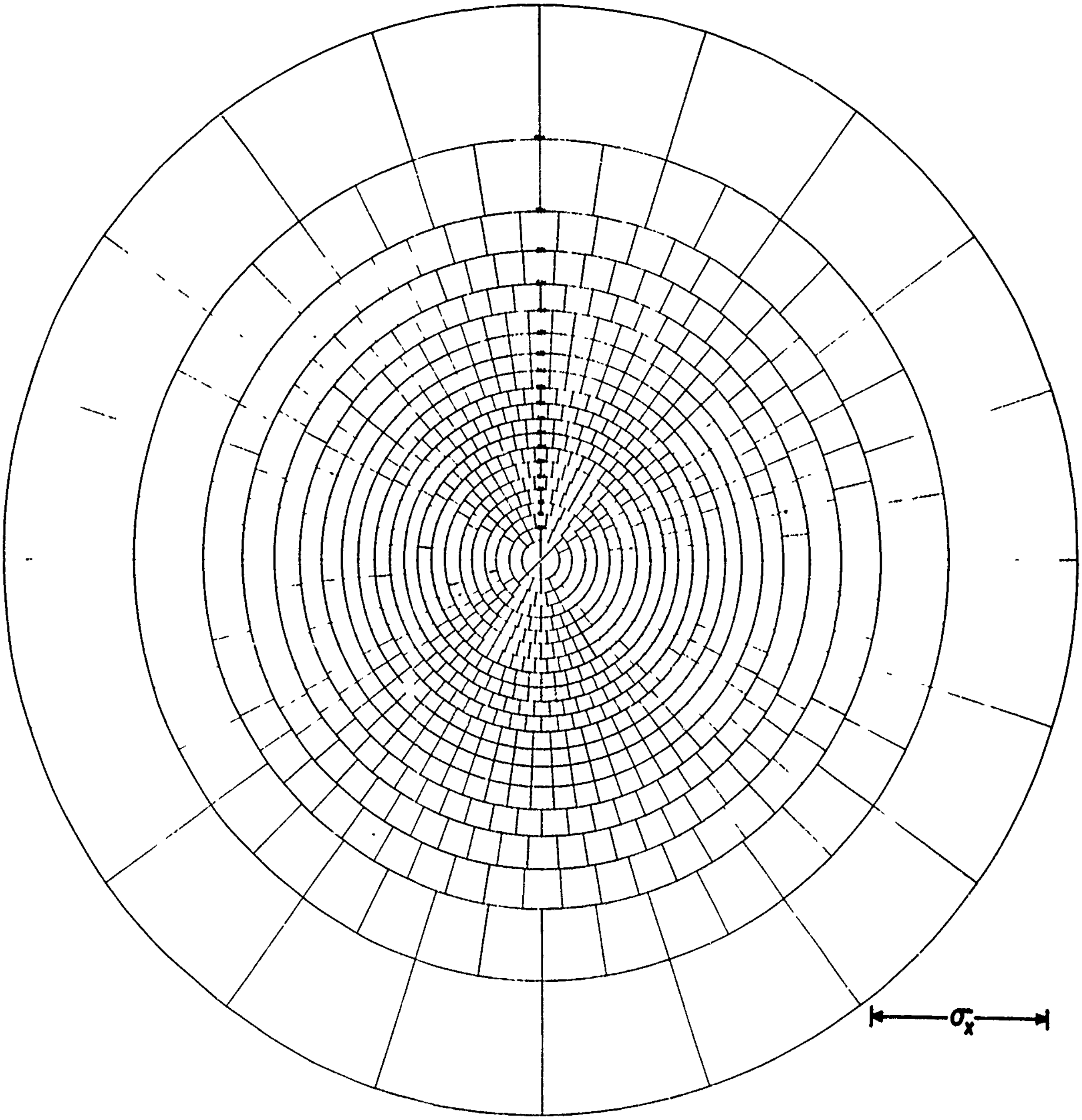

Figure 4. Circular Normal Mil Frequency Diagram.

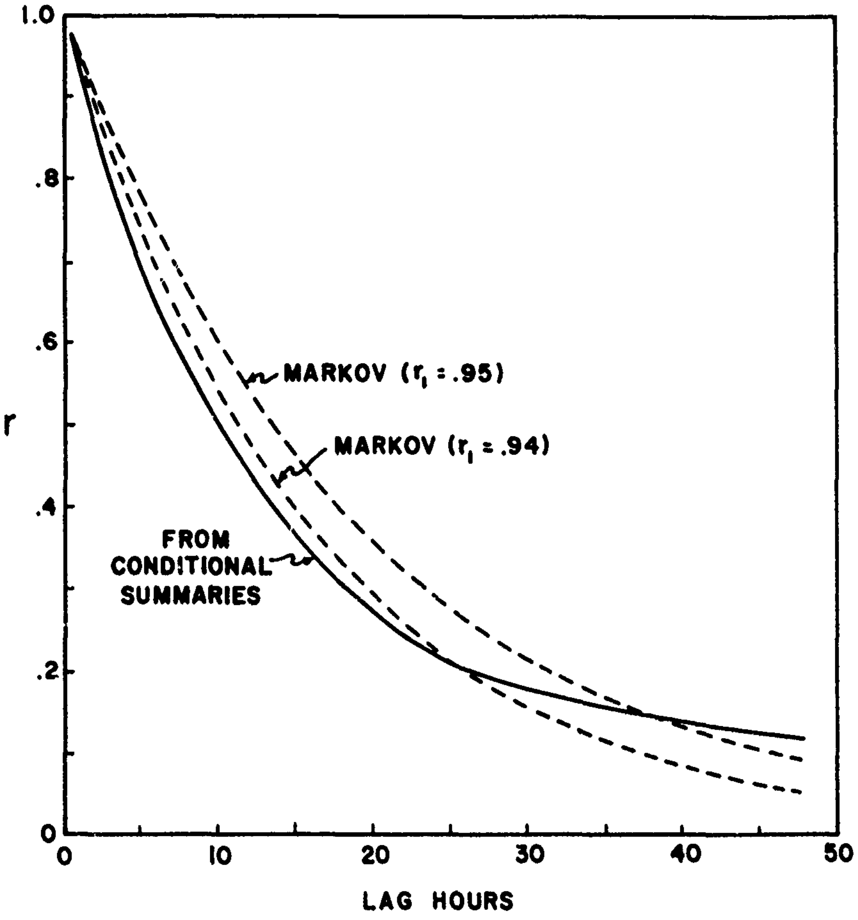

Figure 5. Correlation vs Lag Hours.

or outside of the joint-probability area. The computer program routine permits each point to be tested against as many areas as desired (nine for a three-category event, 16 for a four-category event, etc.), thus providing in one loop the values needed to determine conditional probabilities for all category combinations. Figure 6 is an example of the four-category print-out. The program permits a complete choice of initial LST hours, with one table being prepared for each initial hour specified. In addition to identification data (location, month/season, categories) the only inputs required are the 24 (each) hourly values of the unconditional probabilities of the various categories. These can often be estimated from only three-hourly data, or even six-hourly data associated with times of sunrise and sunset.

```
                                 MCCOY AFB, FLA * JUNE

                           ESTIMATE OF CONDITICNAL PRCBABILITY

                 THE PROBABILITY THAT, FOR A GIVEN INITAL CATEGORY  (A, B, C, D),
                EACH OF THE CATEGORIES (A, B, C, D) OCCURS AT SPECIFIEC LAG TIMES.

                                    VALUES IN PERCENT

GATEGORY DEFINITION
    CAT A * CLOUD AMOUNT 2/10 OR LESS
    CAT B * CLOUD AMOUNT 3/10 THRU 5/10
    CAT C * CLOUD AMOUNT 6/10 OR 7/10
    CAT D * CLOUD AMOUNT 8/10 OR MORE

INITAL HOUR  3 LOCAL STANDARD TIME

                                            H O U R S   L A T E R
INITIAL     SUBSEQUENT
CATEGORY     CATEGORY    1   2   3   4   5   6   7   8   9  10  11  12  15  18  21  24  30  36  42  48

    A           A       89  76  64  53  42  31  21  16  12   9   8   8  18  37  56  57  22   6  32  54
    A           B       10  18  21  24  29  36  37  37  37  36  36  34  21  18  15  18  29  28  16  18
    A           C        1   5  10  12  15  15  21  22  23  23  21  18  16  13  10  11  16  17  13  11
    A           D        0   2   5  10  14  18  21  25  28  33  35  39  45  33  19  15  33  50  38  17

    B           A       24  21  19  14  12   8   6   4   3   3   3   3   9  24  42  47  16   4  27  47
    B           B       56  37  28  27  25  28  23  23  22  21  20  20  16  16  17  21  25  21  16  20
    B           C       18  25  26  21  22  23  26  23  22  22  19  18  14  16  15  12  17  16  14  12
    B           D        3  17  27  38  41  42  46  50  53  54  57  60  62  44  26  20  43  58  43  20

    C           A        2   6   8   6   6   4   2   1   3   2   1   2   9  20  39  43  16   3  23  44
    C           B       31  25  18  19  17  17  18  17  18  15  16  15  13  15  14  17  24  23  16  18
    C           C       41  29  23  21  20  20  22  21  21  19  17  15  13  14  12  14  15  16  12  14
    C           D       27  41  51  54  57  59  59  61  59  65  66  68  66  51  35  26  45  59  50  25

    D           A        0   1   1   1   2   1   1   0   0   0   1   1   4  14  31  35  11   4  23  42
    D           B        3   5   6   6   7   9   8   8   7   7   8   8   8  13  15  18  22  18  15  19
    D           C       14  12  11  10  12  12  14  13  13  13  13  10  10  11  14  15  16  14  12  13
    D           D       83  82  81  83  79  78  78  79  79  79  78  80  79  63  40  31  51  64  50  27

 UNCONDITIONAL PROBABILITY
    AT INITIAL HOUR                      A T   L A T E R   H O U R S
    CATEGCRY
       A      49        48  42  36  30  24  18  12   9   7   5   5   5  12  28  46  49  18   5  28  49
       B      18        19  20  19  20  22  26  26  26  26  25  25  24  16  16  15  18  26  24  16  18
       C      12        11  13  14  14  16  16  20  20  20  20  19  16  14  13  12  12  16  16  13  12
       D      21        22  26  31  36  38  40  42  45  47  50  52  55  58  43  27  21  40  55  43  21
```

Figure 6. Estimate of Conditional Probability (Cloud Amount).

Accuracy of Conditional Estimates

Conditional probability estimates were made for four ceiling-visibility categories at Hamilton AFB, California for the month of January. The 1-, 3-, 6-, 12-, 24-, 36-, and 48-hour lag estimates for initial times of 0000, 0600, 1200, and 1800 LST were compared with the conditional frequencies as calculated from 23 years (1940-1962) of January observations. Table 1 shows the January diurnal variability of the frequency of occurrence (unconditional) of each of the four categories considered. Conditional estimates were computed from inputs of these unconditional frequencies into the automated program.

Root-mean-square (RMS) differences between estimated and observed conditional percentages were calculated. RMS differences between observed and assumptions of persistence and of unconditional probability were also determined. Table 2 compares these three sets of RMS differences as a function of lag time. The overall RMS difference between the estimates and observed frequencies is 7.5%. For none of the lag times considered was the difference of

TABLE 1

Percent Frequency of Occurrence of Ceiling/Visibility Categories for Hamilton AFB, California.

January (Period of Record 1940-1962)

	LST Hour											
Category	00	02	04	06	08	10	12	14	16	18	20	22
A	9	11	15	17	18	9	4	2	2	3	3	5
B	17	20	21	22	26	26	24	20	17	15	13	16
C	19	18	17	19	18	22	23	24	20	20	21	18
D	55	51	47	42	38	43	49	54	61	62	63	61

Category Definitions:

A. Ceiling less than 300 feet and/or visibility less than one mile.

B. Above Category A but ceiling less than 1500 feet and/or visibility less than three miles.

C. Above Category B but ceiling less than 5000 feet and/or visibility less than five miles.

D. Ceiling 5000 feet or higher (or no ceiling) and visibility five miles or more.

TABLE 2

RMS Percent Differences from Observed Conditional Frequencies for the Four Categories of Ceiling/Visibility for Hamilton AFB, California.

January (1940-1962 Historical Data)

	Lag Hours						
	1	3	6	12	24	36	48
Conditional Estimate	5.8	9.4	10.1	8.6	6.8	5.6	6.0
Persistence Assumption	15.3	23.1	31.6	35.4	39.0	42.0	42.8
Unconditional Assumption	34.4	27.1	20.1	17.4	10.9	7.4	6.5

an assumption of either persistence or unconditional probability less than that of the conditional estimate.

Cloud Amount Test

Conditional probability estimates were made by the same program for four categories of total cloud amount at McCoy AFB, Florida for the month of June. Figure 6 is one of the computer-generated tables and shows the diurnal variability of these cloud amount categories. RMS difference, similar to those of the Hamilton AFB sample, are given in Table 3. The overall RMS difference of conditional estimation here is 4.1%; and for all lag periods considered, the conditional estimates have smaller differences than an assumption of either persistence or unconditional probability.

TABLE 3

RMS Percent Differences from Observed Conditional Frequencies for the Four Categories of Cloud Cover for McCoy AFB, Florida.
June (1946, 1953-1965 Historical Data)

	Lag Hours						
	1	3	6	12	24	36	48
Conditional Estimate	4.9	3.2	4.0	3.1	6.6	3.3	3.3
Persistence Assumption	23.6	35.4	41.4	35.5	38.6	44.7	43.0
Unconditional Assumption	28.7	20.4	15.0	10.4	11.4	6.1	6.0

SECTION B — ESTIMATING PERSISTENCE PROBABILITY

The method set forth in Section A provides a means of estimating one-hour conditional probabilities which can vary by hour of the day as a result of the diurnal variation of the categorized event. If one assumes, as in a Markovian process, that future development is determined by the present state and not by the way in which the present state arose, then the probability of any sequence of hour-to-hour events can be specified as the product of the appropriate hour-to-hour conditional probabilities. One particular sequence of meteorological interest is that of "persistence," here defined as the repeated observation of the same event category at hourly intervals. (By this definition, changes that may occur within these hourly intervals, but not affecting the hourly recordings, are not identified as terminating a persistence run.)

A computer program was written which calculated estimates of persistence of categorized events from the hourly unconditional probabilities according to the hour-to-hour conditional estimates. Persistence probability estimates were printed out for hourly intervals to 24 hours for each hour of the day as an initial time.

Test of the Markov Assumption

Persistence probability estimates were made for three cloud-cover categories from the unconditional statistics for Wright-Patterson AFB, Ohio. These were compared with persistence figures determined by hourly historical data for the ten years, 1957-1966. Several one-hour correlation coefficients between 0.94 and 0.99 were used to generate the Markovian persistence probability estimates. By comparison, the historical persistence frequencies appeared quite non-Markovian. The low (0.94) correlation estimates were in good agreement with the historical data for short-period (one to three hour) persistence but grossly underestimated the long-period (21 to 24 hour) persistence. The high (0.99) correlation gave good estimates for the long-period persistence but its estimates were much too high for the short periods. Thus, it appears that for this set of data the one-hour correlation was indeed a function of how long the event had already persisted, the correlation increasing as the persistence period lengthened.

Correlation Function

The program for estimating persistence probability was modified to incorporate a one-hour correlation coefficient which was a function of how long the event had already persisted. Several correlation functions were tested. Persistence estimates for three categories of cloud cover, made according to these functions, were compared with the 1957-1966 Wright-Patterson AFB historical data for April. One correlation function which gives estimates close to the historical persistence frequencies is

$$(10) \qquad \rho_L = 0.95 + 0.04\,[(L - 1)/23]^{.2}$$

where ρ_L is the one-hour correlation coefficient for an event that is known to have persisted for L - 1 hours. In the program, L can be any integer from 1 to 24. Thus, $\rho_1 = 0.95$, the first hourly conditional step having no known prior persistence, and $\rho_{24} = 0.99$, the 24th hourly step knowing that the event has already persisted 23 hours. Figure 7 is an example of the program print-out.

WRIGHT-PATTERSON AFB, OHIO • APRIL

ESTIMATE OF DURATION (PERSISTENCE) PROBABILITY

(THE PROBABILITY THAT THE EVENT BEING CONSIDERED, IF IT OCCURS AT THE INITIAL TIME,
WILL ALSO OCCUR AT SUCCESSIVE HOURLY OBSERVATION TIMES FOR THE PERIODS INDICATED.)

(VALUES IN PERCENT)

EVENT BEING CONSIDERED-
CAT A • TOTAL CLOUD AMOUNT ZERO THRU 2/8

INITIAL TIME LST HR	HOURS PERSISTING 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	UNCONDITIONAL PROBABILITY AT INITIAL HR
0	87	77	69	62	57	42	35	32	28	25	22	16	14	11	10	8	8	8	7	7	7	7	6	6	41
1	84	76	67	62	45	37	34	30	26	23	17	15	12	10	8	8	8	8	8	7	7	7	7	6	41
2	87	77	71	51	42	38	34	30	26	19	16	13	11	9	9	9	8	8	8	8	7	7	7	6	39
3	84	77	56	46	42	37	32	28	21	17	14	12	10	9	9	9	9	8	8	8	8	7	6	6	39
4	87	63	52	46	40	35	30	22	19	15	13	11	10	10	9	9	9	9	8	8	7	7	6	6	37
5	70	57	51	44	38	33	24	20	17	14	11	11	10	10	10	9	9	8	8	8	7	7	6	6	38
6	79	69	60	52	44	33	27	22	18	15	14	14	13	13	12	12	11	11	10	9	9	8	8	6	28
7	85	73	62	53	39	32	26	21	17	17	16	15	15	14	13	13	12	12	11	10	9	9	7	6	25
8	81	68	59	42	34	28	23	18	18	17	16	15	15	14	13	13	12	11	11	10	9	7	6	6	26
9	80	67	49	40	32	26	20	20	19	17	17	16	16	15	14	13	12	12	11	10	7	6	6	6	25
10	79	57	46	37	30	23	23	21	20	19	18	18	16	16	15	14	13	12	11	8	7	7	6	6	24
11	69	55	43	35	27	26	24	22	22	21	20	18	18	17	15	14	13	13	9	8	8	7	6	6	23
12	77	60	49	37	35	33	30	29	28	27	25	24	23	21	19	18	17	13	11	10	9	8	7	6	18
13	73	59	45	41	39	35	34	33	31	29	28	27	24	22	21	19	14	12	12	11	10	9	7	6	16
14	73	55	51	48	43	42	40	38	35	34	32	29	27	25	23	17	15	14	13	11	10	8	7	6	14
15	72	66	61	56	54	51	49	45	44	41	37	35	31	29	22	19	18	16	14	13	10	8	7	6	13
16	86	80	72	69	65	62	57	56	52	47	44	40	37	27	23	22	20	18	16	12	10	9	8	6	11
17	87	77	73	69	65	60	59	55	49	46	41	39	29	24	23	21	19	16	12	11	9	8	7	6	14
18	85	81	76	71	65	63	59	53	49	44	41	30	26	24	22	20	17	13	11	10	8	7	7	6	18
19	91	86	79	73	70	65	58	54	48	45	33	28	26	24	21	19	14	12	10	9	7	7	7	7	20
20	89	81	74	72	66	58	54	48	45	33	28	26	23	21	18	14	12	10	9	7	7	7	7	7	26
21	88	79	77	70	62	57	51	47	35	29	27	24	22	19	14	12	10	9	7	7	7	7	7	7	31
22	87	83	76	67	62	55	51	37	31	29	26	23	20	15	13	11	9	8	7	7	7	7	7	7	34
23	92	84	74	67	60	56	41	34	31	28	25	22	16	14	12	10	8	8	8	7	7	7	7	6	35

Figure 7. Estimate of Duration (Persistence) Probability.

Accuracy of Persistence Estimates

The above correlation function was used to calculate persistence estimates for three categories of cloud amount for the midseason months at Tinker AFB, Oklahoma; Wright-Patterson AFB, Ohio; and Minot, North Dakota. The cloud-cover amount categories were:

Category A — Zero through 2/8

Category B — 3/8 through 5/8

Category C — 6/8 through 8/8

Persistence estimates for Categories A and C, referred to as "clear" and "cloudy" weather, respectively, for eight initial times of day (00, 03, ..., 21 LST hours) were compared with the persistence frequencies shown in ten years (1957-1966) of historical data. Analysis of the percentage differences of the estimates from the historical data frequencies (estimate minus historical frequency) revealed the following:

a. For the set of data as a whole (all three stations, all four months, all eight initial hours, and both categories), the correlation function gave estimates that were relatively unbiased, with an overall mean difference of -0.2% and RMS difference of 6.6%.

b. For clear persistence, the mean difference was -0.9%; for cloudy, it was +0.5%. This indicates that, with this model, the clear weather was slightly higher correlated than the cloudy weather; i.e., higher correlations would have given clear weather persistence estimates closer to the historical data frequencies; while lower correlations would have given cloudy weather persistence estimates closer to the historical data frequencies.

c. Similarly, of the three locations, Tinker AFB showed the highest correlation and Minot showed the lowest for both clear and cloudy conditions.

d. Of the four midseason months, October showed the highest correlation and July showed the lowest for both clear and cloudy conditions.

e. Persistence of clear conditions verifying between 0300 and 0900 LST hours showed higher correlation than those verifying between 1700 and 2100 LST hours. The opposite was true, to a lesser degree, for cloudy conditions.

Also, the RMS differences between persistence estimates and historical frequencies were compared with the standard deviations of the persistence frequencies for the clear and cloudy categories. (These standard deviations are equivalent to the RMS differences between the historical frequencies and the mean persistence of each category and duration period considered.) Comparisons for the 1-, 3-, 6-, 12-, and 24-hour durations showed RMS differences about 45% as great as the standard deviations for clear weather persistence and about 62% as great for cloudy weather persistence. Figures 8 and 9 show graphically the clear and cloudy RMS differences compared with the standard deviations in relation to the mean persistence as a function of the persistence period.

SECTION C — CONCLUSIONS

Information concerning conditional probability and persistence of meteorological events can be used both as an aid to forecasting and as a planning tool. Where large volumes of data are not available for summarization, the

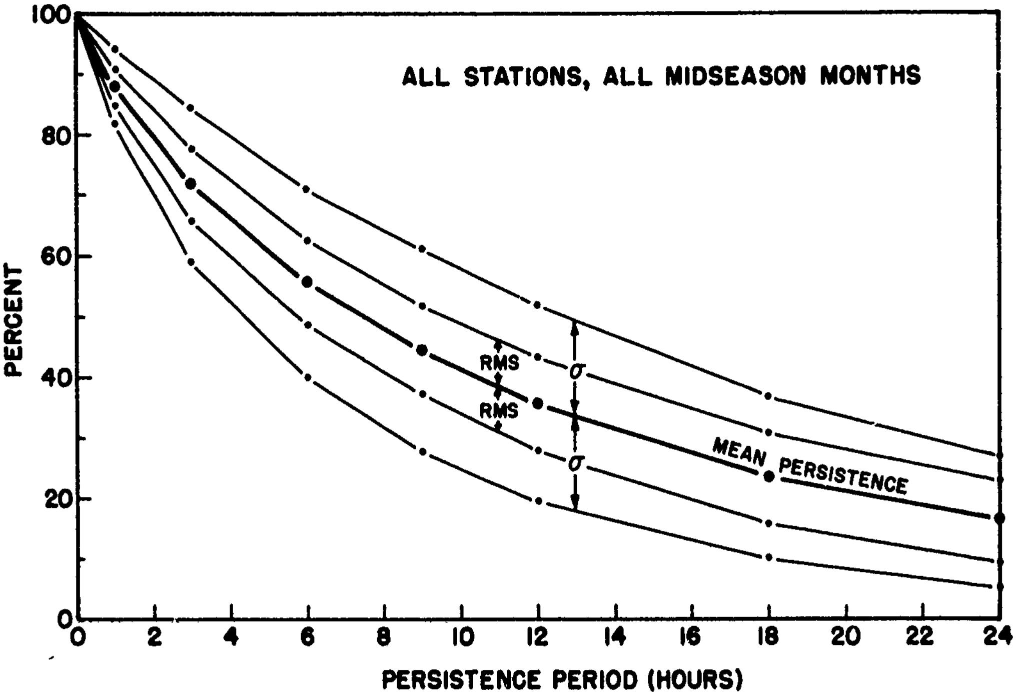

Figure 8. Mean and Standard Deviation of Clear Condition (0-2/8 Cover) Persistence with RMS Difference Between Estimate and Ten Years Data.

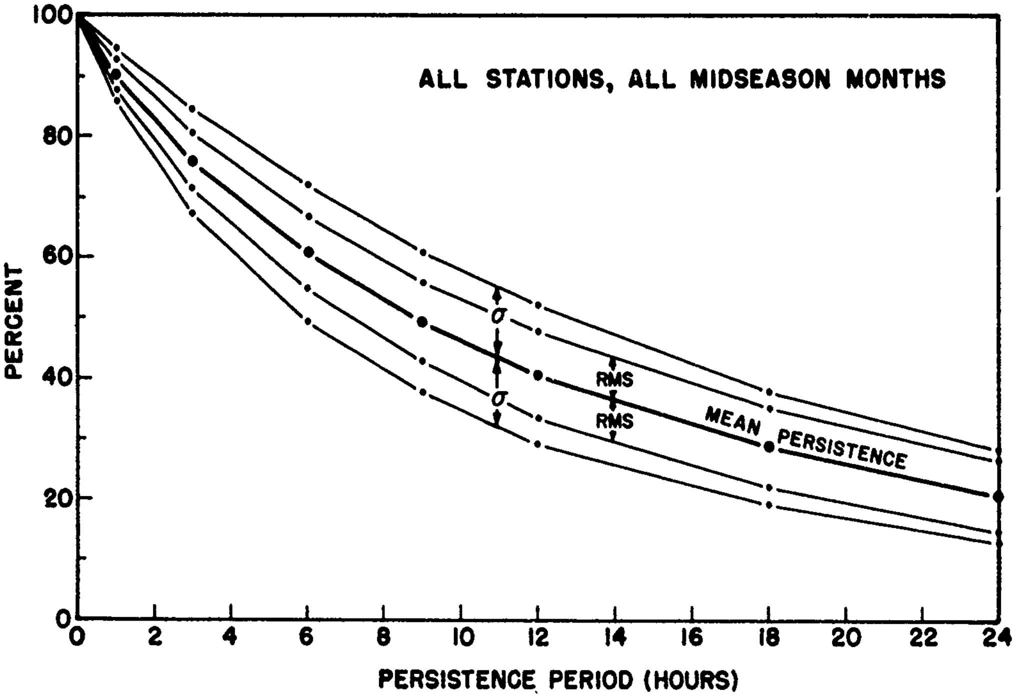

Figure 9. Mean and Standard Deviation of Cloudy Condition (6/8-8/8 Cover) Persistence with RMS Difference Between Estimate and Ten Years Data.

statistical models described in Sections A and B provide estimates of conditional probability and persistence which take into account the diurnal variability of the event being considered.

Although the models have had only a limited test with cloud cover and ceiling/visibility variables, they probably can be used, with little or no modification, to provide estimates concerning other meteorological parameters such as precipitation, temperature, humidity, and wind, so long as the diurnal variation of the event categories can be provided as input to the model.

The models can also be used with meteorological events which do not show significant diurnal variability, as perhaps certain upper-air parameters. However, for such cases, the computer programs can be greatly simplified in input and print-out as well as in the calculation routines. If conditional or persistence estimates are made by these models for variables lacking significant diurnal variability, it would be wise to compare the results with estimates obtained by Gringorten's method [5].

The methods described in this report are not meant to eliminate a need for conditional and persistence frequencies which can be derived from sufficient historical data. Indeed, there is hope for improvement of these models by use of additional historical frequencies to better describe the correlations as a function of a wider range of variables and locations, season, time of day, etc.

For the present, at least, the methods provide first estimates of conditional and persistence probability whenever adequate data are not available for summarization, but the event's diurnal variability is known or can be estimated. When the estimation of the input statistics is necessary, it is a problem that should be left to the well-trained meteorologist-climatologist who is familiar with the climatology of the area of interest.

REFERENCES

[1] Restivo, J. S.: "Persistence Probability," 4th Weather Wing Technical Paper 66-1, August 1966, 34 pp.

[2] Restivo, J. S.: "Onset-Duration Tables of Specified Ceiling and Visibility Categories," 4th Weather Wing Technical Paper 67-1, August 1967, 17 pp.

[3] Brooks, C. E. P., et al.: "Upper Winds Over the World," Geophysical Memoirs No. 85, M.O. 499e, Meteorological Office, Great Britain, pp. 1-50.

[4] Crutcher, H. L. and Baer, L.: "Computations from Elliptical Wind Distribution Statistics," J. Appl. Meteor., Vol. 1, 1962, pp. 522-530.

[5] Gringorten, I. I.: "A Stochastic Model for the Frequency and Duration of Weather Events," J. Appl. Meteor., Vol. 5, 1966, pp. 606-624.

[6] U.S. Department of Commerce, National Bureau of Standards: Tables of the Bivariate Normal Distribution Function and Related Functions, Applied Mathematics Series - 50, Washington, U.S. Government Printing Office, 1959, pp. 1-258.

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)*
Hq USAF Environmental Technical Applications Center (MAC), Bldg 159, Navy Yard Annex
Washington, D. C. 20333

2a. REPORT SECURITY CLASSIFICATION
Unclassified

2b. GROUP
N/A

3. REPORT TITLE
Estimating Conditional Probability and Persistence

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*
N/A

5. AUTHOR(S) *(First name, middle initial, last name)*
McCabe, John T., Colonel, USAF

6. REPORT DATE	**7a. TOTAL NO. OF PAGES**	**7b. NO. OF REFS**
June 1968	20	6

8a. CONTRACT OR GRANT NO.
N/A

b. PROJECT NO.

c.

d.

9a. ORIGINATOR'S REPORT NUMBER(S)
Air Weather Service Technical Report 208

9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*
N/A

10. DISTRIBUTION STATEMENT
This document has been approved for public release and sale; its distribution is unlimited.

11. SUPPLEMENTARY NOTES
N/A

12. SPONSORING MILITARY ACTIVITY
Hq Air Weather Service (MAC)
Scott AFB Illinois 62225

13. ABSTRACT

This report describes a statistical model and automated techniques which provide estimates of conditional and persistence probability of meteorological events for periods to 48 and 24 hours, respectively. The products are designed to assist forecasters and planners in instances where conditional and persistence frequencies are not obtainable by directly processing observational data.

The model considers the diurnal variability of the event by assuming joint probability according to elliptical distributions defined by known (or estimated) hourly unconditional probabilities and lag correlation coefficients obtained from previously summarized data. A computer program performs the integrations by transforming the elliptical distributions to the circular normal with rotated axes, and then counting the number of mil-frequency units in each of the joint probability zones.

Comparison of conditional and persistence probability estimates of categorized cloud cover and ceiling/visibility events with observed frequencies of occurrence reveal root-mean-square differences of the order of 5% to 15%.

DD FORM 1 NOV 65 **1473**

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Statistical model						
Climatology						
Meteorology						
Forecasting						
Probability						
Persistence						
Duration						
Elliptical distributions						