An Analysis of the Relationships of Selected Streamflow Characteristics to Physical Geographic Patterns in the Plains Area of the Canadian Prairie Provinces

EDWARD S. SPENCE

INLAND WATERS DIRECTORATE
DEPARTMENT OF THE ENVIRONMENT
OTTAWA, CANADA, 1972

FOREWORD

This paper, based on Dr. Spence's thesis, is printed, for limited distribution only, by the Network Planning and Forecasting Section, Applied Hydrology Division, Water Resources Branch.

It is hoped that the paper will prove to be of use to studies engineers, particularly in the Water Survey of Canada District Offices.

AN ANALYSIS OF THE RELATIONSHIPS OF SELECTED STREAMFLOW CHARACTERISTICS TO PHYSICAL GEOGRAPHIC PATTERNS IN THE PLAINS AREA OF THE CANADIAN PRAIRIE PROVINCES

by:

Edward S. Spence
Assistant Professor
Department of Geography
York University
Downsview, Ontario

ABSTRACT

This study is an examination of the relationships of selected streamflow characteristics to physical geographic patterns in the plains area of the Canadian Prairie Provinces. The aims of the study were twofold, firstly, to develop a set of statistical relationships for the prediction of streamflow characteristics for ungauged basins, and secondly to add to our understanding of plains hydrology through the identification of physical geographic variables which are related to streamflow patterns.. In view of the multivariate nature of the relationships being considered, a system investigation approach utilizing the statistical techniques of multiple regression and factor analysis was adopted.

A group of 161 study basins was selected for analysis. Each of these basins met defined criteria with regard to basin location, size of drainage area, available streamflow data, and natural flow conditions.

Four dependent variables; the mean annual yield, the mean 10 year yield, the mean annual flood, and the mean 10 year flood; were selected for analysis. The available annual yield and annual flood flow data series for each of the basins were compiled for the base period 1940 to 1969. Wherever possible, short-term records were extended by correlation with records from nearby longer-term stations. The actual estimation of the magnitudes of the dependent variables was made by frequency analysis of the available data series. These analyses were based on the assumption of the lognormal distribution and utilized a least squares curve fitting technique.

Thirty-nine independent variables were estimated for each of the study basins. These variables were chosen on the bases of their theoretical relationships to streamflow and available data sources. The first group of 20 independent variables were measures of climatic patterns and were compiled from published climatic normal data. The other 19 independent variables, measures of other physical geographic patterns, were compiled from 1:250,000 scale topographic maps, and included measures of drainage area, basin topography, and vegetation.

The initial stage of the analysis was an examination of the relationships for the entire study area. Two approaches to these analyses were employed; firstly, a stepwise multiple regression analysis considering all of the independent variables, and secondly, a stepwise multiple regression analysis considering only those independent variables selected after factor analytic screening. The latter approach proved to be more satisfactory in that the signs of the regression coefficients conformed to physical expectations. In an effort to improve the predictive strength of the models, the second stage of the analysis involved the division of the study area into hydrologic regions. Two hydrologic regions were delimited; however, regression analyses for each of these regions did not result in appreciable improvement of the predictive power of the full study area models.

The regression models for each of the four dependent variables, as developed in this study, conform to physical expectations, are stable when tested with independent data and are statistically significant. The standard errors of the estimates for the regression models were relatively large and limit the predictive applications of these relationships. On the basis of the available data, it has not been possible to establish strong predictive models. Several suggestions have been made for possible extensions of the present research with the aim of further improving this predictive strength.

ACKNOWLEDGEMENT

This report is based primarily on the Author's Ph.D. thesis research which was completed at the University of Alberta. The author wishes to acknowledge the assistance of numerous members of the faculty and technical staff of the University of Alberta. Particular thanks are extended to Dr. Arleigh H. Laycock, the chairman of the author's research advisory committee, for his continued encouragement and helpful comments throughout the project.

The financial assistance of the National Research Council of Canada and of the Canada Department of Energy, Mines and Resources is gratefully acknowledged.

TABLE OF CONTENTS

APPENDICES

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION, PROPOSED METHODOLOGY, THE STUDY AREA AND STUDY BASINS

1.1 INTRODUCTION

One of the areas of Canada where regional water shortages are particularly pronounced is the plains region of the prairie provinces. In the face of growing competitive demands, the general public is demanding careful planning, allocation and management of the available water resources. One of the primary requirements of those faced with the task of planning for the future management of our water resources is for accurate estimates of the available naturally occurring supplies. It is to this need that the present study is addressed.

This study is a statistical analysis of the relationships between selected streamflow characteristics and the climatic and other physical geographical patterns for the plains area of the Canadian Prairie Provinces. The hypothetical model for the study is of the form

$$\text{STREAMFLOW CHARACTERISTICS} = f\ (\text{CLIMATIC PATTERNS AND OTHER PHYSICAL GEOGRAPHIC PATTERNS}) \qquad (1\text{-}1)$$

The streamflow characteristics selected for analysis are the annual yields and the annual flood flows. The overall aim of the research is to increase knowledge of plains hydrologic patterns through the establishment of a series of statistical relationships for the prediction of streamflow characteristics in ungauged plains basins. It is intended that the analysis will identify the climatic and other physical geographic variables which are most closely related to the selected streamflow characteristics. These variables once identified will aid in determining the direction of future more specific process oriented research into the hydrologic patterns of the area. The statistical relationships formulated will provide a basis for the preliminary estimation of available water resources.

1.2 PROPOSED METHODOLOGY AND LITERATURE REVIEW

The analysis of the relationships of streamflow to physical geographic patterns is a complex and difficult task. The multi variate nature of the hydrologic processes involved results in the researcher having to examine the integrated effects of numerous variables, none of which may be accurately measured and some of which are, at the present time, unmeasurable. These difficulties associated with hydrologic research have led to the development of two seemingly separate approaches to hydrologic study. The two approaches have been characterisized by Amorocho and Hart (1964) as physical hydrologic investigations, with an emphasis on physical science research into the components of the hydrologic cycle, and hydrologic system investigations, including both stochastic and parametric hydrologies, with an emphasis on the input-output relationships.

In view of the incomplete knowledge of physical hydrologic relationships of the plains and in order to make full use of the available data base, a system investigation approach has been adopted in the present research. Both stochastic and parametric techniques of analysis have been employed in a regional analysis of the hypothetical model (equation 1-1). While it is not possible to accurately predict hydrologic events, stochastic analyses techniques provide for the estimation of the probabilities of occurance of events of given magnitude. In the present study frequency analysis techniques have been employed in the estimation of streamflow characteristics; and the parametric technique of multiple correlation and regression analysis has been utilized in an attempt to develop predictive relationships for ungauged basins.

A combination of stochastic and parametric analysis techniques has resulted in some of the most useful hydrologic systems investigation work contained in the literature. The largest body of such research deals with the frequency analysis of flood flows and has been summarized in papers by Jarvis and others (1936), Benson (1962a) and Wolf (1966).

Probably the most serious problem relating to stochastic analysis of hydrologic events is the lack of long-term data records. Benson (1960) has demonstrated the high degree of variability involved in frequency analysis of relatively short data sequences. He has also demonstrated that greater confidence may be placed in the results of a regional frequency analysis in which the data from several basins are considered together.

Regional flood frequency studies have utilized both the index-flood and multiple regression approaches. The index-flood method has been discussed in detail by Dalrymple (1960) and examples of this approach include the work in Canada by Durrant and Blackwell (1959), Coulson (1967a and 1967b) and Collier and Nix (1967); and numerous studies in the United States conducted by personnel of the geological survey (for a partial bibliography of these studies see Wong 1963 and Benson 1962a). Benson (1962a) proposed the multiple regression approach as an alternative to the index-flood approach. When sufficient data covering both flood magnitudes and the hydrologic characteristics of the basins are available, the multiple regression technique has several advantages. Examples of this method of regional flood frequency analysis are found in the work of Benson in New England (1962b) and in the Southwest (1964); and in Karuks' (1964) study of floods in Southern Ontario. The multiple regression approach to regional flood frequency analysis has also been extended to other streamflow characteristics. Solomon et.al. (1968) employed a similar method in the analysis of mean annual yields on a grid square basis for Newfoundland; while Benson and Carter (1969) and Thomas and Benson (1971) have analysed data on numerous streamflow characteristics for several regions of the United States.

1.3 THE STUDY AREA

1.3.1 Location and Boundary Definition

For the purposes of the current research, the study area, the Canadian Plains, has been delimited as that area of Alberta, Saskatchewan and

Manitoba which lies east of the Rocky Mountain Foothills and south and west of the Canadian Shield margin. The location of the study area and its boundaries are illustrated on the inset map Figure 1-1.

The study area boundaries are the result of a consideration of both physiographic and political divisions. Both the sections of the western limit, and of the northern and eastern limits, which correspond to the margins of the foothills and shield respectively, are based on physiographic divisions. These physiographic boundaries represent distinctive changes in several physical patterns, including surficial and bedrock geology, and topography, all of which are significant hydrologic variables. The other boundaries of the study area are coincident with political boundaries. These boundaries are arbitrary in their application to the present research as they do not have a hydrologic basis. However, some justification for them can be given on the basis of available data. A general lack of suitable long-term hydrometric records in northeastern British Columbia and in the Northwest Territories precludes the inclusion of these areas in the present project; although physiographically they are included in the Canadian Plains. The southern boundary, corresponding to the international boundary, marked the division of data collection responsibilities between the two nations. The present study is based on available Canadian data and is therefore limited on the south by the international border.

1.3.2. General Physical Geography of the Study Area

The study area physiographically is within the Interior Plains Province as delimited by Bostock (1964) and the Geological Survey of Canada (1970). This large physiographic province is geologically a vast crescent shaped sedimentary basin, which rings the Canadian Shield from the United States border to the Arctic coast (Clibbon and Hamelin 1967, p. 72). The general topography of the study area is flat to rolling with local relief features predominating. Local topographic variations are the result of glacial erosion and deposition and of more recent fluvial action. The magnitude of local relief tends to increase from the southeast toward the north and west.

The climate of the study area in continental. The continental effect on the temperature regime of the study area results in an extremely high annual temperature range which varies from 48°F (Calgary - January 14° July 62°) in the southwest of the study area, to over 70°F (Ft. Vermilion - January -9°, July 62°) in the north and 68°F (Winnipeg - January 0°, July 68°) in the east. Throughout the area the winters are extremely cold but the summers are quite warm. With respect to precipitation the study area is relatively dry; however, local variations and the seasonal distribution provide significant modifications to the pattern. The mean annual precipitation ranges upward from less than 12 inches in Southeastern Alberta and Southwestern Saskatchewan to over 20 inches in parts of Northern Alberta and Southern Manitoba. On the basis of Thornthwaite water balance calculations, the annual water surpluses are small being less than one inch for large areas in the south of the study area (Laycock 1967 and Sanderson and Phillips 1967). The available water surpluses are concentrated in the spring months during the snowmelt period; and, the

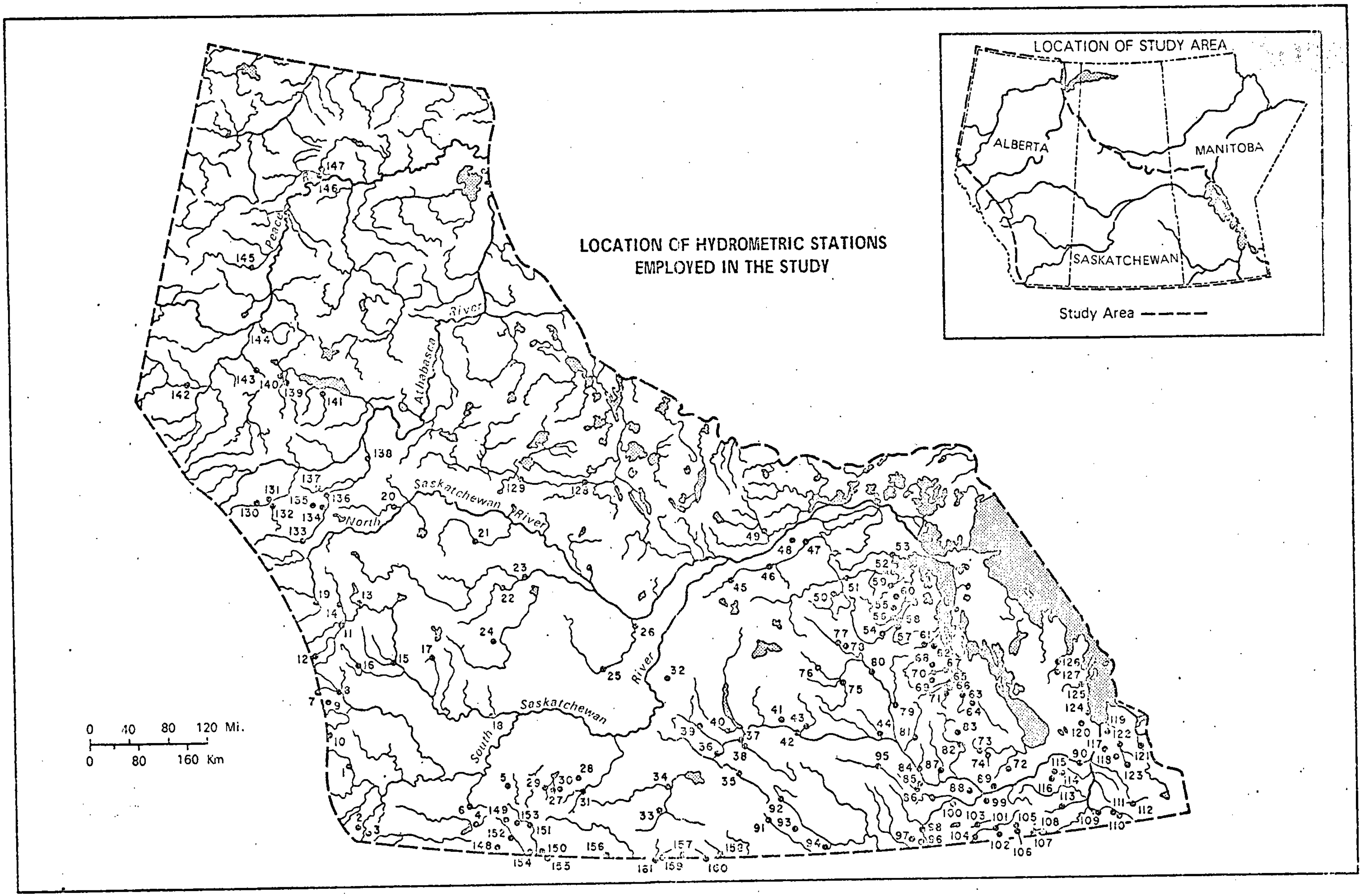

LOCATION OF HYDROMETRIC STATIONS EMPLOYED IN THE STUDY

Figure 1 - 1

annual runoff regimen of the plains streams is characterized by an early spring peak with low summer and winter flows. During the summer season in the drier areas many small streams flow only as the result of heavy and prolonged rainfall.

1.4 CHOICE OF STUDY BASINS

In this study the term "study basin" is employed to refer to the gross topographic drainage area upstream from a given gauging site. It was proposed to include as study basins all basins which met the following five criteria:

1. The stream must have on it at some location an active or discontinued hydrometric station for which daily discharge measurements are available.
2. The hydrometric station must be located within the study area as defined above.
3. The gross topographic drainage area must be greater than 50 square miles and less than 10,000 square miles.
4. Within the base period 1940-1969, the hydrometric station must have a minimum of five years of daily discharge data for the open water season, March 1 to October 31.
5. The streamflow of the basin must be natural flow with no major storage or diversion works upstream from the hydrometric station.

The above five criteria were applied in a stepwise fashion to the data from each of 1,478 active and discontinued hydrometric stations in the Three prairie provinces. The necessary information for this procedure was drawn from two separate sets of publications of the Water Survey of Canada, "1968 Surface Water Data Reference Index" for each of the prairie provinces (Canada, Department of Energy, Mines and Resources, 1969a, 1969b, and 1969c), and the "1967 Surface Water Data" for each of the prairie provinces (Canada Department of Energy, Mines and Resources, 1967a, 1967b, and 1967c). Table 1-1 summarizes the stepwise elimination procedure which resulted in the final selection of 161 study basins from the original total of 1,478 possible.

In order to provide some additional information relating to the selection procedure, it is useful to give brief mention to the rationale underlying the criteria employed. The first criteria relating to available daily discharge measurements was necessary to distinguish hydrometric stations for which discharge measurements are compiled from those measuring lake levels or stream stages. The second criteria relating to the study area imposed the geographical limits on the study. This criteria has been applied in a flexible manner in that the physiographic boundaries of the study area can not be accurately located and the political boundaries have no hydrologic significance. The third criteria imposed arbitrary limits on the size of study basins. While the actual limits are arbitrary they are based on practical considerations. Basins smaller than 50 square miles pose difficult problems in the compilation of physical geographic data from topographic maps. The upper limit of 10,000 square miles has been chosen so as to include most of the large relatively hydrologically homogeneous plains basins while eliminating the larger interregional basins. The fourth criteria relates to the length

TABLE 1-1 - SELECTION OF STUDY BASINS FROM AVAILABLE HYDROMETRIC STATIONS

Province	Total Number of Hydrometric Stations Active and Discontinued	Criteria for Rejection of Basin					Selected Study Basins
		1) Lack of Discharge Data	2) Outside Study Area	3) Gross Drainage Area	4) Length of Records	5) Natural Flow	
Alberta	585	80	129	42	183	107	44
Saskatchewan	428	86	13	47	151	80	51
Manitoba	465	210	36	49	63	41	66
Totals	1478	376	178	138	397	228	161

of record for the open water season within the base period 1940-1969. The specification of the open water season March 1 to October 31 is necessary in that hydrologic stations on many small plains streams are not operated during the November to February period. The final criteria relating to natural streamflow was necessary in that the analysis in the present research was intended to study natural conditions which would be disturbed in basins which have major storage or diversion works.

Figure 1-1 is a map showing the locations of the 161 hydrometric stations selected for inclusion in the analysis. Each of these study basins is identified by number in Table 1-2.

TABLE 1-2

HYDROMETRIC STATION IDENTIFICATION

STATION NUMBER	WATER SURVEY STATION NUMBER	STATION NAME	STATION NUMBER	WATER SURVEY STATION NUMBER	STATION NAME
1	05AB021	WILLOW CK	82	05MF001	MINNEDOSA R
2	05AE002	LEE CK	83	05MF008	ROLLING R
3	05AE005	ROLPH CK	84	05MG001	ARROW R
4	05AF010	MANYBERRIES CK	85	05MG002	BOSSHILL CK
5	05AH002	MACKAY CK	86	05MG003	GOPHER CK
6	05AH041	PEIGAN CK	87	05MG004	OAK R
7	05BJ004	ELBOW R	88	05MH006	LITTLE SOURIS R
8	05BJ005	ELBOW R	89	05MH007	EPINETTE CK
9	05BK001	FISH CK	90	05MJ004	STURGEON CK
10	05BL007	STIMSON CK	91	05NA005	GIBSON CK
11	05CB001	LITTLE RED DEER	92	05NB011	YELLOWGRASS DITC
12	05CB002	LITTLE RED DEER	93	05NB014	JEWEL CK
13	05CC001	BLINDMAN R	94	05NB019	COULEE WEST
14	05CC007	MEDICINE R	95	05NE003	PIPESTONE CK
15	05CE002	KNEEHILLS CK	96	05NF002	ANTLER R
16	05CE006	ROSEBUD R	97	05NF007	GAINSBOROUGH CK
17	05CG002	BULLPOUND CK	98	05NF008	GRAHAM CK
18	05CK005	ALKALI CK	99	05NG010	OAK CK
19	05DB002	PRAIRIE CK	100	05NG012	ELGIN CK
20	05EA001	STURGEON R	101	05OA006	WHITEMUD CK
21	05EE001	VERMILION R	102	05OA007	BADGER CK
22	05FD001	RIBSTONE CK	103	05OA008	PEMBINA
23	05FE001	BATTLE R	104	05OA009	WAKOPA
24	05GA003	MONITER CK	105	05OB006	CRYSTAL CK
25	05GC005	EAGLE CK	106	05OB010	LONG R
26	05GC006	EAGLE CK	107	05OB016	SNOWFLAKE CK
27	05HA003	BEAR CK	108	05OB021	MOWBRAY CK
28	05HA015	BRIDGE CK	109	05OD001	ROSEAU R
29	05HA062	PIAPOT CK	110	05OD004	ROSEAU R
30	05HA075	SKULL CK	111	05OD014	ROSEAU R
31	05HD036	SWIFTCURRENT CK	112	05OE004	RAT R
32	05HG002	BRIGHTWATER CK	113	05OF015	SHANNON CK
33	05JA002	WOOD R	114	05OG004	ELM CK 1
34	05JB003	NOTUKEU CK	115	05OG005	ELM CK 2
35	05JE004	MOOSE JAW R	116	05OG006	ELM CK 3
36	05JE006	MOOSE JAW R	117	05OJ002	COOKS CK
37	05JF005	WASCANA CK	118	05OJ006	COOKS CK
38	05JF006	BOGGY CK	119	05OJ008	NETLEY CK
39	05JG004	QU'APPELLE R	120	05OJ009	NETLEY CK
40	05JH001	ARM R	121	05PH003	WHITEMOUTH R
41	05JK004	JUMPING DEER CK	122	05SA002	BROKENHEAD R
42	05JL002	INDIANHEAD CK	123	05SA004	BROKENHEAD R
43	05JL005	PHEASANT CK	124	05SB002	OSIER CK
44	05JM015	CUTARM CK	125	05SC002	ICELANDIC R
45	05KA001	CARROT R	126	05SD003	FISHER R
46	05KB003	CARROT R	127	05SD004	EAST FISHER R
47	05KC001	CARROT R	128	06AD001	BEAVER R
48	05KD002	PETAIGAN R	129	06AD006	BEAVER R
49	05KE002	TORCH R	130	07AF002	MCLEOD R
50	05LB002	ETOMAMI R	131	07AG001	MCLEOD R
51	05LC001	RED DEER R	132	07AG003	WOLF CK
52	05LC004	RED DEER R	133	07BA001	PEMBINA R
53	05LD001	OVERFLOWING R	134	07BB002	PEMBINA R
54	05LE001	SWAN R	135	07BB003	LOBSTICK R
55	05LE003	BIRCH R	136	07BB004	PADDLE R
56	05LE004	WOODY R	137	07BB005	LITTLE PADDLE R
57	05LE005	ROARING R	138	07BC002	PEMBINA R
58	05LE006	SWAN R	139	07BF001	EAST PRAIRIE R
59	05LF001	STEEP ROCK R	140	07BF002	WEST PRAIRIE R
60	05LF002	BELL R	141	07BJ001	SWAN R
61	05LG001	PINE R	142	07GE001	WAPITI R
62	05LG002	GARLAND R	143	07GH002	LITTLE SMOKY R
63	05LJ005	OCHER R	144	07HA003	HEART R
64	05LJ007	TURTLE R	145	07HC001	NOTIKEWIN R
65	05LJ011	WILSON R	146	07JF002	BOYER R
66	05LJ012	VERMILION R	147	07JF003	PONTON R
67	05LJ015	FISHING R	148	11AA026	SAGE CK
68	05LJ016	FORK R	149	11AB009	MIDDLE CK
69	05LJ017	DRIFTING R	150	11AB075	LYONS CK
70	05LJ019	MINK R	151	11AB076	BATTLE CK
71	05LJ022	EDWARDS CK	152	11AB082	LODGE CK
72	05LL007	PINE CK	153	11AB087	MIDDLE CK
73	05LL009	NEEPAWA CK	154	11AB105	WOODPILE COULEE
74	05LL013	WHITEMUD R	155	11AB107	EAST BATTLE CK
75	05MB001	YORKTON CK	156	11AD001	WHITEWATER CK
76	05MB004	WHITESAND R	157	11AE002	POPLAR R W
77	05MC001	ASSINIBOINE R	158	11AE003	POPLAR R E
78	05MC002	STONY CK	159	11AE005	ROCK CK
79	05MD005	SHELL R	160	11AE008	POPLAR R M
80	05MD006	LITTLE BOGGY CK	161	11AE009	ROCK CK
81	05ME003	BIRDTAIL CK			

CHAPTER II

THE SELECTED STREAMFLOW CHARACTERISTICS: DATA COMPILATION AND SINGLE STATION FREQUENCY ANALYSIS

2.1 INTRODUCTION

This chapter contains discussions of the data compilation and frequency analysis of the selected streamflow characteristics for the study basins. As outlined in Chapter 1, two sets of hydrologic data, annual yields and the annual flood flows, were selected for examination in the present study. These data were analysed by the probabilistic techniques of frequency analysis. These analyses resulted in the estimation of parameters describing the magnitudes and frequencies of both annual yields and flood flows for each of the study basins. These parameters were subsequently employed as dependent variables in the regional parametric analyses of the relationships of streamflow characteristics to climatic and other physical geographic patterns.

2.2 HYDROMETRIC DATA COMPILATION FOR THE STUDY BASINS

A total of 161 hydrometric stations were selected for inclusion in the present study. These stations include all those within the study area which meet the criteria relating to type of records, gross drainage area, length of record, and natural flow conditions, as outlined in Chapter 1. For each of the selected hydrometric stations, the gross topographic drainage area upstream is assumed to delineate a study basin.

2.2.1 Data Sources Employed

One of the criteria applied in the selection of study basins was that within the 1940 to 1969 base-period, a minimum of five years of daily discharge records for the open water season (March 1 to October 31) be available. The source of hydrometric data employed in the present study is the records of the Water Survey of Canada, a branch of the Department of the Environment. All available daily discharge data for the period prior to 1966 were available on magnetic tapes prepared by the Water Survey of Canada. Daily discharge data for the study basins for the years 1967 and 1968 were available in published form in the annual data publications of the Water Survey of Canada for the provinces of Alberta, Saskatchewan and Manitoba (Canada, Department of Energy, Mines and Resources 1967a, b and c; and 1968a, b, c). The daily discharge data for 1969 were supplied in unpublished form by the Calgary and Winnipeg district offices of the Water Survey of Canada. Data for 1969 which were not available prior to May 1, 1970 have not been employed in the present study. The actual data required in the present study have been compiled directly from the magnetic tapes for the period prior to 1966 and from the printed records for later years.

2.2.2 Compilation of the "Annual" Yield Data Series

Daily discharge data for many small streams in the prairie provinces are not collected during the winter period from November 1 to the end of

February. During this period, the hydrometric stations are closed down in response to the severe winter conditions. Under normal conditions, there is little or no flow in these streams during the winter period. Since winter data are not available for many of the selected study basins, the period from March 1 to October 31 has been defined as the "annual" period applied to the present study. Therefore throughout this project, the term "annual" as applied to streamflow data refers only to the March 1 to October 31 period. Undoubtedly, in some of the larger study basins, there is some significant discharge during the winter season; however for the purposes of the present study ,this flow has not been measured, and the eight month flow has been considered as representative of the annual flow at least in relative if not in absolute magnitude.

The annual yield data for this study were compiled from the daily discharge records for each of the study basins. The annual yield figures for each year have been arrived at by summing the daily discharges (c.f.s.) for the eight month period, and converting the sum to acre feet. In years for which entire months of daily discharge data were missing, no annual yield data were compiled. Where shorter periods of records were missing, the available daily data were examined and where possible the missing data were estimated by interpolation between preceding and succeeding measurements.

Table 2-1 has been prepared to indicate for each basin the years for which annual yield data have been compiled. The "X" symbols denote the years for which measured data are available, and the total years with measured data are listed in the column headed "ACTUAL YEARS."

2.2.3 Compilation of the "Annual" Flood Peak Data Series

In the compilation of any series of flood events, it is necessary to carefully define the term flood. Two different flood series, the annual series and the partial-duration series, have been proposed (Langbein 1949 and Chow 1950). The present research is concerned with the annual series which includes only the largest flood event of each year. The "annual" period employed is the same eight month, March to October, period adopted for the annual yield data. This eight month period is acceptable for the compilation of flood data since no significant flood flows would be expected during the winter season. Since available hydrometric records were in terms of daily discharges, it was necessary to further define an annual flood event as the highest daily discharge recorded during the "annual" period. For planning purposes, this value is not as useful as the maximum instantaneous flow. Unfortunately ,this measure cannot be estimated from the available daily discharge data. For stations with recording gauges, it is possible to establish a correction factor to convert maximum daily discharges into instantaneous peaks; however, no such correction has been attempted in the present study.

From the available daily discharge records, the annual flood series for these study basins has been compiled. The highest daily discharge measured in c.f.s. has been recorded for each year of record. Where months or several days of record are missing, the available data have been examined; and where the missing data seemed to constitute a low flow

TABLE 2-1

ANNUAL YIELD RECORDS *

	STATION ID		YEARS	ACTUAL YEARS	EXTENDED YEARS	TOTAL YEARS
			1940 1969			
1	05AB021	WILLOW CK	----XXXXXXXXXXXXXXXXXXXXXXXXX-	25	0	25
2	05AE002	LEE CK	XXXXXXXXXXXXXXXXXXXXXXXXXXXXX-	29	0	29
3	05AE005	ROLPH CK	XXXXXXXXXXX0XXXXXXXXXXXXXXXXX-	28	1	29
4	05AF010	MANYBERRIES CK	00--0-00000000000XXXXXXXXXXXXX	13	14	27
5	05AH002	MACKAY CK	0000000000000000XXXX0XXXXXXXXX	13	17	30
6	05AH041	PEIGAN CK	00000000000000000000XXXXXXXXXX	10	20	30
7	05BJ004	ELBOW R	000000000000XXXXXXXXXXXXXXXXXX	18	12	30
8	05BJ005	ELBOW R	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	30	0	30
9	05BK001	FISH CK	0000000000000000XXXXXXXXXXXXXX	14	16	30
10	05BL007	STIMSON CK	X--XXXXXXXXXXXX0XXXXXXXXXXXX0-	25	2	27
11	05CB001	LITTLE RED DEER	000000000000000000000XXXXXXXXX	9	21	30
12	05CB002	LITTLE RED DEER	000000000000000000000000XXXXXX	6	24	30
13	05CC001	BLINDMAN R	00000000000000000000000XXXXXXX	7	23	30
14	05CC007	MEDICINE R	000000000000000000000000XXXXXX	6	24	30
15	05CE002	KNEEHILLS CK	0000000000000000000XXXXXXXXXXX	11	19	30
16	05CE006	ROSEBUD R	00--0-0000000000000XXXXXXXXXXX	11	16	27
17	05CG002	BULLPOUND CK	0000000000000000000000XXXXXXXX	8	22	30
18	05CK005	ALKALI CK	00000000000000000000000XXXXXXX	7	23	30
19	05DB002	PRAIRIE CK	000000000000XXXXXXXXXXXXXXXXXX	18	12	30
20	05EA001	STURGEON R	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	30	0	30
21	05EE001	VERMILION R	----000000000000000XXXXXXXXXXX	11	15	26
22	05FD001	RIBSTONE CK	----000000000000000000XXXXXXXX	8	18	26
23	05FE001	BATTLE R	----XXXXXXXXXXXXXXXXXXXXXXXXX-	25	0	25
24	05GA003	MONITER CK	----0000000000XXXXXXXXXXXXXXXX	16	10	26
25	05GC005	EAGLE CK	----0000000000000000000XXXXXX-	6	19	25
26	05GC006	EAGLE CK	----0000000000000000000XXXXXX-	6	19	25
27	05HA003	BEAR CK	00000000000000000000000XXXXXXX	7	23	30
28	05HA015	BRIDGE CK	00000000000000000000000XXXXXX-	6	23	29
29	05HA062	PIAPOT CK	00000000000000000000000XXXXXX-	6	23	29
30	05HA075	SKULL CK	00000000000000000000000XXXXXXX	7	23	30
31	05HD036	SWIFTCURRENT CK	000000000000000XXXXXXXXXXX0XX-	13	16	29
32	05HG002	BRIGHTWATER CK	00000000000000000000X0XX0XXXX-	7	22	29
33	05JA002	WOOD R	0000-000000000000XXXXXXXXXXXX-	12	16	28
34	05JB003	NOTUKEU CK	0000-000000000000XXXXXXXXXXXX-	12	16	28
35	05JE004	MOOSE JAW R	------0000000XXXXXXXXXXXXXXXX-	16	7	23
36	05JE006	MOOSE JAW R	------0000000X0XXXXXXXXXXXXXX-	15	8	23
37	05JF005	WASCANA CK	--00--XXXXXX0XXXXXXXXXXXXXXXX-	22	3	25
38	05JF006	BOGGY CK	------0000X0000XXXXXXXXXX000X-	12	11	23
39	05JG004	QU'APPELLE R	----XXXXXXXXXXXXXXXXXXXXXXXX0-	24	1	25
40	05JH001	ARM R	------000000000XXXXXXXXXXXXXX-	14	9	23
41	05JK004	JUMPING DEER CK	-XXXXXXXXXXXXXXXXXXXXXXXXXXXX-	28	0	28
42	05JL002	INDIANHEAD CK	-0XX00XXXXX0XXXXXXXXXX0XXXXXX-	23	5	28
43	05JL005	PHEASANT CK	-000000XXXX0XXXXXXXXXXXXXXXXX-	21	7	28
44	05JM015	CUTARM CK	-------0000000000XXXXXXXXXXXX-	12	10	22
45	05KA001	CARROT R	-0000000000000000XXXXXXXXXXXXX	13	16	29
46	05KB003	CARROT R	-00000000000000XXXXXXXXXXXXXXX	15	14	29
47	05KC001	CARROT R	-00000000000000XXXXXXXXXXXXXXX	15	14	29
48	05KD002	PETAIGAN R	-000000000000000XXXXXXX000000-	7	21	28
49	05KE002	TORCH R	-000000000000XXXXXXXXXXXXXXXX-	16	12	28
50	05LB002	ETOMAMI R	-------0000000000XXXXXXXX0XXX-	11	11	22
51	05LC001	RED DEER R	-------0000000000XXXXXXXXXXXX-	12	10	22
52	05LC004	RED DEER R	-------0000000000XXX0XXXXXXXX-	11	11	22
53	05LD001	OVERFLOWING R	---------------00XXXXXXXXX0XX0	11	4	15
54	05LE001	SWAN R	-------0000XXXXXXXXX000000000-	9	13	22
55	05LE003	BIRCH R	-------000000000XXXXXXXXXX0XX-	12	10	22
56	05LE004	WOODY R	-------0000000000XXXXXXXXXXXX-	12	10	22
57	05LE005	ROARING R	-------0000000000000XXXX0XXXX-	8	14	22
58	05LE006	SWAN R	-------00000000000000XXXXXXXX-	8	14	22
59	05LF001	STEEP ROCK R	-------000000000XXX0XXXXXX0XX-	11	11	22
60	05LF002	BELL R	-------0000000000XXXXXXXX0XX0-	10	12	22
61	05LG001	PINE R	-------0000000000XXXXXXXXXXXX-	12	10	22
62	05LG002	GARLAND R	-------000000000XXX0XX0XXXXXX-	11	11	22
63	05LJ005	OCHER R	-------000000000XXXXXXXXXXXXX-	13	9	22
64	05LJ007	TURTLE R	-------0000000000XXXXXXXXXXXX-	12	10	22
65	05LJ011	WILSON R	-------0000000000XXXXXXXXXXXX-	12	10	22
66	05LJ012	VERMILION R	-------0000000000XXXXXXXXXXXX-	12	10	22
67	05LJ015	FISHING R	-------0000000000XXXX0XXXXXXX-	11	11	22
68	05LJ016	FORK R	-------000000000XXXXXXXXXXXXX-	13	9	22
69	05LJ017	DRIFTING R	--00--0000000000XXXXXXXXXXXXX-	13	12	25
70	05LJ019	MINK R	-000000000000000XXXXXXXXXXXXX-	13	15	28
71	05LJ022	EDWARDS CK	-------0000000000XXXXXXXXXXXX-	12	10	22
72	05LL007	PINE CK	-0-000-----------00XXXXXXXXXX-	10	6	16
73	05LL009	NEEPAWA CK	-------000000000000XXXXXXXXXX-	10	12	22
74	05LL013	WHITEMUD R	---------000000000000XXXXXXXX-	8	12	20
75	05MB001	YORKTON CK	-X-XXX-0000000000XXXXXXXXXXXX-	16	10	26
76	05MB004	WHITESAND R	-------0000000000XX0X0XX0XXXX-	9	13	22
77	05MC001	ASSINIBOINE R	-------0000000000XXXXXXX0XXXX-	11	11	22
78	05MC002	STONY CK	--00--00000000000XXXX0XXXXXXX-	11	14	25
79	05MD005	SHELL R	-------0000000000XXXXXX0XXXXX-	11	11	22
80	05MD006	LITTLE BOGGY CK	-------0000000000XXXXXXXXXXXX-	12	10	22
81	05ME003	BIRDTAIL CK	-------0000000000XXXXXXXXXXXX-	12	10	22

* THE SYMBOL X INDICATES YEARS WITH ACTUAL MEASURED RECORDS AND THE SYMBOL O INDICATES YEARS FOR WHICH RECORDS HAVE BEEN EXTENDED BY CORRELATION ANALYSIS

TABLE 2-1 CONTINUED

	STATION ID		YEARS	ACTUAL YEARS	EXTENDED YEARS	TOTAL YEARS
			1940 1969			
82	05MF001	MINNEDOSA R	-------000000000000XXXXXXXXXX-	10	12	22
83	05MF008	ROLLING R	-------00000000000000XXXXXXXX-	8	14	22
84	05MG001	ARROW R	-------000000000000XXXXXXXXXX-	10	12	22
85	05MG002	BOSSHILL CK	-------000000000000XXXXXXXXXX-	10	12	22
86	05MG003	GOPHER CK	-------000000000000XXXXXXXXXX-	10	12	22
87	05MG004	OAK R	-------000000000000XXXXXXXXXX-	10	12	22
08	05MH006	LITTLE SOURIS R	-------------------00XXXXXXXX-	8	2	10
89	05MH007	EPINETTE CK	-------------------0000XXXXXX-	6	4	10
90	05MJ004	STURGEON CK	-------------------000XXXXXXX-	7	3	10
91	05NA005	GIBSON CK	-------000000000000XXXXXXXXXX-	10	12	22
92	05NB011	YELLOWGRASS DITC	-------0000000000XXXXXXXXXXXXX	13	10	23
93	05ND014	JEWEL CK	-------000000000000XX0XXXXXX0-	8	14	22
94	05NB019	COULEE WEST	-----------------000XXX00XXXX-	7	5	12
95	05NE003	PIPESTONE CK	-------0000000000000XXXXXXXXX-	9	13	22
96	05NF002	ANTLER R	-------000000000XXXXXX0XXXXXX-	12	10	22
97	05NF007	GAINSBOROUGH CK	----------------0XXXX000XXXXXX	10	4	14
98	05NF008	GRAHAM CK	----------------00XXX00XXXXXX0	9	5	14
99	05NG010	OAK CK	-------------------00XXXXXXXX-	8	2	10
100	05NG012	ELGIN CK	-------------------00XXXXXXX0-	7	3	10
101	050A006	WHITEMUD CK	---------0000000000XXXXXXXXXX-	10	10	20
102	050AC07	BADGER CK	---------0000000000XXXXXXXXXX-	10	10	20
103	050A008	PEMBINA	-------000000000000XXXXXXXXXX-	10	12	22
104	050A009	WAKOPA	---------0000000000000XXXXXXX-	7	13	20
105	050B006	CRYSTAL CK	-------------------XXXX0XXXXX-	9	1	10
106	050B010	LONG R	-------000000000000XXXXXXXXXX-	10	12	22
107	050B016	SNOWFLAKE CK	---------000000000000XXXXXXXX-	8	12	20
108	050B021	MOWBRAY CK	---------0000000000000XXXXXXX-	7	13	20
109	050D001	ROSEAU R	---------XXXXXXXXXXXXXXXXXXXX-	20	0	20
110	050D004	ROSEAU R	---------00000000000000XXXXXX-	6	14	20
111	050D014	ROSEAU R	---------0000000000000XXXXXXX-	7	13	20
112	050E004	RAT R	---------00000000000XXXXXXXXX-	9	11	20
113	050FC15	SHANNON CK	---------00000000000XXXXXXXXX-	9	11	20
114	050G004	ELM CK 1	-------------------0XXX0XXXXX-	8	2	10
115	C5CG005	ELM CK 2	-------------------0XXXXXXXXX-	9	1	10
116	050G006	ELM CK 3	-------------------0XXX0XXXXX-	8	2	10
117	050J002	COOKS CK	---------000000000XXXXXXXXXXX-	11	9	20
118	050J006	COOKS CK	---------00000000000XXXXXXXXX-	9	11	20
119	050J008	NETLEY CK	----------------0000XXXXXXXXX-	9	4	13
120	05CJ009	NETLEY CK	-----------------000XXXX0XXXX-	8	4	12
121	05PH003	WHITEMOUTH R	---------0000000XXXXXXXXXXXXX-	13	7	20
122	05SA002	BROKENHEAD R	---------0000000XXXXXXXXXXXX0-	12	8	20
123	05SA004	BROKENHEAD R	---------000000000000XXXXXXXX-	8	12	20
124	05SB002	OSIER CK	-----------------000XXX000XXX-	6	6	12
125	C5SC002	ICELANDIC R	-----------------00XXXXXXXXXX-	10	2	12
126	05SD003	FISHER R	-----------------0000XXXXXXXX-	8	4	12
127	05SD004	EAST FISHER R	-------------------0000XXXXXX-	6	4	10
128	06AD001	BEAVER R	----0000000000000000000XXXXXX-	6	19	25
129	06ADC06	BEAVER R	0000000000000000XXXXXXXXXXXXXX	14	16	30
130	07AF002	MCLEOD R	000000000000000XXXXXXXXXXXXXXX	15	15	30
131	07AG001	MCLEOD R	000000000000000000XXXXXXXXX000	9	21	30
132	07AGC03	WOLF CK	000000000000000XXXXXXXXXXXXXXX	15	15	30
133	07EA001	PEMBINA R	000000000000000000XXXXXXXXXX00	10	20	30
134	07BB002	PEMBINA R	000000000000000XXXXXXXXXXXXXXX	15	15	30
135	07BB003	LOBSTICK R	000000000000000XXXXXXXXXXXXXXX	15	15	30
136	07BB004	PADDLE R	00000000000000000000000XXXXXXX	7	23	30
137	07BB005	LITTLE PADDLE R	00000000000000000000000XXXXXXX	7	23	30
138	07BC002	PEMBINA R	000000000000000000XXXXXXXXXXXX	12	18	30
139	07BF001	EAST PRAIRIE R	0000000000000000000XXXXXXXXXXX	11	19	30
140	07BF002	WEST PRAIRIE R	0000000000000000000XXXXXXXXXXX	11	19	30
141	07BJ001	SWAN R	0000000000000000000000XX0XXXXX	7	23	30
142	07GE001	WAPITI R	---------------000000XXXXXXXXX	9	6	15
143	C7GH002	LITTLE SMOKY R	00000000000000000000XXX0XXXXXX	9	21	30
144	07HA003	HEART R	-------------------0000XXXXXXX	7	4	11
145	C7HC001	NOTIKEWIN R	----------------------XXXXXXX-	7	0	7
146	07JF002	BOYER R	-----------------------XXXXXXX	7	0	7
147	07JF003	PONTON R	-----------------------XXXXXXX	7	0	7
148	11AA026	SAGE CK	XX--X-XXXX00XXXXXXXXXXXXXXXXX-	24	2	26
149	11AB009	MIDDLE CK	0000000000XXXXXXXXXXXXXXXXXXX-	19	10	29
150	11AB075	LYONS CK	XXXXXXXX00XXX0XXXXXXXXXX0XXXX-	25	4	29
151	11ABC76	BATTLE CK	XXXXXXXXXXXXXXXXXXXXXXXXXXX00-	27	2	29
152	11ABC82	LODGE CK	00--0-00000XXXXXXXXX0XXXXXXXX-	17	9	26
153	11ABC87	MIDDLE CK	000000000000000000000XXXXXX0X-	7	22	29
154	11AB105	WOODPILE COULEE	XXXXXXXX00XXXXXXXXXXXXXX0XXXX-	26	3	29
155	11AB107	EAST BATTLE CK	XXXXXXXX00XXXXXXXXXXXXXX0XXXX-	26	3	29
156	11AD001	WHITEWATER CK	X-XXXXXXXXXXXXXXXXXXXXXXXXXXX-	28	0	28
157	11AE002	POPLAR R W	XXXXXXXXXXXXX0000000000000000-	13	16	29
158	11AE003	POPLAR R E	XXXX-XXXXXXXXXXXXXXXXXXXXXXXX-	28	0	28
159	11AE005	ROCK CK	XXXXXXXXXXXXXXXXXXXXXX0000X00-	23	6	29
160	11AE008	POPLAR R M	XXXXXXXXXXX0XXXXXXXXXXXXXXXXX-	28	1	29
161	11AE009	ROCK CK	00000000000000000XXXXXXXXX0XX-	11	18	29

period, it has been assumed that the annual flood is contained in the existing record. Where the missing data seemed to constitute a period of high flow, no annual flood has been recorded for that year.

Table 2-2 has been prepared to indicate for each study basin, the years for which annual flood data have been compiled. The "X" symbols denote the years for which measured data are available; and the total years of measured data are listed in the column headed "ACTUAL YEARS."

2.2.3.1 The Seasonal Distribution of Flood Flows

The seasonal distribution of flood flows was examined with the intention of considering the homogeneity of the flood series. In any frequency analysis, it is assumed that the data set being analysed is homogeneous, that is, it is representative of a single population. Previous work on prairie flood flow analyses has identified three distinct types of flood (Durrant 1959, p. 96). The three types are identified on the basis of causal factors, the first type being the result of snowmelt, the second type being the result of snowmelt and rainfall combined and the third type being the result of rainfall. The examination of the seasonal distribution of flood flows was intended to provide a basis for some general conclusions regarding the relative importance of the three flood types. For each of the 161 study basins, the flood flow data series was examined and a tally of the flood events occurring in each month compiled. The pattern was generally consistent from basin to basin and may be illustrated by the combined tally for all basins as reproduced in Table 2-3. Data in the table indicate that approximately 90% of the peaks included in the annual flood series occurred during the spring months from March to June with a definite peak in April. Floods are most likely to be of the first two types resulting from snowmelt, and snowmelt in combination with rainfall. The remaining 10% of the annual peaks were recorded during the summer and fall season from July to October. These events are most likely of the third type resulting from rainfall. Thus in the present study ,no attempt has been made to separate out types of flood events. The inclusion of Table 2-3 is intended only to provide some indication of the relative importance of summer and fall peaks. In the present analysis, the records have been treated as if drawn from a single homogeneous population.

2.2.4 Test of Serial Correlation in the Study Basin Data Series

One of the assumptions basic to any frequency analysis is that the sample data are a series of independent events. Some hydrologic data series such as daily flows do not meet this assumption; however, other series in which the basic time interval is longer, as in the case of annual yields and annual flood peaks, are often assumed for analysis purposes to satisfy the assumption of independent events.

A simple measure of the degree of dependence between successive events in a sample data series is provided by serial correlation coefficients. In the present study ,serial correlation coefficients based on a lag of one year were computed for each of the 161 study basins for both the annual yield and the annual flood flow data. The serial correlation coefficients were then tested for significance. Results of the computations

TABLE 2-2

ANNUAL PEAK RECORDS *

	STATION ID		YEARS 1940 1969	ACTUAL YEARS	EXTENDED YEARS	TOTAL YEARS
1	05AB021	WILLOW CK	----XXXXXXXXXXXXXXXXXXXXXXXXX-	25	0	25
2	05AE002	LEE CK	XXXXXXXXXXXXXXXXXXXXXXXXXXXXX-	29	0	29
3	05AE005	ROLPH CK	XXXXXXXXXXXXXXXXXXXXXXXXXXXXX-	29	0	29
4	05AF010	MANYBERRIES CK	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	30	0	30
5	05AH002	MACKAY CK	0000000000000000XXXXX0XXXXXXXX	13	17	30
6	05AH041	PEIGAN CK	00000000000000000000XXXXXXXXXX	10	20	30
7	05BJ004	ELBOW R	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	30	0	30
8	05BJ005	ELBOW R	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	30	0	30
9	05BK001	FISH CK	0000000000000000XXXXXXXXXXXXXX	14	16	30
10	05BL007	STIMSON CK	XXXXXXXXXXXXXXXXXXXXXXXXXXXX0-	28	1	29
11	05CB001	LITTLE RED DEER	------------000000000XXXXXXXXX	9	9	18
12	05CB002	LITTLE RED DEER	000000000000000000000000XXXXXX	6	24	30
13	05CC001	BLINDMAN R	0000000000000000000000XXXXXXXX	8	22	30
14	05CC007	MEDICINE R	0000000000000000000000XXXXXXXX	8	22	30
15	05CE002	KNEEHILLS CK	0000000000000000000XXXXXXXXXXX	11	19	30
16	05CE006	ROSEBUD R	0000000000000000000XXXXXXXXXXX	11	19	30
17	05CG002	BULLPOUND CK	----000000000000000000XXXXXXXX	8	18	26
18	05CK005	ALKALI CK	00000000000000000000000XXXXXXX	7	23	30
19	05CB002	PRAIRIE CK	------------XXXXXXXXXXXXXXXXXX	18	0	18
20	05EA001	STURGEON R	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	30	0	30
21	05EE001	VERMILION R	----000000000000000XXXXXXXXXXX	11	15	26
22	05FD001	RIBSTONE CK	----000000000000000000XXXXXXXX	8	18	26
23	05FE001	BATTLE R	----XXXXXXXXXXXXXXXXXXXXXXXXX-	25	0	25
24	05GA003	MONITER CK	----0000000000XXXXXXXXXXXXXXXX	16	10	26
25	05GC005	EAGLE CK	----000000000000000000XXXXXXX-	7	18	25
26	05GC006	EAGLE CK	----0000000000000000000XXXXXX-	6	19	25
27	05HA003	BEAR CK	00000000000000000000000XXXXXXX	7	23	30
28	05HA015	BRIDGE CK	00000000000000000000000XXXXXX-	6	23	29
29	05HA062	PIAPOT CK	00000000000000000000000XXXXXX-	6	23	29
30	05HA075	SKULL CK	00000000000000000000000XXXXXXX	7	23	30
31	05HD036	SWIFTCURRENT CK	000000000000000XXXXXXXXXXX0XX-	13	16	29
32	05HG002	BRIGHTWATER CK	----0000000000000000X0XXXXXXX-	8	17	25
33	05JA002	WOOD R	----XXXXXXXXXXXXXXXXXXXXXXXXX-	25	0	25
34	05JB003	NOTUKEU CK	X---XXXXXXXXXXXXXXXXXXXXXXXXX-	26	0	26
35	05JE004	MOOSE JAW R	----XXXXXXXXXXXXXXXXXXXXXXXXX-	25	0	25
36	05JE006	MOOSE JAW R	----XXXXXXXXXXXXXXXXXXXXXXXXX-	25	0	25
37	05JF005	WASCANA CK	-----XXXXXXXXXXXXXXXXXXXXXXXX-	24	0	24
38	05JF006	BOGGY CK	-----XXXXXX00XXXXXXXXXXXX000X-	19	5	24
39	05JC004	QU'APPELLE R	----XXXXXXXXXXXXXXXXXXXXXXXX0-	24	1	25
40	05JH001	ARM R	----000000000XXXXXXXXXXXXXXXX-	16	9	25
41	05JK004	JUMPING DEER CK	-XXXXXXXXXXXXXXXXXXXXXXXXXXXX-	28	0	28
42	05JL002	INDIANHEAD CK	-XXX00XXXXX0XXXXXXXXX0XXXXXXX-	24	4	28
43	05JL005	PHEASANT CK	-000000XXXXXXXXXXXXXXXXXXXXXX-	22	6	28
44	05JM015	CUTARM CK	-000--0000000000XXXXXXXXXXXXX-	13	13	26
45	05KA001	CARROT R	-000000000000000XXXXXXXXXXXXXX	14	15	29
46	05KB003	CARROT R	-00000000000000XXXXXXXXXXXXXXX	15	14	29
47	05KC001	CARROT R	-00000000000000XXXXXXXXXXXXXXX	15	14	29
48	05KD002	PETAIGAN R	-00000000000000XXXXXXX0000000-	7	21	28
49	05KE002	TORCH R	----0000000XXXXXXXXXXXXXXXXXX-	18	7	25
50	05LB002	ETOMAMI R	----00000000000XXXXXXXXXXXXXX-	14	11	25
51	05LC001	RED DEER R	----0000000000XXXXXXXXXXXXXXX-	15	10	25
52	05LC004	RED DEER R	----000000000000XXXXXXXXXXXXX-	13	12	25
53	05LD001	OVERFLOWING R	----000000000000XXXXXXXXXXXXX-	13	12	25
54	05LE001	SWAN R	----0000000XXXXXXXXXX00000000-	10	15	25
55	05LE003	BIRCH R	----0000000000XXXXXXXXXXXXXXX-	15	10	25
56	05LE004	WOODY R	----0000000000XXXXXXXXXXXXXXX-	15	10	25
57	05LE005	ROARING R	-000000000000000000XXXXXXXXXX-	10	18	28
58	05LE006	SWAN R	----00000000000000000XXXXXXXX-	8	17	25
59	05LF001	STEEP ROCK R	-0000000000000XXXXXXXXXXXX0XX-	14	14	28
60	05LF002	BELL R	-0000000000000XXXXXXXXXXX0XXX-	14	14	28
61	05LG001	PINE R	-0000000000000XXXXXXXXXXXXXXX-	15	13	28
62	05LG002	GARLAND R	-0000000000000XXXXXXXX0XXXXXX-	14	14	28
63	05LJ005	OCHER R	-00000000XXXXXXXXXXXXXXXXXXXX-	20	8	28
64	05LJ007	TURTLE R	-00000000XXXXXXXXXXXXXXXXXXXX-	20	8	28
65	05LJ011	WILSON R	-00000000XXXXXXXXXXXXXXXXXXXX-	20	8	28
66	05LJ012	VERMILION R	-0000000XXXXXXXXXXXXXXXXXXXXX-	21	7	28
67	05LJ015	FISHING R	----0000XXXXXXXXXXXXX0XXXXXXX-	20	5	25
68	05LJ016	FORK R	-0000000000000XXXXXXXXXXXXXXX-	15	13	28
69	05LJ017	DRIFTING R	-0000000000000XXXXXXXXXXXXXXX-	15	13	28
70	05LJ019	MINK R	-000000000000000XXXXXXXXXXXXX-	13	15	28
71	05LJ022	EDWARDS CK	-0000000000000000XXXXXXXXXXXX-	12	16	28
72	05LL007	PINE CK	-000000000000000000XXXXXXXXXX-	10	18	28
73	05LL009	NEEPAWA CK	0000000000000000000XXXXXXXXXX-	10	19	29
74	05LL013	WHITEMUD R	000000000000000000000XXXXXXXX-	8	21	29
75	05MB001	YORKTON CK	-XXXXX0X0X0XXXXXXXXXXXXXXXXXX-	25	3	28
76	05MB004	WHITESAND R	-0000000000000000XX0X0XX0XXXX-	9	19	28
77	05MC001	ASSINIBOINE R	-000XXXXXXXXXXXXXXXXXXXX0XXXX-	24	4	28
78	05MC002	STONY CK	----0000000000000XXXX0XXXXXXX-	11	14	25
79	05MD005	SHELL R	-0000000XXXXXXXXXXXXXXX0XXXXX-	20	8	28
80	05MD006	LITTLE BOGGY CK	-0000000000000000XXXXXXXXXXXX-	12	16	28
81	05ME003	BIRDTAIL CK	-0000000000000XXXXXXXXXXXXXXX-	15	13	28

* THE SYMBOL X INDICATES YEARS WITH ACTUAL MEASURED RECORDS AND THE SYMBOL 0 INDICATES YEARS FOR WHICH RECORDS HAVE BEEN EXTENDED BY CORRELATION ANALYSIS

TABLE 2-2 CONTINUED

	STATION ID		YEARS 1940 1969	ACTUAL YEARS	EXTENDED YEARS	TOTAL YEARS
82	05MF001	MINNEDOSA R	----OOOOOOOOOOOOOOOXXXXXXXXXX-	10	15	25
83	05MF008	ROLLING R	----OOOOOOOOOOOOOOOOOXXXXXXXX-	8	17	25
84	05MG001	ARROW R	---OOOOOOOOOOOOOOOOXXXXXXXXXXO	10	17	27
85	05MG002	BOSSHILL CK	---OOOOOOOOOOOOOOOOXXXXXXXXXXO	10	17	27
86	05MG003	GOPHER CK	----OOOOOOOOOOOOOOOXXXXXXXXXX-	10	15	25
87	05MG004	OAK R	----OOOOOOOOOOOOOOOXXXXXXXXXX-	10	15	25
88	05MH006	LITTLE SOURIS R	---OOOOOOOOOOOOOOOOOOXXXXXXXX-	8	18	26
89	05MH007	EPINETTE CK	OOOOOOOOOOOOOOOOOOOOOXOXXXXXX-	7	22	29
90	05MJ004	STURGEON CK	OOOOOOOOOOOOOOOOOOOOOXXXXXXXX-	8	21	29
91	05NA005	GIBSON CK	-OOOOOOOOOOOOOOOOOOXXXXXXXXXX-	10	18	28
92	05NB011	YELLOWGRASS DITC	OOOO-OOOOOOOOOOOOXXXXXXXXXXXXX	13	16	29
93	05NB014	JEWEL CK	OOOO-OOOOOOOOOOOOOOXXOXXXXXXO-	8	20	28
94	05NB019	COULEE WEST	---OOOOOOOOOOOOOOOOOXXOOXXXXX-	7	19	26
95	05NE003	PIPESTONE CK	---OOOOOOOOOOOOOOOOOXXXXXXXXXO	9	18	27
96	05NF002	ANTLER R	---XXXXXXXXXXXXXXXXXXXOXXXXXX-	25	1	26
97	05NF007	GAINSBOROUGH CK	---OOOOOOOOOOOOOXXXXXOOOXXXXXX	11	16	27
98	05NF008	GRAHAM CK	---OOOOOOOOOOOOOOOXXOOXXXXXXXO	9	18	27
99	05NG010	OAK CK	---OOOOOOOOOOOOOOOOOOXXXXXXXX-	8	18	26
100	05NG012	ELGIN CK	-OOOOOOOOOOOOOOOOOOOOXXXXXXXO-	7	21	28
101	05OA006	WHITEMUD CK	---OOOOOOOOOOOOOOOOXXXXXXXXXX-	10	16	26
102	05OA007	BADGER CK	---OOOOOOOOOOOOOOOOXXXXXXXXXX-	10	16	26
103	05OA008	PEMBINA	---OOOOOOOOOOOOOOOOXXXXXXXXXXO	10	17	27
104	05OA009	WAKOPA	---OOOOOOOOOOOOOOOOOOOXXXXXXX-	7	19	26
105	05OB006	CRYSTAL CK	---OOOOOOOOOOOOOOOOXXXXOXXXXX-	9	17	26
106	05OB010	LONG R	---OOOOOOOOOOOOOOOOXXXXXXXXXXO	10	17	27
107	05OB016	SNOWFLAKE CK	OOOOOOOOOOOOOOOOOOOOOXXXXXXXX-	8	21	29
108	05OB021	MOWBRAY CK	OOOOOOOOOOOOOOOOOOOOOOXXXXXXX-	7	22	29
109	05OD001	ROSEAU R	XXXXXXXXXXXXXXXXXXXXXXXXXXXXX-	29	0	29
110	05OD004	ROSEAU R	OOOOOOOOOOOOOOOOOOOOOOXXXXXXX-	7	22	29
111	05OD014	ROSEAU R	OOOOOOOOOOOOOOOOOOOOOOXXXXXXX-	7	22	29
112	05OE004	RAT R	---OOOOOOOOOOOOOOOOOXXXXXXXXX-	9	17	26
113	05OF015	SHANNON CK	OOOOOOOOOOOOOOOOOOOOXXXXXXXXX-	9	20	29
114	05OG004	ELM CK 1	OOOOOOOOOOOOOOOOOOOOXXXOXXXXX-	8	21	29
115	05OG005	ELM CK 2	OOOOOOOOOOOOOOOOOOOOXXXXXXXXX-	9	20	29
116	05OG006	ELM CK 3	OOOOOOOOOOOOOOOOOOOOXXXOXXXXX-	8	21	29
117	05OJ002	COOKS CK	---OOOOOOOOOOOOOOXXXXXXXXXXXX-	12	14	26
118	05OJ006	COOKS CK	---OOOOOOOOOOOOOOOOOXXXXXXXXX-	9	17	26
119	05OJ008	NETLEY CK	---OOOOOOOOOOOOOOOOOXXXXXXXXX-	9	17	26
120	05OJ009	NETLEY CK	---OOOOOOOOOOOOOOOOOXXXXXXXXX-	9	17	26
121	05PH003	WHITEMOUTH R	---XXXXXXXXXXXXXXXXXXXXXXXXXX-	26	0	26
122	05SA002	BROKENHEAD R	---XXXXXXXXXXXXXXXXXXXXXXXXXX-	26	0	26
123	05SA004	BROKENHEAD R	---OOOOOOOOOOOOOOOOOXXXXXXXXX-	9	17	26
124	05SB002	OSIER CK	---------OOOOOOOOOOOXXXXOXXXX-	8	12	20
125	05SC002	ICELANDIC R	---OOOOOOOOOOOOOOOOXXXXXXXXXX-	10	16	26
126	05SD003	FISHER R	---OOOOOOOOOOOOOOOOOOXXXXXXXX-	8	18	26
127	05SD004	EAST FISHER R	---------OOOOOOOOOOOOXXXXXXXX-	8	12	20
128	06AD001	BEAVER R	OOOOOOOOOOOOOOOOOOOOXXXXXXXXXO	9	21	30
129	06AD006	BEAVER R	----OOOOOOOOOOOOXXXXXXXXXXXXXX	14	12	26
130	07AF002	MCLEOD R	------------OOOXXXXXXXXXXXXXXX	15	3	18
131	07AG001	MCLEOD R	OOOOOOOOOOOOOOOOOOXXXXXXXXXXOO	10	20	30
132	07AG003	WOLF CK	------------OOOXXXXXXXXXXXXXXX	15	3	18
133	07BA001	PEMBINA R	------------OOOOXXXXXXXXXXXXOO	12	6	18
134	07BB002	PEMBINA R	------------OOOXXXXXXXXXXXXXXX	15	3	18
135	07BB003	LOBSTICK R	OOOOOOOOOOOOOOOXXXXXXXXXXXXXXX	15	15	30
136	07BB004	PADDLE R	OOOOOOOOOOOOOOOOOOOOOOOXXXXXXX	7	23	30
137	07BB005	LITTLE PADDLE R	OOOOOOOOOOOOOOOOOOOOOOOXXXXXXX	7	23	30
138	07BC002	PEMBINA R	OOOOOOOOOOOOOOOOOOXXXXXXXXXXXX	12	18	30
139	07BF001	EAST PRAIRIE R	---------------OOOOXXXXXXXXXXX	11	4	15
140	07BF002	WEST PRAIRIE R	OOOOOOOOOOOOOOOOOOOXXXXXXXXXXX	11	19	30
141	07BJ001	SWAN R	----------------------XX-XXXXX	7	0	7
142	07GE001	WAPITI R	-------------------OOXXXXXXXXX	9	2	11
143	07GH002	LITTLE SMOKY R	OOOOOOOOOOOOOOOOOOOOXXXOXXXXXX	9	21	30
144	07HA003	HEART R	-------------------OOOOXXXXXXX	7	4	11
145	07HC001	NOTIKEWIN R	---------------------XXXXXXXXX	9	0	9
146	07JF002	BOYER R	----------------------XXXXXXXX	8	0	8
147	07JF003	PONTON R	-----------------------XXXXXXX	7	0	7
148	11AA026	SAGE CK	XXXXXXXXXXXXXXXXXXXXXXXXXXXXX-	29	0	29
149	11AB009	MIDDLE CK	OOOOOOOOOOXXXXXXXXXXXXXXXXXXX-	19	10	29
150	11AB075	LYONS CK	XXXXXXXOOXXXOXXXXXXXXXXXOXXXX-	25	4	29
151	11AB076	BATTLE CK	XXXXXXXXXXXXXXXXXXXXXXXXXXXOO-	27	2	29
152	11AB082	LODGE CK	OOOOOOOOOOOXXXXXXXXXXXXXXXXXX-	18	11	29
153	11AB087	MIDDLE CK	OOOOOOOOOOOOOOOOOOOOOXXXXXXXX-	8	21	29
154	11AB105	WOODPILE COULEE	XXXXXXXOOXXXXXXXXXXXXXXXOXXXX-	26	3	29
155	11AB107	EAST BATTLE CK	XXXXXXXOOXXXXXXXXXXXXXXXOXXXX-	26	3	29
156	11AC001	WHITEWATER CK	X-XXXXXXXXXXXXXXXXXXXXXXXXXXX-	28	0	28
157	11AE002	POPLAR R W	XXXXXXXXXXXXXOOOOOOOOOOOOOOOO-	13	16	29
158	11AE003	POPLAR R E	XXXX-XXXXXXXXXXXXXXXXXXXXXXXX-	28	0	28
159	11AE005	ROCK CK	XXXXXXXXXXXXXXXXXXXXXXOOOOXOO-	23	6	29
160	11AE008	POPLAR R M	XXXXXXXXXXXXXXXXXXXXXXXXXXXXX-	29	0	29
161	11AE009	ROCK CK	OOOOOOOOOOOOOOOOOXXXXXXXXXOXX-	11	18	29

TABLE 2-3 - SEASONAL DISTRIBUTION OF RECORDED ANNUAL FLOOD FLOW COMBINING RECORDS FOR ALL STUDY BASINS

Month	Annual Flood Events	
	Occurrences	% of Total
March	412	17.5
April	1033	44.0
May	388	16.5
June	284	12.1
July	138	5.9
August	63	2.7
September	22	0.9
October	8	0.3
TOTAL	2348	

TABLE 2-4 - SUMMARY OF THE RESULTS OF THE SERIAL CORRELATION ANALYSIS FOR THE ANNUAL YIELD AND ANNUAL FLOOD PEAK DATA SERIES (Lag = 1 year)

Data Series	No. of Records Tested	Significance of Serial Correlation Coefficients (Lag = 1 year)	
		No. Significant at 5% Level	No. Significant at 1% Level
Annual Yield Data	161	5	3
Annual Flood Flow Data	161	5	2

and significance tests are presented in summarized form in Table 2-4. In all cases, the serial correlation coefficients for a one year lag were weak. Of the 161 annual yield data series tested, only five records resulted in a significant serial correlation coefficient at the 5% significance level and only three were significant at the 1% significance level. Of the 161 annual flood flow data series tested, only five records resulted in significant correlation coefficients at the 5% significance level and only two were significant at the 1% significance level. On the basis of these results, the assumption was confirmed that both the annual yield data series and the annual flood flow data series for the study basins approximate the assumption of independent events.

2.2.5 Extension of Available Records by Correlation and Regression

The frequency analysis of any series of hydrologic data is based on the assumption that the past records represent a sample of random independent events drawn from a very large or infinite population. The available data series usually comprise relatively small samples; and the results of the analysis are sensitive to the chance inclusion in the sample of exceptionally large or small events. The best solution to this problem of variability in short records lies in the compilation of longer data series; however, the time consuming nature of hydrologic data collection usually precludes this solution and it remains for the individual researcher to make the most efficient use possible of the available data. An alternative solution to the problem is in the use of correlation and regression analysis to extend short records on the basis of their relationships to other nearby stations for which longer records are available.

In regional frequency analyses, where frequency curves from several stations are being compared, there is the added consideration of the comparability of curves based on short records of variable length. If a common base period can be defined, the time variability can be minimized and the effects of other factors more easily analysed (Dalrymple 1960, p. 33). Dalrymple (1960) described the use of extended records in a regional analysis of flood frequencies. His method does not employ the extended data directly in the analysis, but rather, uses them to adjust the order numbers assigned to the measured peaks with respect to a selected base period. The base period chosen is normally as long as possible and an attempt is made to extend and fill in all records for the full period.

Langbein (1960, p. 28) suggests several alternative approaches to streamflow record extension by correlation and regression. The simplest approach is that discussed above in which a short-term record is extended by correlation with a long-term record from a nearby station. Other approaches include correlation with long-term precipitation records and multiple correlation with both long-term streamflow and precipitation records.

In the present study, it was considered desirable to employ comparable data sets. The actual data records available for the selected study basins are comparable neither in terms of length nor in terms of the time periods covered (see Tables 2-1 and 2-2 for available actual data records). In order to provide more comparable data series for the study basins, an attempt was made to extend short records and fill in missing records by correlation with the available long-term stations. It was intended that these record extensions would result in the full 30 year record, 1940-1969, for both annual yields and annual flood flows, for each of the study basins. Unfortunately, the limited number of long-term (30-year) records available precluded this aim; and instead, each of the station records of both annual yields and annual flood flows was extended as much as possible up to the full 30 year period. The extended data series provide the basic data sets employed in the frequency analyses for each of the study basins.

Record extension in this study was accomplished by the use of simple linear regression analysis relating the short-term record or dependent variable, to the long-term record or independent variable. The regression equations calculated were based on either the original arithmetic data series or their logarithmic transformations, whichever provided the strongest relationship. The long-term station employed in each case was chosen on the basis of proximity to the station for which the records were being extended, the length of record available, and the strength of the relationship as indicated by the significance level of the correlation coefficient.

The years for which records were extended have been indicated by the "0" symbol in Tables 2-1 and 2-2. Unfortunately, only a small number of the study basins had long enough original records to serve as base stations in the record extension. It was therefore necessary to utilize correlations between widely separated stations. Both the long distances and the small number of suitable base stations would be expected to limit validity of the data extensions. These limitations were recognized and the estimated data were not employed directly in the frequency analyses as discussed in the following section of this chapter; rather, the estimated events were utilized only in adjusting the plotting positions of the actual data.

2.3 SINGLE STATION FREQUENCY ANALYSIS OF THE ANNUAL YIELD AND ANNUAL FLOOD FLOW DATA SERIES

Frequency analysis techniques are a means of analysing the variability of a data sample for the purpose of estimating the population variability. All hydrologic data vary with time; however, this variation is not usually sufficiently regular to be considered as cyclic (Leopold 1959). In the absence of regular variations, it is not possible to employ past records as a basis for forecasting future events. It is, however, possible to employ such historical records as indications of the probabilities of occurrence of future events of given magnitudes. Frequency analysis techniques as applied in the present study are intended to evaluate the variability of the streamflow characteristics under consideration. In general terms, frequency analysis examines the relationship between the magnitude of a variant and its frequency or probability of occurrence (Riggs 1968, p. 1). This relationship is analysed by the fitting of a frequency curve to the sample data series. The frequency curve is fitted so as to estimate the frequency distribution of the population from which the sample has been drawn.

Two basic assumptions underlie the application of frequency analysis to hydrologic data series. The first assumption requires that the data represent discrete independent random events, while the second assumption assumes time-stationarity of the processes. In the present research, the annual yield and annual flood flow data series were assumed to approximate random conditions. A test of the serial correlations with a lag of one year confirmed the validity of this assumption. Elimination of all basins for which natural flow conditions did not exist was intended to help in meeting the second assumption. No attempt has been made to evaluate natural changes in the basins which may upset this assumption.

The literature dealing with the application of frequency techniques to hydrologic data is abundant but scattered through the journals of numerous disciplines. While it is beyond the scope of this report to deal with the subject in detail, the reader is referred to the work of Chow (1964a) and Riggs (1968) which contain a useful review and bibliography on the subject.

2.3.1 Methodology of Frequency Analysis

The frequency analysis of the available annual yield and annual flood flow data series for each of the study basins has been based on the fitting of a 2-parameter lognormal distribution by least squares regression techniques. These analyses resulted in the estimation of the selected dependent variables, the mean annual and 10-year events. The author tested several 2-parameter distributions and found that the 2-parameter lognormal distribution resulted in the best fit for both the annual yield and annual flood flow data series for the study basins. Use of the least square regression technique to fit the frequency curves provided an objective method which was easily adapted to computer calculations and the use of extended data series.

Prior to the actual curve fitting by regression analysis, the magnitudes of the events were transformed to logarithms and the probabilities of occurrence were transformed into a rectangular scaled reduced variate "Y" corresponding to a normal probability scale. This latter transformation was made with the aid of a computer programme function FNUPR written by Cooper and Howells (1969). The equations which resulted from the regression analysis were of the form:

$$X = a + bY$$

where: X is the logarithm to the base ten of the event magnitude,
Y is the reduced variable corresponding to the normal probability scale,
a and b are constants.

These equations may be employed to estimate the magnitude of an event with any given probability of occurrence. In this study, the equations were utilized in the estimation of the magnitudes of the mean annual and 10-year events.

Since the frequency analyses were based on the fitting of a straight line to the lognormal plot of the data, it was necessary only to calculate two points on the line in order to reproduce the entire distribution.

One of the characteristics of the lognormal distribution is that the magnitude of the event with the probability of occurrence of 0.5 or a 2-year return period is the mean of the distribution. This value is a more stable and dependable estimate of the population mean, than is the arithmetic mean (Benson 1960). In the present research this value was defined as the "mean annual event." The second value estimated for each of the frequency curves was the event with the return period of ten years or 0.1 probability of occurrence.

In the present study, a computer programme was prepared to fit the 2-parameter lognormal distribution to a sample data series, and to calculate the mean annual and 10-year events. The major steps in the programme are as follows;

I) A complete data series including both actual measured observations and those estimated by regression is read and sorted into descending order. Each event is assigned an order number, the largest event being 1.

II) Plotting positions are assigned to all actual measured events by the formula:

$$T = \frac{(n - 1)}{m}$$

where: T is the return period in years,
n is the total number of events in the series,
m is the order number of the event.

From T the probability of occurrence P is estimated by the formula:

$$P = 1/T.$$

III) Magnitudes of the events are transformed by taking their logarithms to the base 10. The resulting values are the variates "X."

IV) The probabilities of occurrence "P" are transformed into values of the reduced variate "Y" corresponding to a normal probability scale.

V) The least squares regression equation is calculated for the regression of Y on X.

VI) The regression equation is utilized in the estimation of the mean annual and 10-year events which correspond to P values of 0.500 and 0.100 and Y values of 0.00 and 1.28 respectively.

Tables 2-5 and 2-6 are examples of the computer output resulting from the processing of the annual yield data series for basin number 4, Manyberries Creek. Table 2-5 resulted from the processing of the actual measure data only and Table 2-6 resulted from the processing of the data, including the extended records. The first two columns of the tables list the years and their associated annual yields ordered from smallest to

TABLE 2-5

SAMPLE OUTPUT FOR FREQUENCY ANALYSIS EMPLOYING ACTUAL MEASURED DATA ONLY

```
4  05AF010   MANYBERRIES CK      YIELD DATA LOG-NORMAL DISTN

     YR        DATA      M   PRBOC       TR   YNORPROB

    1961        456.0   13   0.071     1.08   -1.465
    1968        786.0   12   0.143     1.17   -1.068
    1962       1294.2   11   0.214     1.27   -0.792
    1963       3153.7   10   0.286     1.40   -0.566
    1964       3743.4    9   0.357     1.56   -0.366
    1959       6850.1    8   0.429     1.75   -0.180
    1966       8167.9    7   0.500     2.00    0.0
    1969       8330.0    6   0.571     2.33    0.180
    1960       8926.8    5   0.643     2.80    0.366
    1957       9055.9    4   0.714     3.50    0.566
    1958      10628.8    3   0.786     4.67    0.792
    1967      12036.1    2   0.857     7.00    1.068
    1965      17532.8    1   0.929    14.00    1.465

              N =    13   TOTM =    13


 LEAST SQUARES REGRESSION EQUATION TO FIT LOG-NORMAL PLOT IS

       X =     3.6666+    0.5368 Y

       R = 0.9340     SEY =     0.1837     SEB =    0.0619

       T =    8.667     F =  75.121

       VY =     0.2423     VX =     0.7333


EVENT TR2 =     4640.5     EVENT TR10 =    22633.9     RATIO TR2/TR10 =    4.878
```

TABLE 2-6

SAMPLE OUTPUT FOR FREQUENCY ANALYSIS EMPLOYING ESTIMATED DATA

```
4  05AF010   MANYBERRIES CK      YIELD DATA LOG-NORMAL DISTN

     YR        DATA      M   PRBOC       TR   YNORPROB

    1961        456.0   27   0.036     1.04   -1.803
    1968        786.0   26   0.071     1.08   -1.465
    1962       1294.2   24   0.143     1.17   -1.068
    1963       3153.7   22   0.214     1.27   -0.792
    1964       3743.4   19   0.321     1.47   -0.464
    1959       6850.1   14   0.500     2.00    0.0
    1966       8167.9   12   0.571     2.33    0.180
    1969       8330.0   11   0.607     2.55    0.272
    1960       8926.8    9   0.679     3.11    0.464
    1957       9055.9    8   0.714     3.50    0.566
    1958      10628.8    5   0.821     5.60    0.921
    1967      12036.1    4   0.857     7.00    1.068
    1965      17532.8    1   0.964    28.00    1.803

              N =    13   TOTM =    27


 LEAST SQUARES REGRESSION EQUATION TO FIT LOG-NORMAL PLOT IS

       X =     3.6775+    0.4482 Y

       R = 0.9567     SEY =     0.1496     SEB =     0.0411

       T =   10.903     F = 118.880

       VY =     0.2423     VX =     1.1041


EVENT TR2 =     4759.2     EVENT TR10 =    17867.5     RATIO TR2/TR10 =    3.754
```

largest. The third column lists the order numbers assigned to the events with the largest being 1. Order numbers in Table 2-6 have been adjusted on the basis of the extended records and are different from those in Table 2-5. The last three columns of the tables list the calculated probabilities of occurrence, the return periods, and the reduced variables "Y", respectively, left to right. Under the table, the value "N" refers to the number of actual measured data points on which the regression is based; the value "TOTM" refers to the total number of years analysed including extended records. Note that the "TOTM" value in Table 2-6 is larger than N as a result of the inclusion of extended records. The values of "N" and "TOTM," the calculated regression equation and its statistics are presented, and the magnitudes of the mean annual and 10-year events are listed. In addition to the printed output, the programme recorded the regression equation and the calculated event magnitudes on punch cards.

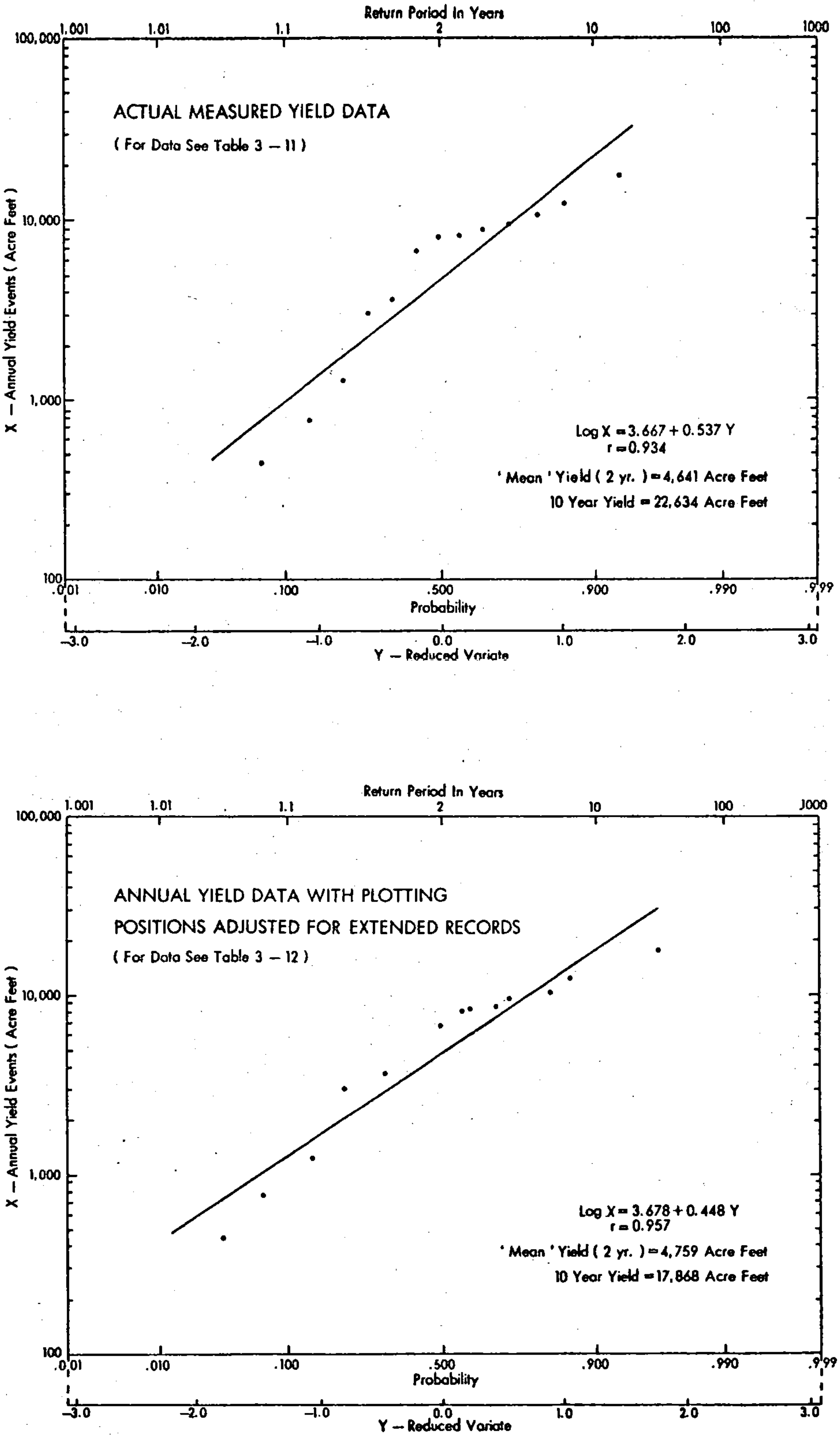
Figure 2-1 illustrates frequency plots of the annual yield data series for study basin number 4, Manyberries Creek. The upper graph is an illustration of the frequency analysis of the actual measured data only, while the lower graph is an illustration of the analysis with plotting positions adjusted on the basis of extended records. Magnitudes of the events are plotted on the logarithmic ordinate scales; and the reduced variate Y is plotted on the abscissa scales. The abscissa scales are also graduated in terms of return period and probability. The linear regression for each of the plots has been estimated by the method of least squares on the transformed data. The adjustment of the probabilities of occurrence on the basis of extended records has resulted in a lowering of the slope of the regression relationship and in an improved correlation coefficient. The use of extended data for basin number 4 resulted in a lower estimate of the 10-year annual yield and a slightly higher estimate of the mean annual yield.

In the present study, the relatively short actual measured data series for many of the basins did not provide a strong basis for frequency analysis. However, record extension by correlation of regression with nearby long-term stations, resulted in more complete data sets for most stations. For the purposes of the present research, these extended data series have been employed in the estimation of the dependent variables for each of the study basins. In all cases, extended data have been employed in adjusting the plotting positions of the actual measured events; however, only the actual measured events have been employed in the frequency analysis by least squares regression.

One of the difficulties inherent in the use of computerized calculations in the frequency analysis of hydrologic data series is that no visual presentation of the data is obtained. Where no visual plots are obtained, it is difficult to identify anomalies in the data. In the present study, as a check for anomalies in the records, and in order to evaluate the curve fitting procedure, the Calcomp Plotter of the University of Alberta Computing Centre was utilized to plot frequency graphs for each of the data sets analysed. Each of the resulting graphs was examined with the aim of identifying anomalous values which obviously did not belong in the same sample as the other observations in the series. In the very

Figure 2 - 1

EXAMPLE FREQUENCY PLOTS OF ANNUAL YIELD DATA
FOR MANYBERRIES CREEK - STATION NO.4



few cases where such points existed they were deleted from the data sets and the frequency analysis was repeated.

The estimated magnitudes of the dependent variables resulting from the frequency analyses for each of the study basins were subsequently employed in the analyses of the hypothetical model (equation 1-1) as discussed in Chapter 4.

CHAPTER III

THE INDEPENDENT VARIABLES: MEASURES OF CLIMATIC AND OTHER PHYSICAL GEOGRAPHIC PATTERNS

3.1 INTRODUCTION

In this Chapter, the selection and subsequent measurement of the independent variables is discussed. The independent variables selected for inclusion in the present study were measures of climatic and other physical geographic patterns for the study basins. Two criteria were considered in the selection of independent variables, the first relating to their theoretical relationships to the four selected dependent variables, and the second related to available data sources. The theoretical relationships were considered in an attempt to avoid the establishment of spurious statistical relationships. The available data sources were considered important in that the relatively short time span and large study area involved in this project precluded the collection of data in the field. This latter consideration led the researcher to rely on available climatic data and existing topographic map coverage as basic data sources.

Based on a consideration of both the theoretical relationships of the variables to streamflow and the available data sources, the independent variables can best be discussed in two groups. The first group, the climatic measures, is closely related to the local moisture balance patterns; and therefore includes factors which control the water supply available for streamflow. Climatic conditions are variable both spatially and temporally and measures must be estimated on the basis of available long-term climatic records. The second group of independent variables, the measures of other physical geographic patterns, included measures of drainage area, basin topography, channel pattern, surficial deposits and vegetation. These variables control the efficiency with which the available moisture supply is collected in channels and conveyed from the basin as streamflow. In this manner, these variables are closely related to the timing and to a lesser extent to the amount of streamflow. While the variables included in this second grouping are not totally time invariant, they are considerably more so than are the climatic patterns considered above. For the purpose of the present study, it was assumed that the measures representative of this second group of independent variables were time invariant with reference to the relatively short periods of streamflow data being analysed. Under this assumption the data for this group of other physical geographic measures have been collected by measurement from available topographic maps.

3.2 CLIMATIC PATTERNS

3.2.1 Theoretical Relationships of Climatic Patterns to Streamflow

Within the hydrologic cycle there are three important processes which are closely related to climatic patterns. These processes are precipitation, snowmelt and evapotranspiration. The precipitation process accounts for the major moisture input to the surface-division of the hydrologic cycle. As such, precipitation is the primary cause of all streamflow. The occurence of precipitation involves both spatial and temporal variations in form, amount and intensity. Consideration of this variability provided a basis for the selection of a set of precipitation variables for inclusion in the present analyses.

The snowmelt process involves the ripening of snow and subsequent release of stored moisture from the snowpack. The processes by which the snowpack gains heat, increases in density, and finally releases water are extremely complex (U.S. Army Corps of Engineers, 1956, p. 141). From a theoretical point of view, the best approach to this problem is through an analysis of the heat balance of the snowpack. Unfortunately, suitable data with respect to the radiation balance and the condition of the snowpack are not normally available; and the energy balance approach to the estimation of snowmelt is not feasible. As an alternative, it is possible to employ the available air temperature and accumulated snowfall data to estimate various empirical indices of melt.

The third hydrologic process which is dependent on climatic conditions is evapotranspiration. The actual amount of moisture lost by evapotranspiration varies as a function of several factors including the available moisture supply, air temperature, humidity, wind conditions, vegetation type, solar radiation and season. The multivariate nature of the relationships involved has resulted in the development of several methods for the empirical estimation of potential and actual evapotranspiration based on available data. These empirical methods range from the formulae developed by Penman (1953, p. 40) which require detailed data on radiation, wind and humidity, to the formulae developed by Turc (1953) and Thornthwaite (1948 and 1957) which are based on temperature and precipitation data only.

3.2.2 Available Climatic Data

The climatic data requirements in the present study were for estimates of the normal climatic patterns of variables related to the precipitation, snowmelt and evapotranspiration processes. The basic climatic data employed in the compilation of climatic variables for each study basin were the published normals of temperature and precipitation (Canada, Department of Transport, Meteorological Branch, 1968a and b). In particular, monthly normal precipitation, monthly normal snowfall, monthly normal mean daily temperatures and monthly normal maximum daily temperatures were employed. These data were compiled for 174 climatological stations within the study area. An examination of the locations of the study basins led to the identification of a need for climatic data from stations outside the study area particularly along the United States and foothill boundaries. For this reason, data were included for an additional 22 climatological stations, nine located in the foothills and mountains of Alberta and 13 in the United States near the border. Data for the United States stations were compiled from publications of the United States, Department of Commerce, Weather Bureau (1962a, b and c). Each of the 196 climatological stations is located on the map Figure 3-1 and identified in Table 3-1.

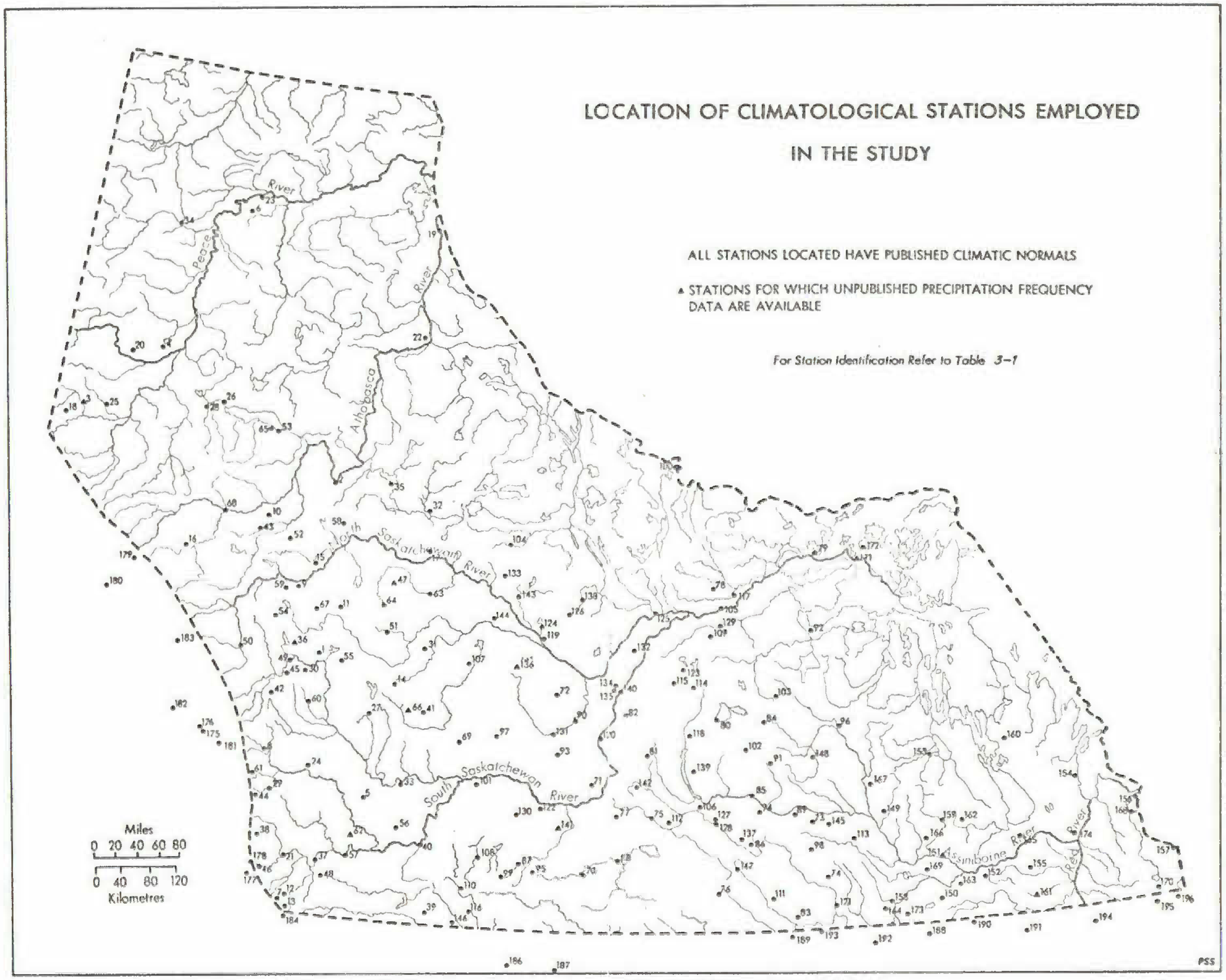

LOCATION OF CLIMATOLOGICAL STATIONS EMPLOYED
IN THE STUDY
ALL STATIONS LOCATED HAVE PUBLISHED CLIMATIC NORMALS
▲ STATIONS FOR WHICH UNPUBLISHED PRECIPITATION FREQUENCY DATA ARE AVAILABLE
For Station Identification Refer to Table 3-1
Peace River
Athabasca River
North Saskatchewan River
South Saskatchewan River
Saskatchewan River
Assiniboine River
Red River
Miles
0 20 40 60 80
0 40 80 120
Kilometres
PSS

Figure 3 - 1

TABLE 3-1

CLIMATOLOGICAL STATION IDENTIFICATION *

STATION NUMBER	STATION NAME	STATION NUMBER	STATION NAME	STATION NUMBER	STATION NAME	STATION NUMBER	STATION NAME
1	ALIX	50	ROCKY MTN HOUSE	99	KLINTONEL	148	YORKTON A
2	ATHABASCA	51	SEDGEWICK	100	LAC LA RONGE	149	BIRTLE
3	BEAVERLODGE CDA	52	SION	101	LEADER	150	BOISSEVAIN 2
4	BERWYN	53	SLAVE LAKE	102	LEROSS	151	BRANDON CDA
5	BROCKS	54	SPRINGDALE	103	LINTLAW	152	CYPRESS RIVER
6	BUFFALO HEAD PRAIRIE	55	STETTLER	104	LOON LAKE CDA	153	DAUPHIN A
7	CALDWELL	56	SUFFIELD A	105	LOST RIVER	154	GIMLI A
8	CALGARY A	57	TABER	106	LUMSDEN	155	GRAYSVILLE
9	CALMAR	58	THORHILD	107	MACKLIN	156	GREAT FALLS
10	CAMPSIE	59	THORSBY	108	MAPLE CREEK	157	INDIAN BAY
11	CAMROSE	60	THREE HILLS	109	MELFORT	158	MELITA
12	CARDSTON	61	TURNER VALLEY	110	MERRYFLAT	159	MINNEDOSA
13	CARWAY	62	VAUXHALL	111	MIDALE	160	MOOSEHORN
14	CORONATION	63	VERMILION A	112	MOOSE JAW A	161	MORDEN CDA
15	EDMONTON INDUSTRIAL	64	VIKING	113	MOOSOMIN	162	NEPAWA A
16	EDSON	65	WAGNER	114	MUENSTER	163	NINETTE
17	ELK POINT	66	WASTINA HEMARUKA	115	MUSKIKE SPRINGS	164	PIERSON
18	ELMNORTH	67	WETASKIWIN	116	NASHLYN	165	PORTAGE LA PRAIRIE A
19	EMBARRAS A	68	WHITECOURT	117	NIPAWIN	166	RIVERS A
20	FAIRVIEW	69	ALSASK HARDENE	118	NOKOMIS	167	RUSSELL
21	FORT MACLEOD	70	ANEROID	119	NORTH BATTLEFORD A	168	SEVEN SISTERS FALLS
22	FORT MCMURRAY A	71	BEECHY	120	OUTLOOK	169	SOURIS
23	FORT VERMILION	72	BIGGAR	121	OXBOW	170	SPRAGUE
24	GLEICHEN	73	BROADVIEW A	122	PENNANT	171	THE PAS
25	GRAND PRAIRIE A	74	CARLYLE	123	PILGER	172	THE PAS A
26	GROUARD	75	CARON	124	PRINCE	173	WASKADA
27	HANNA	76	CEYLON	125	PRINCE ALBERT A	174	WINNIPEG A
28	HIGH PRAIRIE	77	CHAPLIN	126	RABBIT LAKE	175	ANTHRACITE
29	HIGH RIVER	78	CHOICELAND	127	REGINA A	176	BANFF
30	HILLSDOWN	79	CUMBERLAND HOUSE	128	REGINA CDA	177	BEAVER MINES
31	HUGHENDEN	80	DAFOE A	129	RIDGEDALE	178	COWLEY A
32	IRON RIVER	81	DAVIDSON	130	ROADENE	179	ENTRANCE
33	JENNER	82	DUNDURN	131	ROSETOWN	180	JASPER
34	KEG RIVER	83	ESTEVAN A	132	ROSTHERN	181	KANANASKIS
35	LAC LA BICHE	84	FOAM LAKE	133	ST WALBURG	182	LAKE LOUISE
36	LACOMBE	85	FORT QU'APPELLE	134	SASKATOON A	183	NORDEGG
37	LETHBRIDGE A	86	FRANCIS	135	SASKATOON U OF S	184	BABB
38	LUNDBRECK	87	GARDEN HEAD	136	SCOTT CDA	185	GLASGOW
39	MANNYBERRIES	88	GRAVELBOURG	137	SEDLEY	186	HARLEM
40	MEDICINE HAT	89	GRENFELL	138	SPIRITWOOD	187	MALTA
41	NACO	90	HARRIS	139	STRASBOURG	188	BOTTINEAU
42	OLDS	91	HUBBARD	140	SUTHERLAND	189	CROSBY
43	PEAVINE	92	HUDSON BAY	141	SWIFT CURRENT A	190	HANSBORO
44	PEKISKO	93	HUGHTON	142	TUGASKE	191	LANGDON
45	PENHOLD A	94	INDIAN HEAD	143	TURTLEFORD	192	MOHALL
46	PINCHER CREEK	95	INSTOW	144	WASECA	193	PORTAL
47	RANFURLY	96	KAMSACK	145	WHITEWOOD	194	HALLOCK
48	RAYMOND	97	KINDERSLEY	146	WILLOW CREEK	195	ROSEAU
49	RED DEER	98	KIPLING	147	YELLOW GRASS	196	WARROAD

* STATION NOS 1 TO 174 ARE LOCATED WITHIN THE STUDY AREA
STATION NOS 175 TO 183 ARE LOCATED OUTSIDE THE STUDY AREA IN THE FOOTHILLS AND MOUNTAINS OF ALBERTA
STATION NOS 184 TO 196 ARE LOCATED OUTSIDE THE STUDY AREA IN THE UNITED STATES
FOR EXACT STATION LOCATIONS SEE FIGURE 4-1

3.2.3 Interpolation of Climatic Data for the Study Basins

The climatic normal data selected for use in the present research were the result of measurements at point observation sites. The streamflow characteristics being examined were estimated on a drainage basin basis. Before analysing the relationships between streamflow characteristics and various climatic measures, it was necessary to interpolate estimates of the climatic normals for each of the study basins. In the present study the Thiessen Polygon method has been employed in the estimation of basin climatic normals (Thiessen, 1911). This method involves the calculation of the weighted arithmetic average of the available point data. The weight for each climatological station is derived such that the station is considered to be representative of a proportion of the total area, dependent on the spacing of the observation points.

The Thiessen Polygon and Drainage Basin delimitations and the required area measurements were made on 1:250,000 scale topographic maps of the study basins. The drainage area delimited was that of gross drainage area as defined by the topographic divide between adjacent basins. Wherever map coverage was available, the topographic divide was first located on 1:50,000 scale maps and then transferred to the 1:250,000 scale maps. The Thiessen Polygon procedures were employed in the estimation of monthly normals of mean daily temperatures, maximum daily temperatures, monthly precipitation, and monthly snowfall, for each of the 161 study basins. The resulting climatic normal data sets for each of the study basins formed the basis for the subsequent estimation of climatic variables (see Section 3.2.4 following).

3.2.4 The Selection and Calculation of Climatic Variables

For the purpose of discussion the selection and measurement of climatic variables is considered under three headings, precipitation based variables, temperature based variables and composite variables.

3.2.4.1 Precipitation Based Variables

A summary of the selected precipitation based climatic variables and their methods of calculation is contained in Table 3-2. The selected variables are intended to provide indices of the variations in precipitation patterns with respect to amount, form, and seasonal pattern. A brief statement of the theoretical reasoning underlying the choice of the variables follows.

The Mean Annual Precipitation, MAP, was selected as an index of the total water supply available annually within a basin. This variable is logically related to the annual yield of streamflow from an area. It might also be expected that this variable may represent a general index of total water supply as related to the potential magnitude of annual flood events.

TABLE 3-2 - SELECTED PRECIPITATION BASED CLIMATIC VARIABLES
(All Units are Inches)

Variable Abbreviation	Variable Name	Calculation
MAP	Mean Annual Precipitation	Sum of the Basin Mean Monthly Precipitation totals for 12 months
MAS	Mean Annual Snowfall	Sum of the Basin Mean Monthly Snowfall totals for 12 months
MASP	Percentage of MAP as Snowfall	MAS water equivalent expressed as a % of MAP
MWP	Mean Winter Precipitation	Sum of the Basin Mean Monthly Precipitation totals for the 5 months, November to March inclusive
MSP	Mean Spring Precipitation	Sum of the Basin Mean Monthly Precipitation totals for the 3 months, April to June inclusive
MWSP	Mean Winter and Spring Precipitation	Sum of MWP and MSP
A10YP	Annual 10 year Precipitation	Established ratio of 10 year to the mean 300 day from November 1 precipitation for nearest index station. Multiplied the above ratio by the sum of the 10 Monthly Precipitation figures November to August and added September to October normals
W10YP	Winter 10 year Precipitation	Established ratio of 10 year to the mean 160 day precipitation from November. Multiply ratio by MWP

The Mean Annual Snowfall, MAS, was selected on the basis of the fact that most streamflow on the plains is the result of snowmelt. Ideally, this measure requires the estimation of the water equivalent of the snowpack at the end of winter. Such data are not readily available and in the present analysis it has been necessary to assume a constant conversion factor of 0.1 in estimating the water equivalent of the annual snowfall. These data were expressed as two variables, the Mean Annual Snowfall in inches, and the MAS water equivalent as a percentage of the mean annual precipitation. This second snowfall measure has been abbreviated MASP, for Mean Annual Snowfall as a Percentage of MAP.

Another measure of seasonal precipitation is that of the Mean Winter Precipitation, MWP. This variable compiled for the five month period from November to March was based on the consideration that the total winter precipitation is stored in a frozen state to await the spring melt period.

The Mean Spring Precipitation, MSP, was estimated for the three months April to June. This variable was selected as an index of the amount of moisture which is available in spring to supplement the snowmelt runoff. This additional moisture is an important factor during the spring period of high streamflow.

The preceding two variables were combined to estimate another variable, the Mean Winter-Spring Precipitation, MWSP. This variable was proposed as an index of the total moisture input to a basin during the season of high flows.

In addition to the above measures of precipitation amount and seasonal distribution, two variables intended to provide measures of the year-to-year variability of precipitation patterns were included. The first of these, the Annual 10 Year Precipitation, A10YP, was included to provide an index of the variability of the available moisture supply. Such a variable might be related both to the above average annual yields and the above average flood flows. The estimation of this variable was based on a limited number of precipitation frequency data which were available for only ten climatological stations within the study area. The index station closest to the study basin has been used in the precipitation frequency calculations for that basin.

The second measure of precipitation variability was the Winter 10 Year Precipitation, W10YP. The estimation of this variable involved similar calculations to those employed in the estimation of the A10YP above.

3.2.4.2 Temperature Based Variables

A summary of the selected temperature based climatic variables and their methods of calculation is contained in Table 3-3. Temperature patterns and their seasonal variations are primary controls over the processes of snowmelt and evapotranspiration. It is in this context that temperature based variables have been included in the present

TABLE 3-3 - SELECTED TEMPERATURE BASED CLIMATIC VARIABLES
(All Units are Degrees Fahrenheit)

Variable Abbreviation	Variable Name	Calculation
MATR	Mean Annual Temperature Range	Basin Mean Monthly Temperature for warmest month minus the Basin Mean Monthly Temperature for the coldest month
MWT	Mean Winter Temperature	Mean of the Basin Mean Monthly Temperatures for the 5 months November to March inclusive
MJANT	Mean January Temperature Below 32° F.	Basin Mean January Temperature subtracted from 32° F.
MST	Mean Spring Temperature	Mean of the Basin Mean Monthly Temperatures for the 3 months April to June inclusive
MJUNT	Mean June Temperature	Basin Mean June Temperature
WMXT	Mean Winter Maximum Temperature	Mean of the Basin Mean Monthly Maximum temperatures for the 5 months November to March inclusive
JAMXT	Mean January Maximum Temperature Below 32° F.	Basin Mean Monthly Maximum Temperature for January subtracted from 32° F.

study. A brief statement of the theoretical considerations underlying the choice of the temperature variables follows.

The first of the temperature based variables was the Mean Annual Temperature Range, MATR. This measure was intended as a general index of continentality and therefore of basin location. The higher the value of the annual temperature range the greater the seasonal contrast in temperature. This seasonal contrast in temperature represents a factor related to the winter snowfall accumulation and the summer rate of evapotranspiration.

The next two variables, the Mean Winter Temperature, MWT, and the Mean January Temperature, MJANT, were proposed as measures of the intensity of winter and therefore of the permanency of the snowpack. The colder the winter temperatures, the more permanent the snowpack is likely to be, and the greater the potential for spring runoff. The second of these measures, the Mean January Temperature, was included to overcome the limitations of the Mean Winter Temperature, the value of which is affected by the uneven length of the winter season over the study area. In order to avoid the inclusion of zeros and negative numbers in the data, the MJANT has been subtracted from 32.0°F.

The variables, Mean Spring Temperature, MST, and Mean June Temperature, MJUNT, were included to provide indices of the rapidity of the spring warming trend. Spring temperatures were expected to relate both to the rate of spring runoff from snowmelt and to the increase in evapotranspiration. The Mean June Temperature measure was proposed to overcome the season length variations which might affect the Mean Spring Temperature Measure.

The final two temperature based measures were the Winter Maximum Temperature, WMXT, and the January Maximum Temperature, JAMXT. These variables provided further measures of the intensity of winter conditions. The maximum temperatures are particularly important in controlling the snowmelt process. A higher winter maximum temperature indicates a greater potential for melt to occur during winter. In the southwestern portion of the study area, a higher winter maximum might be associated with the occurrence of "chinook" conditions, which may result in the rapid sublimation or melt of a snowpack. The Mean January Maximum Temperature was intended to overcome the limitations imposed by variable season lengths. This variable was calculated by subtracting the Mean January Maximum Temperature from 32.0°F. in order to avoid zeros and negative numbers in the data.

3.2.4.3 Composite Variables Based on Both Temperature and Precipitation Data

A third group of climatic variables for the study basins was derived by combining the available temperature and precipitation data. These composite variables were based on water balance calculations and included estimates of potential evapotranspiration, actual evapotranspiration, and annual water surplus. The estimation of these variables in the present research has been accomplished by the application of empirical formulae developed by Thornthwaite (1948) and Turc (1953). These two procedures are by no means the only available methods for estimation of water balance patterns; however, they have the advantage of entailing relatively simple calculations based on only temperature and precipitation data. The selected composite variables are summarized in Table 3-4.

The Potential Evapotranspiration by Thornthwaite procedures, PE, was introduced as an index of general climatic conditions particularly summer temperature and sunshine patterns. It also represents the potential water loss from the basin by evapotranspiration where water supply is not limited.

Two estimates of Actual Evapotranspiration have been included in the present study. The first such measure, THAET, was calculated according to Thornthwaite (1948) procedures. This method employs the mean monthly temperature and precipitation data and the station latitude, and entails the calculation of a monthly water budget including estimates of precipitation, storage and potential evapotranspiration. A more recent modification of these procedures (Thornthwaite, 1957) has not been employed in the present study. This modified method involves provisions for water use at less than potential rates; however for prairie conditions it has been demonstrated by Holmes and Robertson (1959) that current precipitation from summer storms is used at or near potential rates. Also, the 1948 procedures have been successfully employed for the study area by Laycock (1967). The Thornthwaite calculations in the present analysis were made employing a modification of a computer program written by Black (1966). A second estimate of actual evapotranspiration, TUAET, was made employing procedures developed by Turc (1953). These procedures are simpler than the Thornthwaite method and use only mean annual temperature and mean annual precipitation data. These two estimates of actual evapotranspiration were introduced in the present study as measures of the water loss by the evapotranspiration process.

The final pair of climatic variables were estimates of annual moisture surplus based on the actual evapotranspiration estimates discussed above and the mean annual precipitation. The first of these measures, the Mean Annual Water Surplus by Thornthwaite, THSUR, was estimated by subtracting the actual evapotranspiration estimate by Thornthwaite procedures from the Mean Annual Precipitation. The second water surplus variable, the Mean Annual Water Surplus by Turc, TUSUR, was calculated in the identical manner employing the actual evapotranspiration estimated by Turc's procedures and the MAP. Both of these annual surplus measures were proposed as indices of the total available water supply.

3.3 OTHER PHYSICAL GEOGRAPHIC PATTERNS

The second major group of independent variables included in the present research has been classified under the general heading "Other Physical Geographic Patterns". This grouping includes measures of the physical characteristics of the study basins with the exception of the climatic variables.

TABLE 3-4 - SELECTED COMPOSITE CLIMATIC VARIABLES BASED ON BOTH PRECIPITATION AND TEMPERATURE DATA
(All final units are Inches)

Variable Abbreviation	Variable Name	Calculation
PE	Potential Evapotranspiration	Potential Evapotranspiration calculated by Thornthwaite (1948) procedures based on Basin Monthly Temperature Normals
THAET	Thornthwaite Actual Evapotranspiration	Actual Evapotranspiration calculated by Thornthwaite (1948) procedures based on Basin Monthly Temperature and Precipitation Normals
TUAET	Turc Actual Evapotranspiration	Actual Evapotranspiration according to Turc's formula (1953): $TUAET = (P/\sqrt{0.9 + (P/L)^2})/25.4$ $L = 300 + 25t + 0.05t^3$ where: P is mean annual ppt. in millimeters t is mean annual temp. in °C.
THSUR	Thornthwaite Surplus	THAET subtracted from MAP
TUSUR	Turc Surplus	TUAET subtracted from MAP

3.3.1 Theoretical Relationships to Streamflow Characteristics

The physical geographic characteristics of a watershed are related to the efficiency with which the available moisture supply is collected in channels and conveyed from the basin as streamflow. In particular, the physical characteristics are related to the timing of streamflow, and have a less well-defined influence on the total volume of streamflow which is more directly the result of the climatic parameters. For the purpose of discussion of the theoretical relationships of the physical characteristics to streamflow, three groups of patterns have been considered: measures of drainage area; measures of topographic patterns; and measures of surficial geology, soil, vegetation and land-use.

Drainage area is a measure which is of the utmost importance in any study of streamflow characteristics. In any study of the relationships of physical geographic patterns to streamflow, it is imperative that the effect of drainage basin size be accounted for before considering other factors. One of the most straightforward approaches to this problem is to divide the streamflow volume by the drainage area, thus converting the volume measure to an estimate of depth over the basin. Unfortunately in many areas, particularly in a semi-arid glaciated plains region such as the present study area, it is not possible to accurately delimit the drainage area contributing to streamflow. The problem of defining drainage area as a hydrologic variable in the glaciated Canadian Prairies has been discussed in some detail by Stichling and Blackwell (1958) and by Laycock (1959). Drainage in this region is characterized by large areas of internal drainage. This drainage is directed into local depressions which contain swamps or sloughs. This pattern is particularly pronounced in areas of hummocky dead ice morainic deposits. Many of the local depressions have no outlet except by evaporation, while others overflow and contribute to streamflow in some years. The net result of these patterns is that the drainage area contributing to streamflow is extremely variable both seasonally and from year to year as a function of the moisture supply conditions. There is no simple method of evaluating the area of a drainage basin which contributes to streamflow under given conditions. The best solution probably is the detailed field study of local patterns under varying conditions of moisture supply. Such an approach is impossible for a project such as the present study which involves a large number of basins over a large study area. Durrant and Blackwell (1959, p. 107) in their study of flood flows in the Southern Canadian Prairies prepared detailed drainage maps at a relatively large scale from aerial photographs. These maps were employed as a basis for estimating the drainage area contributing to the mean annual floods. A similar approach has been taken in the present study, and a measure of non-contributing drainage area has been included.

The second group of physical patterns considered, topographic characteristics of drainage basins, is related to the efficiency with which the available moisture is collected in channels and conveyed as streamflow from the basin (Langbein and others, 1947, p. 128). In this context, topographic patterns are closely related to the concentration and timing of streamflow, and indirectly related to the total volume of streamflow. Where topography is such that the movement of available moisture toward channels is retarded, the result may be an increase in infiltration and in actual evapotranspiration. Many hydrologic textbooks contain a general discussion of the influence of topographic patterns on both the timing and volume of runoff (see for example, Ward, 1967, p. 324 and 330ff.; and Wilson, 1969, p. 84). Laycock (1959) has provided a qualitative discussion of the effects of local topographic patterns on available water supplies in the Canadian Prairies.

While the importance of landform and topographic patterns in the determination of the characteristics of streamflow is widely recognized, any analysis of these relationships requires the quantitative measurement of these patterns. The pioneering work in this field was published by Horton in 1945 and updated by Strahler (1964). Many of the measures are not suitable for use in a study on the scale of the present research in which basin sizes range from 50 to 10,000 square miles. It is imperative that straightforward, readily measured indices of topographic conditions be employed. Many such measures have been devised and are in evidence in the work of Benson (1962b and 1964); Karuks (1964); Thomas and Benson (1971) and many others.

The third group of physical geographic patterns considered includes measures of surficial geology, soils, vegetation and landuse. These patterns share a common influence on streamflow in that they are related to variations in the infiltration capacity of the land surface. Unfortunately quantitative data on these patterns are difficult to obtain particularly for a study on the scale of the present research. In the present analysis such measures have been limited to a single vegetation variable.

3.3.2 Available Data Sources

The types of variables which are likely to be most closely linked to the streamflow characteristics being analysed in the present study include measures of basin drainage areas, topography, surficial geology and soils, and prevailing vegetation and landuse patterns. While none of these conditions is completely time invariant, the assumption of time invariance has been made for the time span of the present research. This time span has a maximum limit of 30 years corresponding to the maximum period of hydrometric data compilation. This assumption was required so that estimates of physical characteristics could be made from available data sources and in particular from topographic maps which have been compiled at different times for different sections of the study area.

In the present research, it was considered desirable to employ a data source which provided consistent data over the entire study area. Such a data source exists in the form of the National Topographic Series of Maps. The available topographic map coverage for the study area at scales of 1:50,000 and 1:250,000 is illustrated in Figure 3-2. The entire study area has been mapped at a scale of 1:250 000. For most of the southern part of the study area, topographic map coverage is also available at a scale of 1:50,000. Some of the study basins extend south

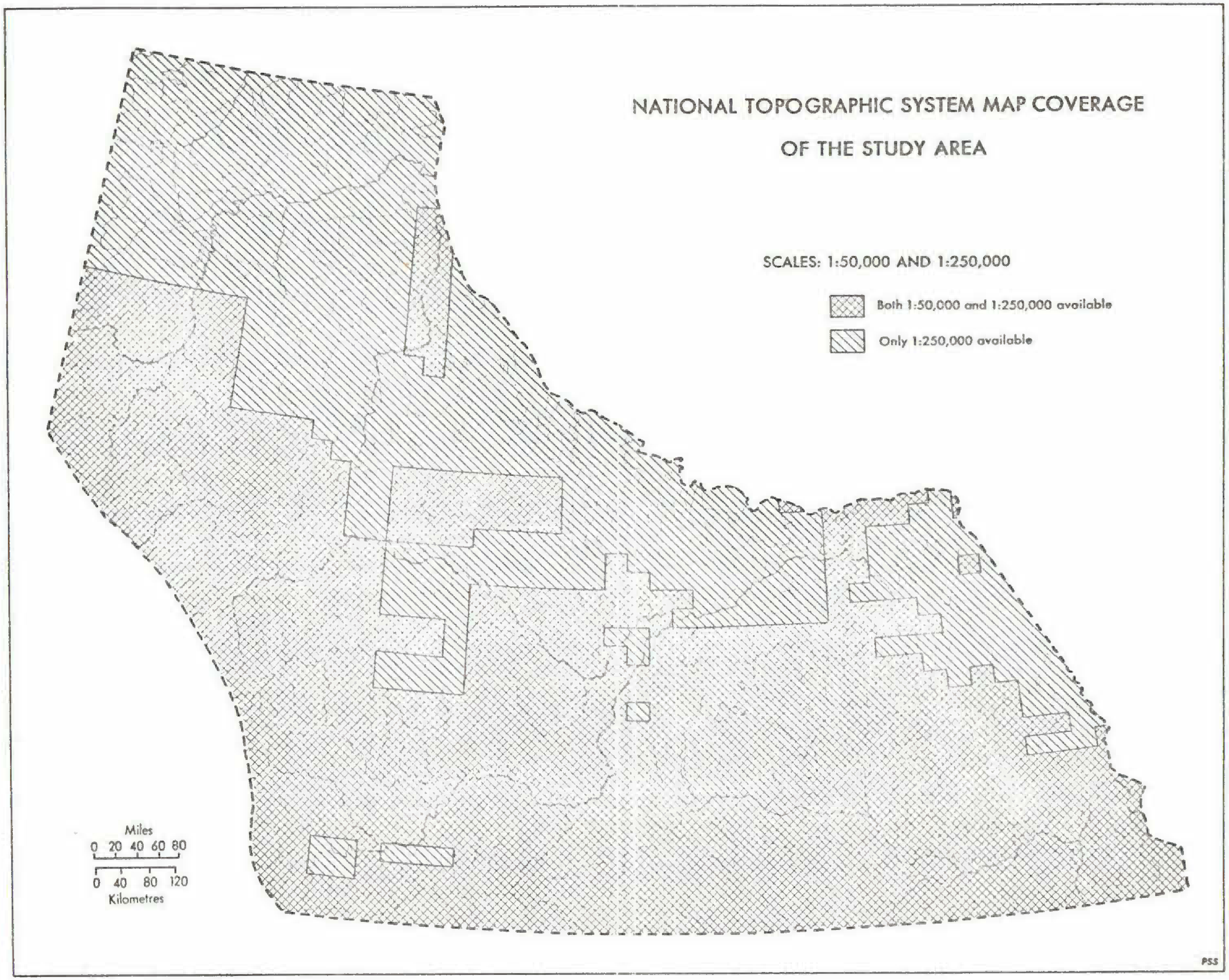

NATIONAL TOPOGRAPHIC SYSTEM MAP COVERAGE
OF THE STUDY AREA
SCALES: 1:50,000 AND 1:250,000
▩ Both 1:50,000 and 1:250,000 available
▧ Only 1:250,000 available
Miles
0 20 40 60 80
0 40 80 120
Kilometres
PSS

Figure 3 - 2

of the United States-Canada border, and topographic map coverage for these areas is available at a scale of 1:250,000. Since 1:250,000 scale topographic maps were available for the entire study area and bordering regions, this scale of maps was chosen as the basic data source to be employed in the compilation of the physical variables. One of the major disadvantages of employing these maps was related to the evaluation of measures which were dependent on the interpolation of elevations from the contour lines. As a partial solution to this problem, it was decided to employ 1:50,000 scale topographic maps, wherever available, to increase the accuracy with which elevations could be estimated.

3.3.3 The Selection and Compilation of Other Physical Variables

The selection of the other physical geographic variables to be included in the analyses of the present project was made on the basis of a consideration of their theoretical relationships to the streamflow characteristics under study, a review of the relevant literature concerning other similar studies, and a recognition of some of the practical problems of data compilation on the scale of the present investigation. The consideration of the theoretical relationships of physical geographic patterns to the streamflow characteristics being studied led to the identification of three main groups of variables for inclusion in the analyses. These groups of variables were measures of drainage basin area with particular reference to the contributing portion of the basin; measures of topographic factors including slopes, elevation, basin shape, channel network and natural storage; and measures of surficial geology, soils, vegetation and landuse patterns. A review was made of the relevant literature dealing with similar studies in other areas and the problems associated with the quantitative measurement of physical geographic patterns.* This literature review introduced the author to the great number of different measurements of physical geographic patterns which have been employed with varying degrees of success in other similar studies. The experience of other researchers, the theoretical relationships to streamflow characteristics, and some consideration of the practical limitations of data collection from the available maps and on the scale of the present study, provided a basis for the selection of a set of physical geographic variables.

---

*The literature review of other similar studies examined a large number of publications including: Benson (1959, 1962a and b, and 1964); Cole (1966); Collier and Nix (1967); Coulson and Gross (1967); Durrant and Blackwell (1959); Golding and Low (1960); Horton (1945); Howe, Slaymaker and Harding (1967); Karuks (1964); Kennard et al. (1963); Lull and Anderson (1967); Lull and Sopper (1966); Morisawa (1959a and b); Nash and Shaw (1966); Schneider (1965); Slaymaker and Jeffrey (1969); Solomon et al. (1968); Strahler (1964); Thomas and Benson (1971); and others.

3.3.3.1 The Compilation of the Basic Physical Data from Topographic Maps

A set of basic physical data was compiled for each of the 161 study basins. These data were obtained by measurement from the 1:250,000 scale topographic maps of the study basins. The measures of physical patterns which were included in the basic data set for each basin, are listed in Table 3-5. The table summarizes the abbreviations attached to the measures, the measure names, and the methods of measurement.

The 13 physical geographic measures which are summarized in Table 3-5 form the basic physical data set as compiled from the 1:250,000 scale maps for each of the 161 study basins.

3.3.3.2 The Final Calculation of the Physical Variables

The physical geographic variables for inclusion in the analysis of streamflow characteristics were calculated from the basic physical data sets, which were compiled for each of the study basins from the 1:250,000 scale topographic maps. A summary of the variable abbreviations, the variable names and their methods of calculation is contained in Table 3-6. A brief consideration of the reasons for the selection of these measures is presented in the following pages.

The first group of physical variables relates to the measurement of drainage basin area. The Topographic Drainage Area, TDA, was the simplest such measure and was compiled directly from the topographic maps. The Non-Contributing Drainage Area as a Percentage of TDA, NCDA, represented an attempt to introduce a measure of the relative area within a basin which is subject to internal drainage and does not normally contribute directly to streamflow. A variation on this variable is the Contributing Drainage Area, CDA, which was the difference between the TDA and the NCDA measures. This CDA measure was proposed as a substitute for TDA particularly when mean streamflow events are being analysed.

The second group of physical variables included a large number of measures of basin topography. The selected variables included measures of basin shape, channel slope, basin elevation and relief, channel development, and natural storage. The first of the topographic variables was the Basin Length, BL, which was defined as the main channel length extended to the divide. This variable represented a very general measure of drainage area, a measure which was not affected by non-contributing sections of the basin. The BL variable was combined with the TDA variable to estimate an index of Basin Shape, BS. There are several possible indices of basin shape referred to in the literature; however, the measure selected for the present study was chosen on the basis of its simple calculation and use of available data. The shape of a basin is theoretically related to the timing of streamflow.

There are several possible methods for the evaluation of slope within a basin. In the present study the estimation of an average basin slope was ruled out because of the time-consuming measurements which

TABLE 3-5 - BASIC PHYSICAL DATA COMPILED FROM TOPOGRAPHIC MAPS OF THE STUDY BASINS

Abbreviation for Measure	Name of Measure	Method of Measurement
TDA	Topographic Drainage Area***	Topographic drainage area was outlined and measured in the calculation of Thiessen weights for climatic data interpolation (see Section 4.2.3.1). Whenever 1:50,000 scale maps were available, the divide was located on the larger scale maps and transferred to 1:250,000 scale
NCDA	Non-contributing Drainage Area***	Non-contributing drainage area was delimited on 1:250,000 scale maps to include all areas of internal drainage not contributing surface runoff to the stream under normal conditions. Such areas include swamp, lake and slough drainage not connected to the main stream
MCL	Main Channel Length*	Measured length of longest channel from hydrometric station site to end of tributary
MCLD	Mean Channel Length to the Divide*	Same as MCL above but extended upstream to the topographic basin divide
TL	Total Tributary Length*	Measured total length of all tributaries (including intermittent streams) with connections to main channel
ELG	Elevation of Gauge** (Hydrometric Station)	Elevation of gauge interpolated between nearest contour lines

Table 3-5, Continued

Abbreviation for Measure	Name of Measure	Method of Measurement
ELD	Elevation of Main Channel intersection with Divide**	Elevation of main channel intersection with divide interpolated between nearest contour lines
ELMX	Maximum Elevation within the Drainage Basin**	Elevation of highest point in the basin
EL10	Elevation of point on Main Channel located 10% of MCLD above gauge**	Elevation of point 10% of length up Main Channel interpolated between nearest contour lines
EL85	Elevation of point on Main Channel located 85% of MCLD above gauge**	Elevation of point 85% of length up Main Channel interpolated between nearest contour lines
ALOS	Surface Area of Lakes on the Stream***	Surface area of lakes located on streams and therefore a part of the stream network
ASOS	Surface Area of Swamps Adjacent to Streams***	Surface area of swamps located on or immediately adjacent to the streams and therefore part of the stream network
AF	Area of Forest Cover within the Topographic Drainage Area***	Total area of forest cover in the basin

* All length measurements were made in inches with an opsometer. These measurements were subsequently converted to miles.

** All elevation measurements were made in feet by interpolation between contour lines. 1:50,000 scale maps were employed wherever such coverage was available.

*** All area measurements were made in square inches with "Bruning" Areagraph Charts. These measurements were subsequently converted to square miles.

TABLE 3-6 - SUMMARY OF SELECTED INDEPENDENT VARIABLE MEASURES OF OTHER PHYSICAL GEOGRAPHIC PATTERNS

Variable Abbreviation	Variable Name	Calculation of Variable (For explanation of variable abbreviation see Table 4-7 and column 1 of this table)
TDA	Topographic Drainage Area	Topographic Drainage Area expressed in square miles
NCDA	Non-contributing Drainage Area as % of TDA	NCDA = ((NCDA/TDA) x 100.0) + 1.0
CDA	Contributing Drainage Area	CDA = TDA - NCDA
BL	Basin Length	Equivalent to the Main Channel Length to the Divide, MCLD
BS	Basin Shape	$BS = BL^2/TDA$
MCS	Main Channel Slope	Estimated as slope of the channel between points located 10% and 85% of total BL measured upstream from the hydrometric station MCS = (EL85-EL10)/(BL x 0.75)
BEL	Basin Elevation	Mean of the elevations at 10% and 85% of the total BL measured from the hydrometric station BEL = (EL85-EL10)/2.0
BR	Basin Relief	Total Basin Relief, the difference between the highest elevation and the elevation of the hydrometric station BR = ELMX - ELG
DDTDA	Drainage Density based on TDA	DDTDA = (MCL + TL)/TDA
DDCDA	Drainage Density based on CDA	DDCDA = (MCL + TL)/CDA

Table 3-6 , Continued

Variable Abbreviation	Variable Name	Calculation of Variable (For explanation of variable abbreviation see Table 4-7 and column 1 of this table)
OFTDA	Average Length of Overland Flow based on TDA	OFTDA = 1.0/(2.0 x DDTDA)
OFCDA	Average Length of Overland Flow based on CDA	OFCDA = 1.0/(2.0 x DDCDA)
ALTDA	Surface Area of Lakes on Stream as % of TDA	ALTDA = ((ALOS/TDA) x 100.0) + 1.0
ALCDA	Surface Area of Lakes on Stream as % of CDA	ALCDA = ((ALOS/CDA) x 100.0) + 1.0
STDA	Surface Area of Swamp on Stream as % TDA	STDA = ((ASOS/TDA) x 100.0) + 1.0
SCDA	Surface Area of Swamp on Stream as % CDA	SCDA = ((ASOS/CDA) x 100.0) + 1.0
SLTDA	Surface Area of Lakes and Swamp on Stream as % TDA	SLTDA = ((ALOS + ASOS)/TDA) x 100.0) + 1.0
SLCDA	Surface Area of Lakes and Swamp on Stream as % CDA	SLCDA = (((ALOS + ASOS)/TDA) x 100.0) + 1.0
FTDA	Area of Forest as % TDA	FTDA = (AF/TDA) x 100.0) + 1.0

would have been required. As an index of slope within a basin, a measure of the Main Channel Slope, MCS, as developed by Benson (1959) was adopted. This measure was expected to be closely related to the timing of runoff and thus to the magnitudes of flood flows.

Two simple measures of basin elevation and relief were included in the analyses. Basin Elevation, BEL, was estimated as the mean of the elevations at 10 percent and 85 percent of the channel length (the same points as employed in the estimate of channel slope) and provided an approximation to the mean elevation of the basin. This relatively crude index was proposed as an alternative to actually measuring the area-elevation relationship for the entire basin. The proposed basin elevation index, BEL, was expected to relate in a general way to the precipitation patterns within the study area. Another measure, Basin Relief, BR, was proposed to account for local orographic effects on the precipitation pattern. This measure was estimated as the difference in elevation between the hydrometric station and the highest point in the basin.

The degree to which the channel network is developed within a basin is related to the efficiency with which the available moisture is collected in streams. In the present analyses, two measures of network development were investigated, the drainage density and the length of overland flow. The Drainage Density which is the mean number of the streams occurring per square mile was calculated both for the topographic drainage area and for the contributing drainage area. The Average Length of Overland Flow was estimated to be equal to one half of the average distance between streams. This measure was also estimated both for the TDA and for the CDA.

The final variables relating to basin topography were measures of the natural storage on or adjacent to the streams. These measures included the Percentage Area of Lakes, the Percentage Area of Swamps, and the Percentage Area of Lakes and Swamps. In each case, these measures were expressed as percentages of both the topographic and the contributing drainage areas. A constant value of 1.0 was added to all percentages to avoid the inclusion of zeros in the data set. The measures of natural storage within a basin were considered to be particularly relevant in view of their dampening effect on peak discharges.

In addition to the physical variables relating to drainage area and topographic patterns, it was considered desirable to include some measures of surficial geology, soils, vegetation, and landuse patterns. Unfortunately, data on these patterns are not readily available on the scale of the present study. A further problem relates to the quantitative measurement of these conditions. For these reasons no attempt was made to include measures of surficial geology and soils in the present analyses. Landuse and vegetation data are also difficult to obtain on the scale of the present study, and only one such variable was included, the Percentage Forest Area in a Basin. This variable was estimated for the topographic drainage area. The forest area variable was proposed on the basis of the effect forest cover has on reducing the seasonal peaks in streamflow by increasing the infiltration capacity of a watershed.

In all, data for 19 physical variables were compiled for each of the 161 study basins. These variables were employed as independent variables in the analyses of the regional streamflow characteristics.

3.4 SUMMARY

In this chapter consideration has been given to the selection and measurement of the independent variables. The variables were selected on the basis of their theoretical relationships to the streamflow characteristics under study and on the basis of available data sources. For the purposes of discussion, the independent variables have been considered in two groups, climatic measures and measures of other physical geographic variables. In all, 39 independent variables, 20 climatic measures, and 19 measures of other physical geographic patterns, have been estimated for each of the 161 study basins. These data are employed in the analyses of Chapter 4.

CHAPTER IV

THE ANALYSES OF THE RELATIONSHIPS OF THE SELECTED STREAMFLOW CHARACTERISTICS TO THE INDEPENDENT VARIABLES

4.1 INTRODUCTION

In this chapter, the statistical analyses of the relationships of the selected streamflow characteristics to the measures of climatic and other physical geographic patterns are described. The analyses were undertaken from the approach of hydrologic system investigation utilizing the multivariate statistical techniques of multiple regression and factor analysis. The data set utilized included 4 dependent and 39 independent variables compiled for each of the 161 study basins. A complete list of all the variables and their abbreviations is presented in Table 4-1. These abbreviations are used interchangeably with the variable names in the discussion of the analyses which follows.

4.2 METHODOLOGY OF ANALYSIS

Statistical methods are widely used in hydrologic research in an attempt to synthesize the information contained in a mass of data, and to make the best possible use of past records in the understanding and prediction of future events (Slivitzky 1966, p. 184). The complex nature of the interrelationships linking streamflow characteristics and the various measures of climatic and other physical geographic patterns are well suited to analysis by multivariate statistical techniques. The basic technique employed in the present analysis has been multiple linear regression.

In the present study, two approaches to the analysis by multiple linear regression techniques were employed. In the first approach, stepwise multiple linear regression analysis was utilized to relate each of the dependent variables to the full set of 39 independent variables. In the second approach, factor analytic methods were employed in the screening of the independent variables prior to the application of stepwise multiple linear regression. This screening procedure was necessary in order to eliminate redundant independent variables, and resulted in simpler, more easily interpreted regression models.

4.2.1 Stepwise Multiple Linear Regression

Stepwise multiple linear regression is a special case of multiple linear regression analysis. The basic regression model on which the technique is based entails the calculation of a prediction equation relating a dependent random variable to one or more independent deterministic variables (Stammers 1966, p. 256). In addition to the regression equation, the analysis results in two measures of the strength of the relationship, the coefficient of multiple correlation and the standard error of the estimate. The theoretical considerations and mathematical calculations which underlie regression analyses are beyond the scope of the present discussion; however, the reader is referred to the work of

TABLE 4-1 - ALPHABETICAL LISTING OF VARIABLE ABBREVIATIONS AND NAMES AS EMPLOYED IN THE ANALYSES

Variable Abbreviation*	Variable Name
ALCDA	Area of Lakes % CDA
ALTDA	Area of Lakes % TDA
A10YP	Annual Precipitation 10 Year Return Period
BEL	Basin Elevation
BL	Basin Length
BR	Basin Relief
BS	Basin Shape
CDA	Contributing Drainage Area
DDCDA	Drainage Density Based on CDA
DDTDA	Drainage Density Based on TDA
FR	Flood Index-Ratio 10 Year to the Mean
FTDA	Area of Forest as % of TDA
JAMXT	Mean January Maximum Temperature below 32°F
MAP	Mean Annual Precipitation
MAS	Mean Annual Snowfall
MASP	Mean Annual Snowfall as % of MAP
MATR	Mean Annual Temperature Range
MCS	Main Channel Slope
✓MF	Mean Annual Flood
MJANT	Mean January Temperature below 32°F
MJUNT	Mean June Temperature
MSP	Mean Spring Precipitation
MST	Mean Spring Temperature
MWP	Mean Winter Precipitation
MWSP	Mean Winter Spring Precipitation
MWT	Mean Winter Temperature
✓MY	Mean Annual Yield
✓M10YF	Annual Flood with 10 Year Return Period
✓M10YY	Annual Yield with 10 Year Return Period
NCDA	Non-contributing Drainage Area % of TDA
OFCDA	Length of Overland Flow Based on CDA
OFTDA	Length of Overland Flow Based on TDA
PE	Potential Evapotranspiration
SCDA	Area of Swamps as % of CDA
STDA	Area of Swamps as % of TDA
SLCDA	Area of Swamps and Lakes as % CDA
SLTDA	Area of Swamps and Lakes as % TDA
TDA	Topographic Drainage Area
THAET	Actual Evapotranspiration by Thornthwaite
THSUR	Annual Water Surplus by Thornthwaite
TUAET	Actual Evapotranspiration by Turc
TUSUR	Annual Water Surplus by Turc
WMXT	Mean Winter Maximum Temperature
W10YP	Winter Precipitation with 10 Year Return Period
YR	Yield Index-Ratio 10 Year to the Mean

✓ dependent vbls.

*The prefix L added to any of these variable abbreviations indicates a transformation to a logarithm with base 10 has been made.

Ezekiel and Fox (1959) and Solomon (1966) for a complete discussion of these aspects of the technique.

The interpretation of a regression analysis for any data set is based on several assumptions regarding the nature of the input data. The most important of these assumptions relate to the sampling procedures and frequency distributions of the variables examined. In most regression analyses the aim is to develop relationships which are applicable beyond the confines of the data sample analysed. In order to ensure the applicability of the results to a particular population of events, it is necessary to ensure that the data analysed represent a random sample of observations drawn from the population of interest. When the observations fail to meet this requirement of random sampling, it is not reasonable to expect the results of the analysis to have general application to the overall population. The second assumption relating to the nature of the basic data requires that the frequency distributions of the variables approximate the normal distribution. The significance tests which are utilized in testing the reliability of a regression analysis are based on this assumption. When the assumption of normality is not met, the use of the standard significance tests is in jeopardy.

Two further problems which are closely related to the selection of input data for a regression analysis are those of spurious correlations and multicollinearity among the independent variables. Both of these problems are closely related to the interpretation of the regression results. A spurious correlation is said to exist when a statistical correlation is found between two variables which are theoretically unrelated. It must be recognized that while the existence of such a relationship is indicative of a statistical covariation between the variables, it in no way indicates a physical or causal relationship. Benson (1965) has discussed the problem of spurious correlation in hydrologic research with particular reference to the use of composite variables which are products, sums or ratios based on several measures which may be common to more than one variable. In order to avoid the misinterpretation of spurious correlations, it is imperative that the input variables are selected on the basis of theoretical or intuitive relationships; otherwise, it is not possible to interpret the physical significance of any relationships which may result.

The existence of multicollinearity among the independent variables in a regression analysis may lead to serious difficulties in the interpretation of the regression results. Multicollinearity is the presence of linear interrelationships among variables in the independent variable set. When such intercorrelations exist, certain elements of information have been measured by more than one variable (Tennessee Valley Authority 1966, p. 134). This problem of multicollinearity is common in hydrologic analysis because of the multivariate nature of the relationships involved. The intercorrelations among independent variables tend to result in unstable regression coefficients which make the interpretation of the functional relationships between the variables most difficult.

In any research project utilizing statistical methods, it is imperative that the researcher carefully evaluates the merits of his data with reference to the assumptions and limitations of the techniques employed. In the case of regression analysis, the implications of the assumptions of random sampling and normal data, and the limitations with respect to spurious correlations and multicollinearity are of particular relevance. The random sampling assumption must be accounted for in the original research design. The problem of normally distributed variables may be attacked by the use of transformations. Many hydrologic variables have right skewed distributions that are limited to positive values. If the data for such variables are transformed to logarithms, the resulting frequency distributions will more closely approximate normality. The spurious correlation problem can best be approached through a theoretical formulation of the hypothetical model and a careful choice of input variables. The final problem, that of multicollinearity, requires the careful selection and screening of independent variables. In the present study the technique of factor analysis has been employed for the purpose of identifying the independent elements of information contained in the independent variable set.

In the present research, the multiple regression analysis has been executed, utilizing the Stepwise Multiple Linear Regression Computer Program, BMD02R, of the Biomedical Computer Program of the University of California (for program documentation refer to Dixon 1970, pp. 233-257). This program computes a sequence of least squares multiple linear regression equations in a stepwise fashion, one variable being added at each step.

4.2.2 Factor Analysis

The second multivariate statistical technique employed in the present research is factor analysis. Factor analysis is a technique which provides a means of collapsing a set of intercorrelated variables into a smaller number of independent dimensions or factors (King 1969, p. 165). It is beyond the scope of this discussion to consider the theoretical background and calculations involved in factor analysis; however, the reader is referred to the work of King (1969, pp. 165-193); Harmon (1967); and the Tennessee Valley Authority (1966, pp. 151-156) for a discussion of the mathematical methodologies involved.

Factor analytic techniques have been employed for two rather different purposes (Tennessee Valley Authority 1966, p. 151). In the first case, the aim is to discover the underlying factors which operate in determining the measurements of the variables and possibly to test hypotheses related to these underlying factors. The ultimate aim in this application is to group variables and identify the underlying dimensions. The second purpose for which factor analysis has been employed is in the screening of an intercorrelated variable set in an effort to identify the independent components of that set; and ultimately, to reduce the dimensionality of the variable set to a few components which may be represented by individual variables. The second approach is more amenable to the building of predictive models in that the variables selected by the screening process represent an economy of explanation. Examples of this approach applied to hydrologic problems are included in the work of the

Tennessee Valley Authority (1966), Wallis (1965) and Wong (1963). The variable screening approach has been employed in the present research in an attempt to identify the variables which most closely represent independent measures of the basic dimensions of the original set of climatic and physical measures. These variables were subsequently employed in the stepwise regression analyses of the streamflow characteristics.

In the present study the factor analyses have been executed utilizing the Factor Analysis Computer Program, BMDX72, of the Biomedical Computer Programs of the University of California (for program documentation see Dixon 1970a, pp. 90-103).

These analyses have utilized a principal component solution and a varimax factor rotation. This method results in a stable factor structure, which is characterized by a strong correspondence between factor dimensions and variables, leading to relatively easy factor identification (Wallis 1965, p. 453; and Tennessee Valley Authority 1966, pp. 155-156).

4.2.3 The Structure of the Analysis

The analyses of the present study have been structured in two stages. The first stage entailed the analysis of the relationships of streamflow characteristics and climatic and other physical geographic patterns at the scale of the entire study area. In this first stage, two approaches to the analysis have been undertaken, firstly a stepwise multiple regression analysis based on all of the independent variables; and secondly, a stepwise multiple regression analysis based on the independent variables as selected by the screening technique of factor analysis.

On the basis of the results of the first stage of the analysis, a second stage in which the study area was divided into hydrologic regions was undertaken. In these analyses, the stepwise multiple regression technique in combination with factor analytic variable screening was employed for each hydrologic region. The second stage of the analysis was undertaken in an effort to improve the relationships developed in the first stage and to evaluate the regional differences, if any, in the relationships.

4.3 DATA PREPARATION

Prior to the commencement of the analyses of the data compiled for the 161 study basins, consideration was given to the relationships of these data to the assumptions underlying the techniques to be employed. In particular, the random sampling and normality assumptions were considered.

4.3.1 Sampling Limitations in the Present Study

In order to ensure the applicability of the results of a statistical analysis, it is necessary to assume that the basic data set represents a random sample drawn from the population in question. The selection of study basins in the present research has been discussed in Chapter I. The population of basins considered in this study includes all basins located within the study area, having a drainage area of between 50 and

10,000 square miles, and having natural streamflow conditions. Unfortunately, it was not possible to draw a random sample from all the basins which met the above three criteria. Rather, it was necessary to impose the criterion of available hydrometric records. The basins which were selected for analysis were all those which met the above criteria with respect to location, drainage area, and flow conditions, and for which a minimum of 5 years of streamflow records were available. Prairie basins for which streamflow data have been collected are not randomly located throughout the study area; but rather, their distribution is closely related to settlement patterns with a much denser coverage in the southern areas. This means that the distribution of study basins available in the present research is also concentrated in the southern regions of the study area, and their selection for analysis cannot be assumed to represent a random sample drawn from the study area. Unfortunately, there is no alternative method of selecting a random sample in that streamflow data are non-existent for many basins. Therefore, the interpretation of the results of the ensuing analyses and their application in the prediction of streamflow characteristics for ungauged basins must remain in doubt. The results must be interpreted carefully with regard to the distribution of the study basins as illustrated in Figure 1-1.

4.3.2. Non-normality and Transformations of the Variables

The significance tests utilized in evaluating the results of a multiple regression analysis are based on the assumption that the input data are normally distributed. Many of the variables employed in the present analysis have a limit and their distributions are skewed. Previous research of a similar nature has led to the conclusion that hydrological data can be made to approximate normality by means of logarithmic transformations (Benson 1962b and 1964 and Karuks 1964).

In the present study ,histograms were plotted of the frequency distributions of each of the independent variables. The climatic variables with the exception of the Annual Surplus calculated by Thornthwaite procedures, THSUR, were found to approximate a normal distribution. The histograms of the measures of the other physical geographic patterns and the THSUR variable exhibited a right skewness and in most cases were limited to the left. These variables were transformed by taking their logarithms to the base 10. A similar transformation was applied to each of the dependent variables. These transformed variables replaced their corresponding original measures in all of the subsequent analyses.

4.3.3. Selection of Comparative Test Sample

The limitations of the basic data set with respect to random sampling and non-normality may introduce some bias into the significance tests of the results of the regression analysis. As a check on the predictive strength of the relationships resulting from the analyses, it was decided to retain a test sample from the original 161 study basins for use in evaluating the relationships. A random sample of 15 basins was selected from the original list of 161 study basins. The test sample was selected on the basis of random numbers with the proviso that no two spatially adjacent basins would be included in the sample. The 15 basins included in this test sample are listed in Table 4-2. The data compiled for these 15 basins were not employed in the subsequent analyses, but rather were retained as an independent sample for the testing of the relationships developed in the analyses.

TABLE 4-2 - BASINS IN TEST SAMPLE FOR
COMPARISON OF ANALYSIS RESULTS
(Not Employed in the Analyses)

Basins Number	Basin Name	Basins Number	Basin Name
4	Manyberries Ck	78	Stony Ck
7	Elbow R	82	Minnedosa R
15	Kneehills Ck	108	Mowbray Ck
25	Eagle Ck	118	Cooks Ck
30	Skull Ck	133	Pembina R
37	Wascana Ck	139	East Prairie R
44	Cutarm Ck	159	Rock Ck
52	Red Deer R		

4.4 ANALYSIS STAGE 1: COMPLETE STUDY AREA

In this initial stage of the analysis, the relationships of the selected streamflow characteristics to the various measures of climatic and other physical geographic patterns were examined on the scale of the complete study area. Two approaches to the building of statistical models were utilized, first, a stepwise multiple regression analysis including all of the independent variables, and secondly, a stepwise multiple regression analysis including only those independent variables selected after screening by factor analytic techniques. In the following sections, each of these approaches is discussed and a comparative summary of the resulting relationships is presented.

4.4.1 Stepwise Multiple Regression Analyses: All Variables

The first approach to the analysis of the data for the entire study area utilized the stepwise multiple linear regression technique to estimate a multiple regression equation for each of the four dependent variables. In each case the full set of 39 independent variables for the 146 study basins was entered in the analysis. The regression relationships were built up in a stepwise fashion, with the variable added at each step being that one which has the highest partial correlation with the dependent variable after the effects of variables already in the model have been accounted for.

Table 4-3 is the simple correlation matrix for all of the variables. The regression equations which resulted from the stepwise regression analysis for each of the dependent variables are summarized in Tables 4-4, 4-5, 4-6 and 4-7. Table 4-4 is discussed in detail below as an example of the format employed throughout this chapter in displaying the results of the stepwise regression analyses.

4.4.1.1 Format of Regression Results

All of the stepwise regression results in the present research are reported in a tabular form similar to Table 5-4. Each table is comprised of 7 columns containing the following information:

Column (1)	The step number in the regression analysis
Column (2)	The regression equation resulting at the given analysis step
Column (3)	The multiple correlation coefficient, R, corresponding to the regression equation at the given analysis step
Column (4)	The coefficient of multiple determination, R^2, corresponding to the regression equation at the given analysis step
Column (5)	The value of the analysis of variance F statistic for the regression equation at the given analysis step
Column (6)	The standard error of the estimate associated with the given analysis step
Column (7)	The standard error of the estimate expressed as a percentage of the mean of the dependent variable.

In Table 4-4 the dependent variable is the LMY. The first row of the table contains entries in Columns 2, 6 and 7 only. The entry in column 2 reports the mean value of the dependent variable, while the entries in columns 6 and 7 report the standard deviation and the standard deviation expressed as a percentage of the mean, respectively. These values are included for comparison with the standard errors resulting for the regression steps. The second and subsequent rows of the table contain entries in all columns and include a complete set of statistics for each step in the regression analysis.

In the tables of results, each step of the regression analysis has been reported until the addition of a variable fails to add at least 1% to the R^2 value. Further steps in the analysis contributed little to the explained variance and often involved regression coefficients which were not significantly different from 0. For reference purposes the regression equation from the final step in each of the regression analyses has been assigned an equation number.

In the interpretation of the results of the regression analyses, it is imperative that consideration be given to the underlying theoretical relationships. In these discussions, references to physical significance

TABLE 4-3

ALL VARIABLE CORRELATION MATRIX FOR THE FULL STUDY AREA

(N=145)

VARIABLE	1	2	3	4	5	6	7	8	9	10	11
1 LHY	1.000	0.918	0.922	0.671	0.669	-0.310	0.729	-0.090	0.005	0.481	0.634
2 LHIDYY		1.000	0.849	0.761	0.751	-0.267	0.811	-0.146	0.059	0.472	0.723
3 LHF			1.000	0.818	0.666	-0.265	0.716	-0.157	0.007	0.434	0.612
4 LHIDYF				1.000	0.690	-0.127	0.719	-0.200	0.153	0.415	0.637
5 LTDA					1.000	0.172	0.500	-0.473	0.033	0.295	0.930
6 LNCDA						1.000	0.019	-0.413	-0.156	-0.368	0.071
7 LCDA							1.000	-0.413	0.048	0.343	0.526
8 LHCS								1.000	0.526	0.558	-0.359
9 LBEL									1.000	0.632	0.518
10 LSR										1.000	0.386
11 LBL											1.000

VARIABLE	12	13	14	15	16	17	18	19	20	21	22
1 LHY	0.217	0.186	0.120	0.633	0.522	0.614	0.584	0.069	-0.069	-0.093	0.061
2 LHIDYY	0.246	0.226	0.159	0.631	0.619	0.615	0.584	0.072	-0.063	-0.096	0.058
3 LHF	0.160	0.115	0.059	0.521	0.505	0.495	0.466	0.087	-0.023	-0.109	0.019
4 LHIDYF	0.175	0.097	0.056	0.358	0.341	0.342	0.316	0.122	0.066	-0.138	-0.060
5 LTDA	0.270	0.231	0.227	0.463	0.475	0.460	0.470	-0.103	-0.167	0.170	0.183
6 LNCDA	-0.185	0.123	0.247	-0.275	-0.238	-0.237	-0.158	-0.380	-0.085	0.406	0.138
7 LCDA	0.302	0.227	0.192	0.560	0.505	0.492	0.424	-0.104	-0.166	0.026	0.177
8 LHCS	0.043	0.532	-0.005	-0.156	-0.176	-0.122	-0.149	0.609	0.533	-0.611	-0.553
9 LBEL	0.195	0.376	0.067	-0.124	-0.127	-0.059	-0.106	0.595	0.623	-0.604	-0.627
10 LSR	0.392	0.251	0.147	0.172	0.157	0.199	0.171	0.576	0.526	-0.558	-0.502
11 LBL	0.405	0.243	0.234	0.461	0.459	0.459	0.464	-0.039	-0.040	0.025	0.056
12 LES	1.000	0.136	0.122	0.224	0.199	0.203	0.198	0.254	0.257	-0.301	-0.249
13 LALTDA		1.000	0.979	0.227	0.233	0.403	0.420	0.132	0.143	-0.129	-0.149
14 LALCDA			1.000	0.132	0.145	0.363	0.401	0.082	0.157	-0.076	-0.140
15 LSTDA				1.000	0.997	0.972	0.959	-0.068	-0.169	0.069	0.157
16 LSCDA					1.000	0.969	0.962	-0.087	-0.175	0.070	0.164
17 LSLTDA						1.000	0.993	-0.030	-0.119	0.016	0.105
18 LSLCDA							1.000	-0.062	-0.114	0.050	0.105
19 LDDTDA								1.000	0.902	-0.936	-0.912
20 LDDCDA									1.000	-0.895	-0.897
21 LOFTDA										1.000	0.913
22 LOFCDA											1.000

VARIABLE	23	24	25	26	27	28	29	30	31	32	33
1 LHY	0.643	0.349	0.365	0.218	0.364	0.096	0.236	0.291	0.302	-0.104	0.061
2 LHIDYY	0.539	0.211	0.275	0.216	0.256	-0.034	0.107	0.151	0.203	-0.114	0.095
3 LHF	0.511	0.232	0.293	0.187	0.293	0.050	0.179	0.222	0.233	-0.076	0.067
4 LHIDYF	0.219	-0.024	0.109	0.171	0.046	-0.133	-0.060	-0.033	0.093	-0.103	-0.015
5 LTDA	0.125	-0.077	-0.050	0.023	-0.062	-0.256	-0.192	-0.125	-0.097	-0.153	0.160
6 LNCDA	-0.336	-0.181	-0.301	-0.279	-0.291	-0.132	-0.229	-0.191	-0.262	-0.163	0.234
7 LCDA	0.245	-0.046	0.022	0.074	-0.019	-0.235	-0.153	-0.095	-0.265	-0.149	0.135
8 LHCS	0.161	-0.076	0.330	0.443	0.188	0.144	0.155	0.074	0.244	0.329	-0.369
9 LBEL	-0.224	-0.272	0.226	0.563	-0.165	0.013	-0.074	-0.216	-0.147	0.227	-0.725
10 LSR	0.290	-0.013	0.410	0.097	0.200	0.050	0.132	0.011	0.238	0.323	-0.576
11 LBL	0.190	-0.070	-0.107	0.024	-0.021	-0.241	-0.165	-0.106	-0.050	-0.162	0.139
12 LES	0.096	-0.017	0.088	0.152	0.065	-0.079	-0.015	-0.035	0.079	-0.030	0.017
13 LALTDA	0.207	-0.196	-0.276	0.091	-0.123	-0.275	-0.234	-0.257	-0.056	-0.321	0.280
14 LALCDA	0.153	-0.159	-0.105	0.049	-0.152	-0.263	-0.242	-0.233	-0.097	-0.326	0.293
15 LSTDA	0.605	0.251	0.160	-0.003	0.217	-0.069	0.078	0.212	0.130	-0.243	0.235
16 LSCDA	0.586	0.265	0.151	-0.013	0.203	-0.063	0.061	0.198	0.112	-0.247	0.261
17 LSLTDA	0.593	0.219	0.118	0.000	0.153	-0.099	0.017	0.148	0.086	-0.279	0.265
18 LSLCDA	0.568	0.207	0.135	-0.022	0.134	-0.109	-0.001	0.137	0.063	-0.293	0.261
19 LDDTDA	-0.027	-0.291	0.173	0.431	-0.019	-0.051	-0.047	-0.136	0.011	0.334	-0.415
20 LDDCDA	-0.167	-0.277	0.069	0.353	-0.134	-0.298	-0.125	-0.212	-0.098	0.376	-0.406
21 LOFTDA	0.022	0.177	-0.194	-0.439	0.005	0.017	0.027	0.124	-0.022	-0.396	0.429
22 LOFCDA	0.170	0.272	-0.071	-0.353	0.133	0.059	0.123	0.210	0.150	-0.354	0.413
23 LHTDA	1.000	0.617	0.526	0.195	0.627	0.295	0.520	0.569	0.597	-0.307	0.265
24 HSF		1.000	0.325	0.271	0.804	0.832	0.925	0.971	0.735	0.021	-0.005
25 HAS			1.000	0.730	0.533	0.706	0.957	0.760	0.816	0.324	-0.333
26 HASP				1.000	0.454	0.149	0.321	0.152	0.491	0.376	-0.433
27 HHP					1.000	0.563	0.855	0.835	0.969	-0.018	0.012
28 HSP						1.000	0.910	0.927	0.545	0.356	-0.247
29 HHSP							1.000	0.939	0.826	0.216	-0.212
30 AIDYP								1.000	0.779	0.071	-0.059
31 WIDYF									1.000	-0.063	0.052
32 HWT										1.000	-0.591
33 HUANT											1.000

VARIABLE	34	35	36	37	38	39	40	41	42	43
1 LHY	-0.488	-0.392	-0.032	0.034	-0.522	0.342	0.099	0.301	0.421	-0.157
2 LHIDYY	-0.463	-0.425	-0.026	0.026	-0.523	0.203	-0.034	0.206	0.292	-0.154
3 LHF	-0.417	-0.343	-0.022	0.022	-0.458	0.260	0.054	0.259	0.236	-0.145
4 LHIDYF	-0.359	-0.359	0.072	-0.062	-0.409	-0.014	-0.150	0.551	0.056	-0.193
5 LTDA	-0.264	-0.257	-0.116	0.111	-0.361	-0.037	-0.234	-0.146	-0.004	-0.022
6 LNCDA	0.229	0.233	-0.243	0.239	0.207	-0.134	-0.130	-0.291	-0.126	0.258
7 LCDA	-0.302	-0.292	-0.083	0.079	-0.394	-0.013	-0.713	-0.102	0.026	-0.056
8 LHCS	-0.269	-0.346	0.351	-0.373	-0.228	-0.060	0.043	0.123	-0.025	-0.396
9 LBEL	-0.415	-0.730	0.784	-0.850	-0.444	-0.302	-0.097	-0.222	-0.334	-0.855
10 LSR	-0.534	-0.645	0.392	-0.456	-0.554	-0.049	-0.048	0.073	0.002	-0.548
11 LBL	-0.291	-0.293	-0.092	0.029	-0.381	-0.028	-0.223	-0.139	-0.002	-0.050
12 LES	-0.164	-0.207	0.009	-0.006	-0.218	0.306	-0.078	-0.045	0.009	-0.082
13 LALTDA	-0.240	-0.183	-0.269	0.244	-0.325	-0.157	-0.327	-0.150	-0.088	0.126
14 LALCDA	-0.174	-0.144	-0.287	0.262	-0.278	-0.153	-0.372	-0.208	-0.098	0.151
15 LSTDA	-0.386	-0.264	-0.173	0.195	-0.444	0.296	-0.016	0.209	0.377	0.029
16 LSCDA	-0.377	-0.259	-0.159	0.202	-0.437	0.285	-0.024	0.192	0.364	0.017
17 LSLTDA	-0.395	-0.271	-0.254	0.219	-0.460	0.242	-0.097	0.143	0.323	0.032
18 LSLCDA	-0.378	-0.255	-0.226	0.235	-0.445	0.235	-0.094	0.119	0.311	0.251
19 LDDTDA	-0.218	-0.275	0.425	-0.431	-0.255	-0.213	-0.193	-0.089	-0.210	-0.462
20 LDDCDA	-0.149	-0.357	0.397	-0.405	-0.197	-0.283	-0.153	-0.206	-0.303	-0.434
21 LOFTDA	0.229	0.406	-0.442	0.443	0.264	0.198	0.091	0.055	0.197	0.478
22 LOFCDA	0.146	0.352	-0.438	0.417	0.192	0.251	0.146	0.199	0.300	0.438
23 LHTDA	-0.502	-0.274	-0.267	0.260	-0.490	0.555	0.247	0.577	0.723	0.049
24 HSF	-0.273	-0.111	-0.031	0.533	-0.158	0.976	0.843	0.776	0.976	-0.091
25 HAS	-0.410	-0.564	0.291	-0.320	-0.517	0.610	0.607	0.307	0.710	-0.497
26 HASP	-0.613	-0.652	0.416	-0.450	-0.600	-0.065	0.034	0.469	0.029	-0.592
27 HHP	-0.374	-0.276	-0.049	0.044	-0.294	0.499	0.605	0.894	0.815	-0.115
28 HSP	-0.200	-0.153	0.273	-0.306	-0.038	0.306	0.761	0.576	0.738	-0.327
29 HHSP	-0.313	-0.209	0.160	-0.170	-0.171	0.957	0.856	0.916	0.072	-0.263
30 AIDYP	-0.262	-0.222	0.014	-0.016	-0.147	0.930	0.947	0.791	0.535	-0.133
31 WIDYF	-0.374	-0.214	-0.292	0.381	-0.269	0.420	0.533	0.850	0.753	-0.067
32 HWT	-0.073	-0.412	0.980	-0.969	-0.007	-0.005	0.415	-0.023	-0.142	-0.956
33 HUANT	0.154	0.494	-0.963	0.936	0.090	0.027	-0.372	0.018	0.144	0.920
34 HST	1.000	0.897	-0.156	0.192	0.970	-0.224	0.033	-0.363	-0.374	0.495
35 HUGNT		1.000	-0.495	0.522	0.503	-0.056	0.025	-0.179	-0.155	0.778
36 WNIT			1.000	-0.992	-0.205	-0.064	0.316	-0.339	-0.170	-0.916
37 LANIT				1.000	0.129	0.072	-0.303	0.020	0.167	0.925
38 PI					1.000	-0.131	0.173	-0.252	-0.275	0.450
39 TWIET						1.000	0.545	0.621	0.549	-0.049
40 TOFET							1.000	0.547	0.705	-0.290
41 LTHSUR								1.000	0.831	-0.093
42 TCSUR									1.000	-0.002
43 HAIR										1.000

TABLE 4-4

SUMMARY OF STEPWISE MULTIPLE REGRESSION RESULTS FOR THE FULL STUDY AREA
ALL VARIABLE ANALYSIS OF LMY
(N=146)

Step Number (1)	Regression Equations[1] (2)	R (3)	R^2 (4)	F*** (5)	S.E. (6)	S.E.% of $\overline{LMY}$ (7)
	$\overline{LMY}$ = 4.14614				.87969	21.2
1	LMY = 1.269 +1.159 LCDA (.09)	.729	.532	163.6	.6040	14.6
2	LMY = -1.074 +0.967 LCDA +0.571 LFTDA (.07) (.05)	.872	.761	227.6	.4331	10.4
3	LMY = 1.316 +0.998 LCDA +0.425 LFTDA -0.231 LNCDA (.06) (.05) (.07)	.888	.788	176.2	.4091	9.9
4	LMY = -0.106 +1.054 LCDA +0.453 LFTDA -0.237 LNCDA +0.122 TUAET (.06) (.05) (.07) (.04)	.895	.801	141.7	.3932	9.6
5	LMY = -0.538 +1.091 LCDA +0.475 LFTDA -0.243 LNCDA +0.145 TUAET -0.484 LOFCDA (.06) (.05) (.06) (.04) (.14)	.904	.817	125.3	.3827	9.2

(4-1)

[1] A t test has been employed to test the significance of each of the regression coefficients. All regression coefficients are significant at the 1% level except where noted by * or **. The * indicates the coefficient is significant at the 5% level only; and the ** indicates the coefficient is not significant at the 5% level.

*** All F statistics are significant at the 1% level.

TABLE 4-5

SUMMARY OF STEPWISE MULTIPLE REGRESSION RESULTS FOR THE FULL STUDY AREA
ALL VARIABLE ANALYSIS OF LM10YY
(N=146)

Step Number (1)	Regression Equations[1] (2)	R (3)	R^2 (4)	F*** (5)	S.E. (6)	S.E.% of $\overline{LM10YY}$ (7)
	$\overline{LM10YY}$ = 4.73958				.67723	14.3
1	LM10YY = 2.276 +0.993 LCDA (.05)	.811	.658	276.9	.3975	8.4
2	LM10YY = 2.166 +0.834 LCDA +0.323 LFTDA (.05) (.04)	.884	.782	255.9	.3187	6.7
3	LM10YY = 2.357 +0.908 LCDA +0.264 LFTDA -0.230 LNCDA (.05) (.04) (.05)	.900	.810	202.2	.2930	6.3
4	LM10YY = 2.254 +0.925 LCDA +0.234 LFTDA -0.205 LNCDA -0.310 LOFCDA (.05) (.04) (.05) (.10)	.907	.822	162.9	.2897	6.1

(4-2)

[1] A t test has been employed to test the significance of each of the regression coefficients. All regression coefficients are significant at the 1% level except where noted by * or **. The * indicates the coefficient is significant at the 5% level only; and the ** indicates the coefficient is not significant at the 5% level.

*** All F statistics are significant at the 1% level.

TABLE 4-6

SUMMARY OF STEPWISE MULTIPLE REGRESSION RESULTS FOR THE FULL STUDY AREA-
ALL VARIABLE ANALYSIS OF LMF
(N=146)

Step Number (1)	Regression Equations[1] (2)	R (3)	R^2 (4)	F*** (5)	S.E. (6)	S.E.% of $\overline{LMF}$ (7)
	$\overline{LMF}$ = 2.66213				.62776	23.6
1	LMF = 0.643 +0.812 LCDA (.07)	.716	.512	151.1	.4400	16.5
2	LMF = 0.547 +0.713 LCDA +0.295 LFTDA (.06) (.04)	.795	.632	122.7	.3835	14.4
3	LMF = 0.509 +0.729 LCDA +0.235 LFTDA -0.405 LOFTDA (.06) (.04) (.12)	.813	.661	92.4	.3692	13.9
4	LMF = 0.341 +0.736 LCDA +0.194 LFTDA -0.451 LOFTDA +0.742 LTHSUR (.06) (.05) (.11) (.21)	.829	.688	77.6	.3558	13.4
5	LMF = 0.473 +0.552 LCDA +0.187 LFTDA -0.567 LOFTDA +0.741 LTHSUR -0.375 LBS (.05) (.05) (.12) (.21) (.13)	.840	.706	67.2	.3466	13.0
6	LMF = 0.953 +0.891 LCDA +0.164 LFTDA -0.748 LOFTDA +1.047 LTHSUR -0.335 LBS -0.021* NASP (.06) (.05) (.14) (.24) (.13) (.01)	.848	.718	59.1	.3403	12.8

(4-3)

[1] A t test has been employed to test the significance of each of the regression coefficients. All regression coefficients are significant at the 1% level except where noted by * or **. The * indicates the coefficient is significant at the 5% level only; and the ** indicates the coefficient is not significant at the 5% level.

*** All F statistics are significant at the 1% level.

TABLE 4-7

SUMMARY OF STEPWISE MULTIPLE REGRESSION RESULTS FOR THE FULL STUDY AREA
ALL VARIABLE ANALYSIS OF LM10YF
(N=146)

Step Number (1)	Regression Equations[1] (2)	R (3)	R^2 (4)	F*** (5)	S.E. (6)	S.E.% of $\overline{LM10YF}$ (7)
	$\overline{LM10YF}$ = 3.29262				.53375	16.2
1	LM10YF = 1.571 +0.694 LCDA (.06)	.719	.517	154.5	.3721	11.3
2	LM10YF = 1.533 +0.711 LCDA -0.403 LOFTDA (.05) (.11)	.747	.557	90.1	.3575	10.8
3	LM10YF = 1.417 +0.727 LCDA -0.429 LOFTDA +0.417* LTHSUR (.05) (.11) (.17)	.759	.577	64.5	.3509	10.7
4	LM10YF = 2.163 +0.744 LCDA -0.495 LOFTDA +1.291 LTHSUR -0.148 W10YP (.05) (.11) (.31) (.04)	.779	.607	54.4	.3393	10.3

(4-4)

[1] A t test has been employed to test the significance of each of the regression coefficients. All regression coefficients are significant at the 1 % level except where noted by * or **. The * indicates the coefficient is significant at the 5% level only; and the ** indicates the coefficient is not significant at the 5% level.

*** All F statistics are significant at the 1% level.

suggest only that the signs of the regression coefficients correspond to physical expectations, rather than implying functional relationships.

4.4.1.2 Summary of the All Variable Stepwise Regression Analyses

The regression equations which resulted from the all variable stepwise regression analysis for each of the four dependent variables were subject to two limitations, spurious correlations and high standard errors. The spurious relationships limit the usefulness of the regression equations as physical models; and the relatively high standard errors limit their application in predicting streamflow characteristics for ungauged basins.

In order to illustrate the spurious nature of the regression equations, the results of the all variable stepwise regression analysis for the dependent variable LMY as detailed in Table 4-4 are discussed. The first independent variable to enter the regression was LCDA and this measure accounts for 53% of the total variance in the dependent variable. In subsequent steps, an additional four independent variables were added to the regression equation (4-1). These variables contributed an additional 29% to the explained variance. All of the regression coefficients of the final equation (4-1) are significant at the 1% level; however, the signs of some of these coefficients seem to contradict physical theory. The positive coefficient associated with the LCDA term is as expected with larger drainage areas being related to larger mean annual yields. The positive coefficient associated with the LFTDA is not as easily explained. On the basis of physical theory, a negative relationship between percentage of forest area and mean annual yield would be expected. The positive relationship in the present analysis is hypothesized to be a somewhat spurious relationship reflecting the geographic distribution of forest within the study area. Forested area is more prevalent in the north and east of the study area in regions which have a greater moisture supply. Therefore in the present analysis, LFTDA may be entering the regression as a substitute for a moisture supply index and does not represent a moisture retention index as might be expected. The third term in the equation, LNCDA, has a negative coefficient which seems to reflect a physical relationship in that the non-contributing drainage area reflects a reduction in runoff. The positive coefficient associated with the fourth term, TUAET, seems to represent a spurious relationship. The actual evapotranspiration is a loss from available moisture and therefore is expected to have a negative relationship to annual yield. However in the relatively dry study area, the precipitation total is closely related to the evapotranspiration, and TUAET is higher in the areas with higher precipitation. Thus, TUAET may be representing an index of moisture supply in the equation. The final term in the equation is LOFCDA and has a negative coefficient. This relationship corresponds to physical theory in that greater distance of overland flow is associated with lower yields. Spurious relationships similar to those discussed above were also identified in the results of the all variable analyses for the other three dependent variables.

The spurious nature of some of the relationships identified in these analyses can be attributed to the multicollinearity among the independent variables. In an effort to overcome this problem, the technique of factor analysis was proposed for the screening of the independent variables. This second approach to the analysis for the full study region is discussed below in Section 4.4.2.

The second limitation associated with the results of the all variable analyses was the relatively high standard errors of the estimates. This problem has no immediate solution. However, it would seem that these errors must be attributed either to the original sample data or to some independent variables which have been omitted from the present analyses. This difficulty will be considered further in Section 4.5.

4.4.2 Stepwise Multiple Regression Analysis in Combination with Factor Analytic Screening of Variables

The second approach to the full study area analysis of the relationships between the selected streamflow characteristics and the independent variables employed the stepwise multiple regression technique in combination with factor analytic screening of the independent variables. This second approach to the analysis was undertaken with the aim of overcoming the multicollinearity problem and developing more meaningful models of the relationships.

The actual methodology employed in the screening of the independent variables was patterned after that utilized by the Tennessee Valley Authority (1966) and by Wallis (1965). The steps in the analysis are as follows:

1) Make a principal component factor analysis with varimax rotation on the complete set of 39 independent variables as compiled for the full sample of 146 basins

2) Examine the rotated factor matrix which resulted from step 1; and select two defining variables for each factor which contributes a minimum of 1% to the cumulative proportion of the total variance explained by the analysis. Generally, only variables with loadings greater than 0.900 are considered in the selection of defining variables; however, when no such loadings exist the highest available are selected

3) Make a second principal components analysis with varimax rotation on the set of defining variables selected in step 2 above

4) Examine the rotated factor matrix which resulted from step 3; and select one defining variable for each factor which contributes at least 1% to the cumulative proportion of the total variance explained by the analysis

5) Employ the variables selected in step 4, as independent variables in a stepwise multiple regression analysis for each of the four dependent variables.

The results of this form of analysis applied to the data for the full study area are reported below.

4.4.2.1 Screening of the Independent Variables by Factor Analytic Techniques

A factor analysis by principal component solution with a varimax factor rotation was made on the full set of 39 independent variables for the 146 study basins. The initial correlation matrix for this analysis is the same as that of Table 4-3 with the omission of the first four variables, the dependent measures. The rotated factor matrix which resulted from this analysis has been reproduced as Table 4-8. Only those factors which add at least 1% to the cumulative proportion of the total variance have been included. The 11 factors included account for a total of 97% of the total variance in the original independent variable set.

The choice of defining variables for each of the factors has been made on the basis of two considerations. First, the magnitudes of the factor loadings have been considered; and second, the problems of data compilation and the ultimate use of the variables have been considered. As pointed out in the Tennessee Valley Authority report (1966) cited earlier, a factor analysis does not intrinsically select the reduced set of orthogonal variables that can replace the original data set. Rather, factor analysis indicates if redundant variables are present, the number of dimensions, and the grouping of the variables. The final selection of variables must be made on the basis of a full consideration of the limitations of the data and their intended use. For this reason, some degree of subjectivity has been utilized along with the factor loadings in the ultimate selection of defining variables.

In the first stage of the variable selection process the highest loadings on each factor were identified. In Table 4-8 all loadings above 0.900 have been underlined; and in the cases of factors for which no loadings of this strength existed, the highest loadings have been indicated. Having identified these highest loadings, one or two defining variables were selected for each factor. The defining variables selected for each factor have been identified by the arrow head symbol in the table. Wherever more than two high loadings were present on a single factor, the author selected the defining variables on the added consideration of the available data and the expected relationships to the dependent variables. For the first factor, MAP was selected because of its high loading and simplicity of measurement, and A10YP was selected on the basis of its anticipated relationship to the 10 year hydrologic events. For the second factor, MJANT and JAMXT were chosen over MWT and WMXT on the basis of the consideration that the former pair of variables are not affected by season length and therefore have more general application over the study area. In the case of the third factor, LSCDA and LSLCDA were selected over their TDA based counterparts on the basis of the

TABLE 4-8

ROTATED FACTOR MATRIX RESULTING FROM THE FACTOR ANALYSIS OF ALL INDEPENDENT VARIABLES FOR THE FULL STUDY AREA

(N=146)

	FACTOR										
VARIABLE	1	2	3	4	5	6	7	8	9	10	11
1 LTDA	-0.11709	-0.03183	0.28428	-0.13421	-0.14265	0.91782◄	0.06821	-0.10875	0.00053	0.01797	-0.05265
2 LHCDA	-0.16048	-0.14951	-0.22284	-0.18298	0.08466	0.10464	0.19021	-0.89192◄	0.06626	-0.07971	-0.05594
3 LCDA	-0.08541	-0.01760	0.30265	-0.09802	-0.15253	0.92599◄	0.05665	0.04895	-0.02132	0.02152	-0.03926
4 LMCS	0.08215	0.22815	-0.08148	0.47057	-0.19895	-0.38063	0.03787	0.19358	-0.12349	0.03836	0.67251◄
5 LBEL	-0.20974	0.73751	-0.07041	0.37473	-0.38311	0.05234	0.09046	-0.02063	-0.04592	0.07680	0.21393
6 LBR	0.06600	0.31037	0.08843	0.44816	-0.34053	0.33241	0.10655	0.18816	-0.15639	0.13176	0.32565
7 LBL	-0.09255	-0.03348	0.27709	-0.01587	-0.14060	0.86577	0.07015	-0.05854	-0.00688	0.36286	-0.03383
8 LBS	0.01166	-0.01919	0.10950	0.24963	-0.05895	0.27959	0.03579	0.03113	-0.01891	0.91303◄	0.02472
9 LALTDA	-0.16328	-0.17863	0.20068	0.11625	-0.14938	0.08966	0.92171◄	-0.00257	-0.02636	0.01643	0.01533
10 LALCDA	-0.16239	-0.18597	0.17729	0.10658	-0.12017	0.07452	0.92224◄	-0.13916	-0.00443	0.02656	0.01059
11 LSTDA	0.13451	-0.09793	0.94515	-0.04668	-0.13409	0.20674	0.00573	0.08734	0.00076	0.03930	-0.01845
12 LSCDA	0.12004	-0.09321	0.94771◄	-0.06903	-0.13032	0.21067	0.01147	0.05127	-0.00015	0.03983	-0.02363
13 LSLTDA	0.08420	-0.11549	0.93621	-0.01435	-0.13637	0.18970	0.18704	0.07907	0.01202	0.02917	0.00138
14 LSLCDA	0.07329	-0.12475	0.93506◄	-0.02486	-0.12661	0.16505	0.21000	-0.00404	0.01096	0.03572	-0.00202
15 LODTDA	-0.07835	0.25062	-0.01650	0.91577	-0.10937	-0.04921	0.04586	0.20422	-0.06387	0.06392	0.05186
16 LODCDA	-0.15306	0.24538	-0.07209	0.92016◄	-0.09735	-0.07917	0.06882	-0.12002	-0.01112	0.07182	0.01765
17 LOFTDA	0.06576	-0.26641	0.00361	-0.91149	0.10887	0.03453	-0.04453	-0.22668	0.06380	-0.06500	-0.05513
18 LOFCDA	0.14987	-0.25573	0.05612	-0.92441◄	0.08546	0.09297	-0.07707	0.09830	0.01004	-0.06239	-0.03746
19 LFTDA	0.55286	-0.26216	0.44783	0.00692	-0.26637	0.10182	0.10558	0.17274	-0.04824	-0.03424	0.17283
20 MAP	0.97070◄	-0.01886	0.14990	-0.11725	-0.03673	-0.02580	-0.05304	0.01355	0.09211	-0.00694	-0.03235
21 MAS	0.77952	0.25170	0.03105	0.09753	-0.37033	0.00717	0.01202	0.05393	-0.40444	0.02167	0.03691
22 MASP	0.17529	0.33964	-0.06483	0.25875	-0.48372	0.05056	0.07183	0.08859	-0.71719◄	0.03368	0.10034
23 MWP	0.85149	-0.08622	0.06254	0.01394	-0.14574	0.01494	-0.05185	0.10915	-0.35166	0.01565	0.06170
24 MSP	0.86128	0.28130	-0.11616	-0.05410	-0.00702	-0.12934	-0.04140	-0.02227	0.11190	-0.01745	0.00649
25 MWSP	0.96766	0.13340	-0.04165	-0.02702	-0.07775	-0.07398	-0.05198	0.04081	-0.10614	-0.00306	0.03506
26 AIOYP	0.97020◄	0.01053	0.09350	-0.06763	-0.03482	-0.04800	-0.09447	-0.00129	0.01933	0.00652	0.02429
27 WIOYP	0.81083	-0.14548	-0.02973	0.04225	-0.16900	-0.00224	0.00198	0.10901	-0.40446	0.03103	0.10782
28 MWT	0.07668	0.95712	-0.13776	0.16487	0.03942	-0.07388	-0.12746	0.00816	-0.02658	-0.01380	-0.01276
29 MJANT	-0.06470	-0.95715◄	0.14390	-0.18434	0.04412	0.05798	0.10136	-0.02555	0.03576	0.01793	-0.01166
30 MST	-0.24401	-0.12739	-0.23388	-0.10035	0.89359◄	-0.12641	-0.10218	-0.05001	0.08513	-0.01442	-0.04062
31 MJUNT	-0.10987	-0.46561	-0.17036	-0.21110	0.80123	-0.16045	-0.09419	-0.01055	0.08234	-0.04771	-0.03806
32 WMXT	0.01411	0.96491	-0.08093	0.18250	-0.03946	-0.02758	-0.11183	0.06471	-0.02267	-0.00694	0.01148
33 JAMXT	-0.01890	-0.96850◄	0.11110	-0.18574	0.07942	0.01872	0.08629	-0.06633	0.03573	0.01477	-0.03145
34 PE	-0.12330	-0.08384	-0.29146	-0.14834	0.89055◄	-0.19973	-0.13958	-0.04035	0.09686	-0.03241	-0.03169
35 THAET	0.93076	-0.03526	0.17319	-0.12006	0.02003	0.00869	-0.02768	-0.00994	0.22417	0.01216	-0.03251
36 TUAET	0.86034	0.33268	-0.04632	-0.08526	0.26082	-0.08834	-0.12990	-0.02894	0.08713	-0.00639	-0.04787
37 LTHSUR	0.78976	-0.07840	0.08404	-0.06897	-0.16706	-0.10284	-0.10473	0.10768	-0.33412	-0.04207	-0.07611
38 TUSUR	0.93012	-0.16024	0.21689	-0.12012	-0.15717	0.00210	-0.01716	0.02966	0.08625	-0.00684	-0.02314
39 MATR	-0.12593	-0.88941	-0.01159	-0.22505	0.35994	-0.05004	0.03863	-0.02702	0.04351	-0.00325	-0.01500
CUMULATIVE PROPORTION OF TOTAL VARIANCE											
	0.27171	0.50758	0.68937	0.77035	0.82277	0.86637	0.90586	0.92722	0.94554	0.96101	0.97289

◄ DEFINING VARIABLES SELECTED FOR FURTHER ANALYSIS

greater confidence the writer has in the CDA based measures. Similar considerations led to the selection of defining variables for each factor.

The third step in the variable screening process was to make a second factor analysis of the defining variables selected on the basis of Table 4-8. The rotated factor matrix resulting from this analysis has been reproduced as Table 4-9. This second factor analysis was employed to examine the stability of the factor structure before selecting the final set of variables for inclusion in the regression analyses.

The highest factor loadings for each factor have been identified by underlining in Table 4-9. The actual selection of the final set of independent variables was made on the basis of considerations similar to those employed in the selection of defining variables from Table 4-8. For example, the MAP was selected over M10YP in factor 3, on the basis of its ease of measurement; LALCDA was selected over LALTDA in factor 4, on the basis of the author's expressed confidence in the CDA based measure; and LSLCDA was selected over LSCDA in factor 5 on the basis of its possible wider application under more varied lake and swamp conditions in future studies. A list of the final selection of independent variables is presented in Table 4-10.

TABLE 4-10 - FINAL SELECTION OF INDEPENDENT VARIABLES
BASED ON FACTOR ANALYSIS SCREENING

Factor Number	Selected Variable	Factor Number	Selected Variable
1	LCDA	7	ST
2	LDDCDA	8	LNCDA
3	MAP	9	JAMXT
4	LALCDA	10	MASP
5	LSLCDA	11	LMCS
6			

4.4.2.2 Stepwise Multiple Regression Analysis after Variable Screening

The factor analytic screening of the original set of 39 independent variables resulted in the identification of 11 orthogonal factors, and the subsequent selection of a set of 11 independent variables. While these variables are not totally free of multicollinearity since their factor loadings were all somewhat less than 1.0, they do represent a set of variables which approximate the assumption of independence among the independent variables. These eleven variables were employed in a stepwise multiple regression analysis for each of the four dependent variables.

TABLE 4-9

ROTATED FACTOR MATRIX RESULTING FROM THE FACTOR ANALYSIS OF THE SELECTED DEFINING VARIABLES

(N=146)

	FACTOR										
VARIABLE	1	2	3	4	5	6	7	8	9	10	11
1 LTDA	0.92727	0.09777	-0.08019	0.09267	0.22407	-0.10192	0.15001	0.12477	-0.04108	-0.00046	-0.10767
2 LNCDA	0.08205	0.04016	-0.09530	0.13257	-0.15257	0.09600	-0.11944	0.93970◄	-0.12399	0.06901	-0.12078
3 LCDA	0.93346◄	0.10502	-0.06137	0.08376	0.22841	-0.11815	0.16901	-0.02675	-0.02635	-0.02036	-0.08405
4 LMCS	-0.36151	-0.37511	0.02520	0.03464	-0.07949	-0.03750	0.20122	-0.24528	0.19429	-0.15351	0.74633◄
5 LBS	0.18133	-0.17664	0.00440	0.05342	0.08592	-0.95400◄	0.07283	-0.09144	-0.02367	-0.03011	0.01963
6 LALTDA	0.08219	-0.07438	-0.12513	0.95118	0.13862	-0.03195	0.13969	0.00733	-0.13848	-0.02328	0.01746
7 LALCDA	0.07172	-0.09880	-0.11060	0.94791◄	0.13107	-0.03242	0.10358	0.13655	-0.14554	-0.00839	0.00201
8 LSCDA	0.25868	0.08328	0.13531	0.06401	0.90932	-0.06140	0.19218	-0.11805	-0.11510	0.02173	-0.03957
9 LSLCDA	0.23370	0.04062	0.10527	0.26451	0.89285◄	-0.05320	0.18104	-0.06442	-0.13047	0.03688	-0.01690
10 LONCDA	-0.09218	-0.93769◄	-0.13900	0.08653	-0.06265	-0.10790	0.09330	-0.01332	0.20604	-0.06498	0.09068
11 LOFCDA	0.10628	0.93293	0.13814	-0.09472	0.04671	0.09355	-0.08519	0.03238	-0.21940	0.06203	-0.10695
12 MAF	-0.04893	0.14854	0.96180◄	-0.08842	0.12323	0.00490	0.11625	-0.05576	-0.00304	0.02048	-0.01727
13 MASP	0.02671	-0.19271	0.07231	0.04774	-0.08037	-0.05434	0.48979	-0.12257	0.29167	-0.76473◄	0.14230
14 AIGYP	-0.07321	0.09633	0.96649	-0.13239	0.07281	-0.00834	0.10298	-0.04639	0.01681	-0.06449	0.03289
15 MJANT	0.05722	0.21229	-0.03127	0.15927	0.11992	-0.01742	-0.06730	0.05201	-0.94196	0.09013	-0.05482
16 ST	-0.12330	0.06203	-0.17572	-0.10869	-0.16471	0.03523	-0.93254◄	0.06923	-0.10642	0.11092	-0.06919
17 JAMNT	0.01605	0.20469	0.01602	0.13903	0.09453	-0.00975	-0.10220	0.09062	-0.94554◄	0.08566	-0.06548
18 PE	-0.20172	0.11692	-0.07044	-0.15927	-0.21700	0.05950	-0.91159	0.06895	-0.05136	0.12705	-0.04644

CUMULATIVE PROPORTION OF TOTAL VARIANCE

1	2	3	4	5	6	7	8	9	10	11
0.24282	0.46959	0.62705	0.72677	0.79578	0.84957	0.89984	0.93791	0.96158	0.97959	0.99496

◄ DEFINING VARIABLES SELECTED FOR INCLUSION IN THE MULTIPLE REGRESSION ANALYSES

The basic 15 x 15 simple correlation matrix has been reproduced in Table 4-11; and the results of the regression analysis for each of the dependent variables have been summarized in the following sections.

4.4.2.2.1. Summary of the Stepwise Regression Analysis, After Variable Screening, for LMY. A summary of the results of the stepwise multiple regression analysis, after variable screening by factor analysis, for the dependent variable LMY is presented in Table 4-12. The first variable to enter the regression was LCDA which accounts for 53% of the total variance in the dependent variable. The subsequent addition of four other independent variables accounted for an additional 25% of the variance.

All of the regression coefficients in the final equation (5-6) are significant at the 1% level, and their signs seem to correspond to physical theory. The LCDA variable has a positive sign as anticipated in that larger drainage areas are associated with larger annual yields. The positive coefficient of the MAP term is in accordance with that variable's role as an index of available moisture supply. The negative effect of LNCDA is in accordance with the expected relationship. The LMCS has a positive coefficient which is indicative of the greater efficiency in the collection and transport of runoff in basins with higher slopes. The final term in the equation, LSLCDA, has a positive coefficient which demands some further explanation. The possibility of a spurious relationship is evident in that the area of lakes and swamps represents a measure of moisture storage, and therefore, might be expected to have a negative effect on runoff volume. However, this measure as defined in the present study included only those lakes and swamps adjacent to and part of the drainage network. In this respect these areas may indicate areas of highly efficient runoff collection in that moisture available in these areas is in fact already in the drainage network. On this basis it may be suggested that the positive relationship of LSLCDA to the LMY may have a physical basis and is not a spurious relationship.

The statistical significance of the multiple regression relationship is confirmed by the highly significant F statistic. The R^2 statistic indicates that 78% of the total variance in LMY is accounted for by the regression. This is only slightly less than the 82% accounted for by the corresponding equation 4-1 in the all variable analysis. The equation 4-5 resulting after the variable screening process has the great advantage that the signs of the regression coefficients conform to physical expectations.

4.4.2.2.2 Summary of the Stepwise Regression Analysis, After Variable Screening, for LM10YY. A summary of the results of the stepwise multiple regression analysis, after variable screening, for the dependent variable LM10YY is presented in Table 4-13. The first variable to enter the regression is again LCDA, which in this case, accounts for 66% of the variance in the dependent variable. The next four variables contribute an additional 15% to this explained variance.

TABLE 4-11

CORRELATION MATRIX FOR DEPENDENT VARIABLES AND SELECTED INDEPENDENT MEASURES FOR THE FULL STUDY AREA

(N=146)

VARIABLE	1	2	3	4	5	6	7	8
1 LMY	1.000	0.918	0.922	0.671	-0.310	0.729	-0.090	0.212
2 LM10YY		1.000	0.849	0.761	-0.267	0.811	-0.146	0.266
3 LMF			1.000	0.818	-0.265	0.716	-0.157	0.160
4 LM10YF				1.000	-0.127	0.719	-0.200	0.175
5 LNCDA					1.000	0.019	-0.413	-0.186
6 LCDA						1.000	-0.413	0.302
7 LMCS							1.000	0.083
8 LBS								1.000

VARIABLE	9	10	11	12	13	14	15
1 LMY	0.120	0.584	-0.069	0.349	0.218	-0.488	0.034
2 LM10YY	0.159	0.584	-0.063	0.211	0.216	-0.463	0.026
3 LMF	0.059	0.466	-0.023	0.272	0.187	-0.419	0.022
4 LM10YF	0.056	0.316	0.066	-0.004	0.171	-0.339	-0.062
5 LNCDA	0.247	-0.158	-0.086	-0.181	-0.279	0.229	0.239
6 LCDA	0.192	0.484	-0.166	-0.046	0.074	-0.302	0.079
7 LMCS	-0.006	-0.149	0.533	-0.006	0.483	-0.269	-0.373
8 LBS	0.122	0.198	0.257	-0.017	0.152	-0.164	-0.006
9 LALCDA	1.000	0.401	0.157	-0.189	0.048	-0.194	0.262
10 LSLCDA		1.000	-0.114	0.207	-0.022	-0.378	0.238
11 LDDCDA			1.000	-0.277	0.353	-0.149	-0.405
12 MAP				1.000	0.071	-0.273	0.033
13 MASP					1.000	-0.613	-0.450
14 ST						1.000	0.192
15 JAMXT							1.000

Table 4-12

- SUMMARY OF STEPWISE MULTIPLE REGRESSION RESULTS FOR THE FULL STUDY AREA
ANALYSIS OF LMY AFTER INDEPENDENT VARIABLE SCREENING
(N=146)

Step Number (1)	Regression Equations[1] (2)	R (3)	R^2 (4)	F*** (5)	S.E. (6)	S.E.% of $\overline{LMY}$ (7)
	$\overline{LMY}$ = 4.14614				.87969	21.2
1	LMY = 1.269 +1.159 LCDA (.09)	.729	.532	163.6	.6040	14.6
2	LMY = -0.874 +1.187 LCDA +0.119 MAP (.08) (.01)	.824	.679	151.1	.5020	12.1
3	LMY = -0.285 +1.192 LCDA +0.104 MAP -0.433 LNCDA (.07) (.01) (.07)	.864	.746	138.9	.4482	10.8
4	LMY = -1.080 +1.304 LCDA +0.110 MAP -0.313 LNCDA +0.330 LMCS (.07) (.01) (.06) (.10)	.875	.765	114.9	.4322	10.4
5	LMY = -0.780 +1.190 LCDA +0.106 MAP -0.280 LNCDA +0.329 LMCS +0.250 LSLCDA (.08) (.01) (.07) (.09) (.08) (4-5)	.885	.782	100.7	.4176	10.1

---

[1] A t test has been employed to test the significance of each of the regression coefficients. All regression coefficients are significant at the 1% level except where noted by * or **. The * indicates the coefficient is significant at the 5% level only; and the ** indicates the coefficient is not significant at the 5% level:

*** All F statistics are significant at the 1% level.

Table 4-13

- SUMMARY OF STEPWISE MULTIPLE REGRESSION RESULTS FOR THE FULL STUDY AREA
ANALYSIS OF LM10YY AFTER INDEPENDENT VARIABLE SCREENING
(N=146)

Step Number (1)	Regression Equations[1] (2)	R (3)	R^2 (4)	F*** (5)	S.E. (6)	S.E. % of $\overline{LM10YY}$ (7)
	$\overline{LM10YY}$ = 4.73958				.67723	14.3
1	LM10YY = 2.276 +0.993 LCDA (.05)	.811	.658	276.9	.3975	8.4
2	LM10YY = 2.544 +0.999 LCDA -0.359 LNCDA (.05) (.05)	.859	.738	201.4	.3491	7.4
3	LM10YY = 1.631 +1.010 LCDA -0.312 LNCDA +0.048 MAP (.05) (.05) (.01)	.882	.778	166.1	.3223	6.8
4	LM10YY = 1.848 +0.919 LCDA -0.287 LNCDA +0.041 MAP +0.183 LLSCDA (.05) (.05) (.01) (.05)	.891	.794	135.7	.3119	6.6
5	LM10YY = 1.337 +0.992 LCDA -0.210 LNCDA +0.045 MAP +0.182 LSLCDA +0.212 LMCS (.06) (.05) (.01) (.05) (.07) (4-6)	.899	.807	117.3	.3026	6.4

---

[1] A t test has been employed to test the significance of each of the regression coefficients. All regression coefficients are significant at the 1% level except where noted by * or **. The * indicates the coefficient is significant at the 5% level only; and the ** indicates the coefficient is not significant at the 5% level.

*** All F statistics are significant at the 1% level.

All of the regression coefficients are significant at the 1% level. The independent variables and the signs of their coefficients correspond with those of equation 5-6 as discussed in the preceding section (5.4.1.1). The physical bases for these relationships are similar to those discussed with respect to the LMY analysis in the preceding section.

The regression equation is highly significant on the F test of an analysis of variance. The R^2 value of .807 indicates that 81% of the total variance in the dependent variable has been explained by the regression. This figure is only slightly lower than the 82% explained by the corresponding multiple regression equation (4-2) calculated without variable screening. The equation (4-6) developed in this section has the advantage of conforming to physical theory.

4.4.2.2.3. Summary of the Stepwise Regression Analysis, After Variable Screening, for LMF. Table 4-14 contains the results of the stepwise multiple regression analysis, after variable screening, for the dependent variable LMF. The first variable to enter the regression is the LCDA which explains 51% of the total variance in the dependent variable. The subsequent entry of five additional variables contributes a further 21% to this explained variance.

The regression coefficients for the first five dependent variables are significant at the 1% level, while that of the sixth variable, the JAMXT, is significant at the 5% level. The signs associated with each of these coefficients correspond to the relationships anticipated on the basis of physical theory. The LCDA and LNCDA terms with their positive and negative coefficients, respectively, indicate the relationship of drainage area to flood magnitude. The MAP term provides an index of moisture supply and is therefore positively related to the LMF. The LDDCDA is positively related to LMF, indicating that a denser drainage network has a greater flood potential. The LBS index has a negative coefficient which indicates that basins with an elongated shape have lower flood flows than do basins with more rotund shapes. The final variable in the equation, JAMXT, has a positive coefficient. This relationship might be anticipated in that the higher values of JAMXT, which is estimated in degrees below freezing, are related to areas with a more permanent snowpack. This more permanent snowpack provides a basis for larger spring floods.

The analysis of variance F statistic indicates that the regression equation is highly significant. Unfortunately, the coefficient of multiple determination, R^2, is 0.72 indicating that only 72% of the total variance of the dependent variable is accounted for by the regression. The standard error of the estimate is correspondingly large being 12.7% of the mean. In summary, the relationship expressed in equation 4-7 conforms to theoretical expectations; however, its predictive value is limited by the low value of R^2 and the high standard error of the estimate.

4.4.2.2.4. Summary of Stepwise Regression Analysis, After Variable Screening, for LM10YF. Table 4-15 summarizes the results of the stepwise multiple regression analysis, after variable screening by factor analysis,

Table 4-14

SUMMARY OF STEPWISE MULTIPLE REGRESSION RESULTS FOR THE FULL STUDY AREA
ANALYSIS OF LMF AFTER INDEPENDENT VARIABLE SCREENING
(N=146)

Step Number (1)	Regression Equations[1] (2)	R (3)	R^2 (4)	F*** (5)	S.E. (6)	S.E. % of $\overline{\text{LMF}}$ (7)
	$\overline{\text{LMF}}$ = 2.06213				.62776	23.6
1	LMF = 0.048 +0.812 LCDA (.07)	.716	.512	151.1	.4400	16.5
2	LMF = -0.563 +0.828 LCDA +0.059 MAP (.06) (.01)	.778	.605	109.5	.3973	14.9
3	LMF = -0.198 +0.830 LCDA +0.063 MAP -0.272 LNCDA (.06) (.01) (.05)	.810	.657	90.6	.3716	14.0
4	LMF = -0.407 +0.866 LCDA +0.070 MAP -0.244 LNCDA +0.437 LDDCDA (.05) (.01) (.05) (.13)	.827	.683	76.0	.3583	13.5
5	LMF = -0.235 +0.943 LCDA +0.073 MAP -0.278 LNCDA +0.591 LDDCDA -0.476 LBS (.06) (.01) (.06) (.13) (.13)	.844	.712	69.1	.3431	12.9
6	LMF = -0.435 +0.951 LCDA +0.074 MAP -0.311 LNCDA +0.733 LDDCDA -0.528 LBS +0.013* JAMXT (.06) (.01) (.06) (.14) (.13) (.01)	.851	.724	60.7	.3370	12.7
	(4-7)					

[1] A t test has been employed to test the significance of each of the regression coefficients. All regression coefficients are significant at the 1% level except where noted by * or **. The * indicates the coefficient is significant at the 5% level only; and the ** indicates the coefficient is not significant at the 5% level.

*** All F statistics are significant at the 1% level.

Table 4-15

- SUMMARY OF STEPWISE MULTIPLE REGRESSION RESULTS FOR THE FULL STUDY AREA
ANALYSIS OF LMIOYF AFTER INDEPENDENT VARIABLE SCREENING
(N=146)

Step Number (1)	Regression Equations[1] (2)	R (3)	R^2 (4)	F*** (5)	S.E. (6)	S.E.% of $\overline{\text{LMIOYF}}$ (7)
	$\overline{\text{LMIOYF}}$ = 3.29262				.53375	16.2
1	LMIOYF = 1.571 +0.694 LCDA (.05)	.719	.517	154.4	.3721	11.3
2	LMIOYF = 1.595 +0.724 LCDA +0.407 LDDCDA (.05) (.12)	.743	.553	88.3	.3595	10.9
3	LMIOYF = 1.688 +0.725 LCDA +0.394 LDDCDA -0.125* LNCDA (.05) (.12) (.05)	.754	.568	62.3	.3544	10.8
4	LMIOYF = 1.840 +0.778 LCDA +0.494 LDDCDA -0.151 LNCDA -0.330* LBS (.06) (.12) (.06) (.13)	.766	.587	50.1	.3478	10.6
	(4-8)					

[1] A t test has been employed to test the significance of each of the regression coefficients. All regression coefficients are significant at the 1% level except where noted by * or **. The * indicates the coefficient is significant at the 5% level only; and the ** indicates the coefficient is not significant at the 5% level.

*** All F statistics are significant at the 1% level.

for the dependent variable LM10YF. The LCDA variable is again dominant, and is first to enter the regression explaining 52% of the total variance in the LM10YF. The addition of three more variables contributed a further 7% to this explained variance.

The first three regression coefficients in equation 5-9 are significant at the 1% level, while the fourth, which is associated with the LBS measure, is significant at the 5% level. The signs of all the coefficients conform to anticipated relationships which are the same as discussed for equation 4-7 in the preceding section.

The analysis of variance F statistic confirms the significance of the regression equation at the 1% level. However, the low R^2 value indicates the limited usefulness of the relationship for predictive purposes.

4.4.4. Summary of Results for the Full Study Area Analysis

Two approaches to the analysis of the relationships of the selected streamflow characteristics to the various measures of climatic and other physical geographic patterns for the entire study area, have been discussed in the preceding sections of this chapter. The first approach involved the utilization of stepwise multiple regression analysis to examine the relationships of each of the dependent variables to the full set of independent measures; while the second approach employed factor analytic techniques in the screening of the independent variables prior to a regression analysis for each of the dependent variables. In order to facilitate a comparative summary of these two approaches to the analysis, Table 4-16 has been compiled, summarizing the results from the two sets of analyses.

Several comparative observations may be made based on the data in the table. In the case of the all variable regression approach, there is a tendency for spurious relationships to be developed as indicated when regression coefficients have the opposite sign to that expected on the basis of physical theory. In the case of regression analysis after variable screening, the signs of the regression coefficients corresponded to those expected on the basis of physical theory. The R^2 values, indicating the proportion of the total variance in the dependent variable explained by the selected independent variables, were observed to be slightly higher for relationships developed by the first approach. The standard errors of the estimates were similar for the two approaches with those of the first approach being slightly lower in three out of the four cases examined.

The stated aims of the present study are twofold; first, to examine some of the physical relationships which underlie prairie streamflow patterns, and second, to develop statistical relationships for the prediction of streamflow characteristics for ungauged streams. On the basis of the data in Table 4-16, it seems that the first aim is best served by the second approach to the analysis in that the relationships derived in this manner correspond more closely to physical expectations. The second

TABLE 4-16 - COMPARATIVE SUMMARY OF ANALYSIS RESULTS FOR THE FULL STUDY AREA (N=146)

Dependent Variable	Approach I: Stepwise Multiple Regression All Variables					Approach II: Stepwise Multiple Regression After Factor Analytic Variable Screening				
	Equat No.	No. of Steps	Reg. Coef. Conform to Theory	R^2	S.E. % Mean	Equat No.	No. of Steps	Reg. Coef. Conform to Theory	R^2	S.E. % Mean
LMY	4-1	5	No	.82	9.2	4-5	5	Yes	.78	10.1
LM10YY	4-2	4	No	.82	6.1	4-6	5	Yes	.81	6.4
LMF	4-3	6	No	.72	12.8	4-7	6	Yes	.72	12.7
LM10YF	4-4	4	No	.61	10.3	4-8	4	Yes	.59	10.6

aim, that of prediction, is slightly better served by the first approach to the analysis; however, this slight advantage is offset by the possibility of spurious relationships. The relatively large standard errors and low values for R^2 derived from all of the analyses limit the possible use of the regression equations for prediction purposes. This is particularly true when it is recognized that the dependent variables have been transformed to logarithmic units, and the process of converting these measures back to arithmetic values results in considerably larger percentage standard errors than are indicated by the logarithmic measures in the table.

The limitations of the regression equations for predictive purposes must initially be attributed to either errors in variable measurements or to variables which have not been considered in the analysis. Several groups of possible variables including measures of precipitation intensity, soil infiltration rates, and landuse patterns, have not been included in the present analysis because of a lack of available data. While it was not considered feasible to add measures of these variables at this stage of the study, an alternative approach to increasing the strength of the regression relationships was proposed. This proposal involved the division of the study area into regions of hydrologic similarity and the subsequent repetition of the multiple regression analyses on a regional basis. The division of the study area into hydrologic regions and the subsequent regression analyses are discussed in the following sections of this chapter.

4.5 ANALYSIS STAGE II: HYDROLOGIC REGIONS

In an effort to improve the predictive value of the regression equations resulting from the analyses, it was proposed to divide the study area into hydrologically similar regions. Such a division of the study area in order to improve the strength of relationships has been employed with some success in several similar studies in other areas. Examples of such an approach are included in the work of Durrant and Blackwell (1959), Dalrymple (1960), Benson (1964), Coulson (1967b), Solomon et al. (1968), and Canada, Department of Energy, Mines and Resources (1970).

Regions of hydrologic similarity are delimited for the purpose of accounting for some of the regional variations in the relationships of streamflow characteristics to physical geographic patterns. There are several possible methods for the delimitation of such regions. Two such methods which have been employed in other studies are regionalization on the basis of residual plots, and regionalization on the basis of index-ratios of selected hydrologic events. A third alternative approach involves the use of multi-variate techniques to delimit regions of similarity. Each of these three methods has been considered and tested in the present study.

4.5.1 Regional Subdivision on the Basis of Residual Plots

The residuals which result from the estimation of a regression equation for a given set of data may indicate some geographical pattern to these errors. Several other researchers working in different study areas have successfully employed the spatial distribution of residual errors as a basis for the subdivision of their study areas (see for example the work of Benson 1964; Solomon et al. 1968; and Canada, Department of Energy, Mines and Resources 1970).

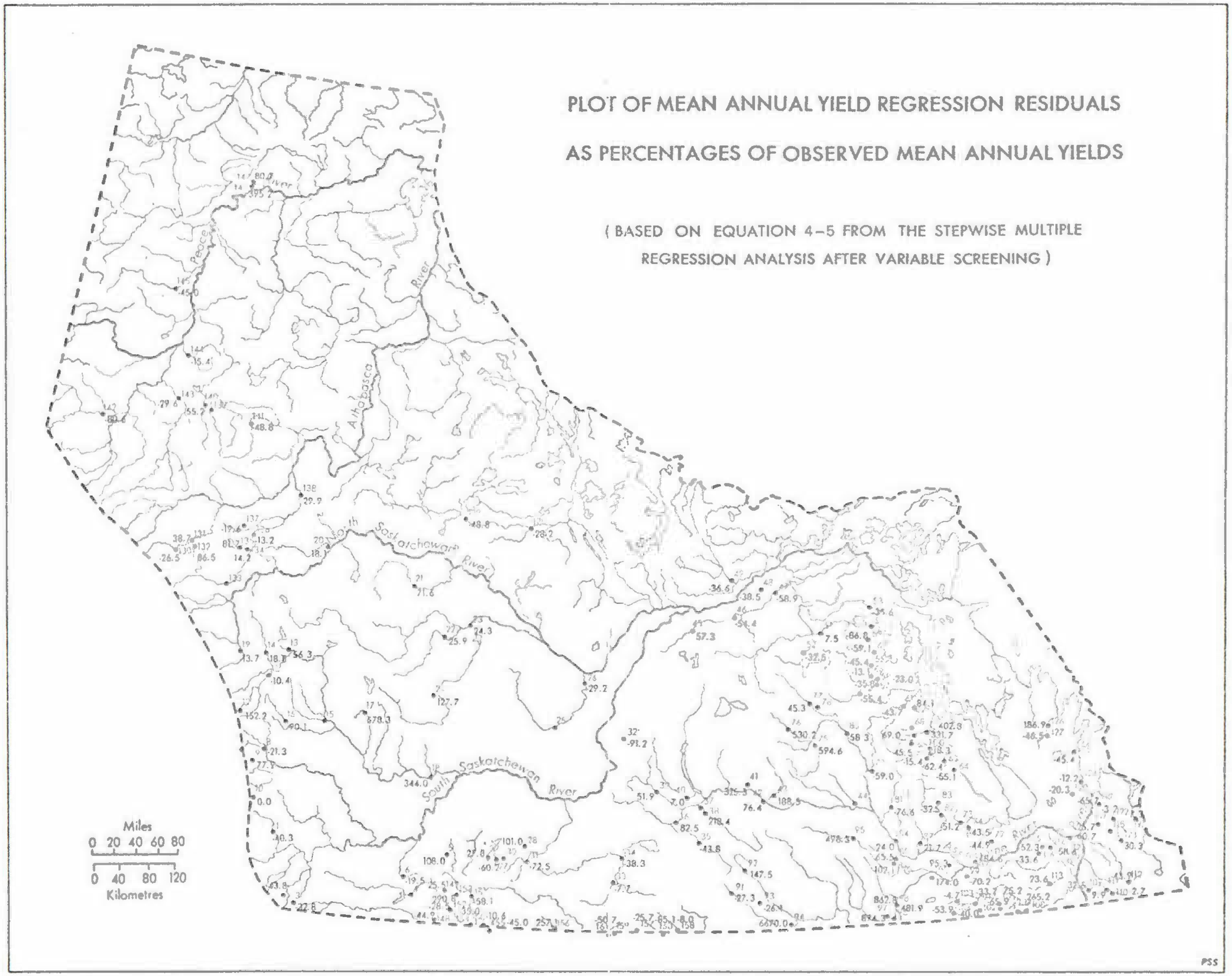
In the present study, the residuals which resulted from the stepwise multiple regression analyses after variable screening were considered as a possible basis for the subdivision of the study area. For each of the four dependent variables, a map was plotted to show the spatial distribution of the residuals from the regression analysis. The residuals plotted were expressed as percentages of the observed values of the dependent variables. The map which resulted from the plotting of the residuals for the LMY relationship (equation 4-5), has been reproduced as Figure 4-1. The pattern of residuals on this map is similar to those obtained for the regression relationships of the other three dependent variables.

The pattern of the residuals as plotted on Figure 4-1 does not reveal any obvious regional divisions. The only residual groupings which are observed are local clusters of only a few basins. These local patterns were considered to represent divisions on too small a scale for use in the present research. Such small divisions include so few of the study basins that meaningful statistical analyses would be impossible without the addition of more basins. On the basis of the residual plots for the four regression relationships which resulted from the stepwise multiple regression analyses after variable screening, the use of residual plots for the subdivision of the study area was ruled out.

4.5.2 Regional Subdivision on the Basis of Index-Event Ratios

The index-event method of regional subdivision has been widely employed in flood frequency studies by the United States Geological Survey; and the method has been outlined in detail by Dalrymple (1960). A similar method was employed by Blackwell and Durrant (1959) in their flood frequency study for the southern prairies. In this method, regional subdivision is based on the ratio of some index-event to the mean event. The index-event is usually an event with a return period of 10 years. The index-ratio is a measure of the year to year variability in the magnitudes of the particular event being analysed. If the spatial distribution of index-ratios is such as to indicate regions of hydrologic similarity, the analyses for each area can be concentrated on explaining the mean magnitudes of events in terms of the physical basin characteristics. The estimated mean events in combination with regional index-ratios provide a basis for the estimation of the magnitudes of events with any required return period.

In the present study, the ratios of the 10 year to the mean events were estimated for both annual yields and annual flood flows for each of the study basins. These ratios were then plotted on maps of the study area. The map of the index-ratios for the annual yield events has been reproduced as Figure 4-2. A similar pattern resulted from the plot of the index-ratios for annual flood flow events. On the basis of these maps, it was observed that the spatial pattern of index-ratios as plotted for the study basins did not lend itself to regional grouping except on



PLOT OF MEAN ANNUAL YIELD REGRESSION RESIDUALS
AS PERCENTAGES OF OBSERVED MEAN ANNUAL YIELDS

(BASED ON EQUATION 4–5 FROM THE STEPWISE MULTIPLE
REGRESSION ANALYSIS AFTER VARIABLE SCREENING)

Figure 4 - 1

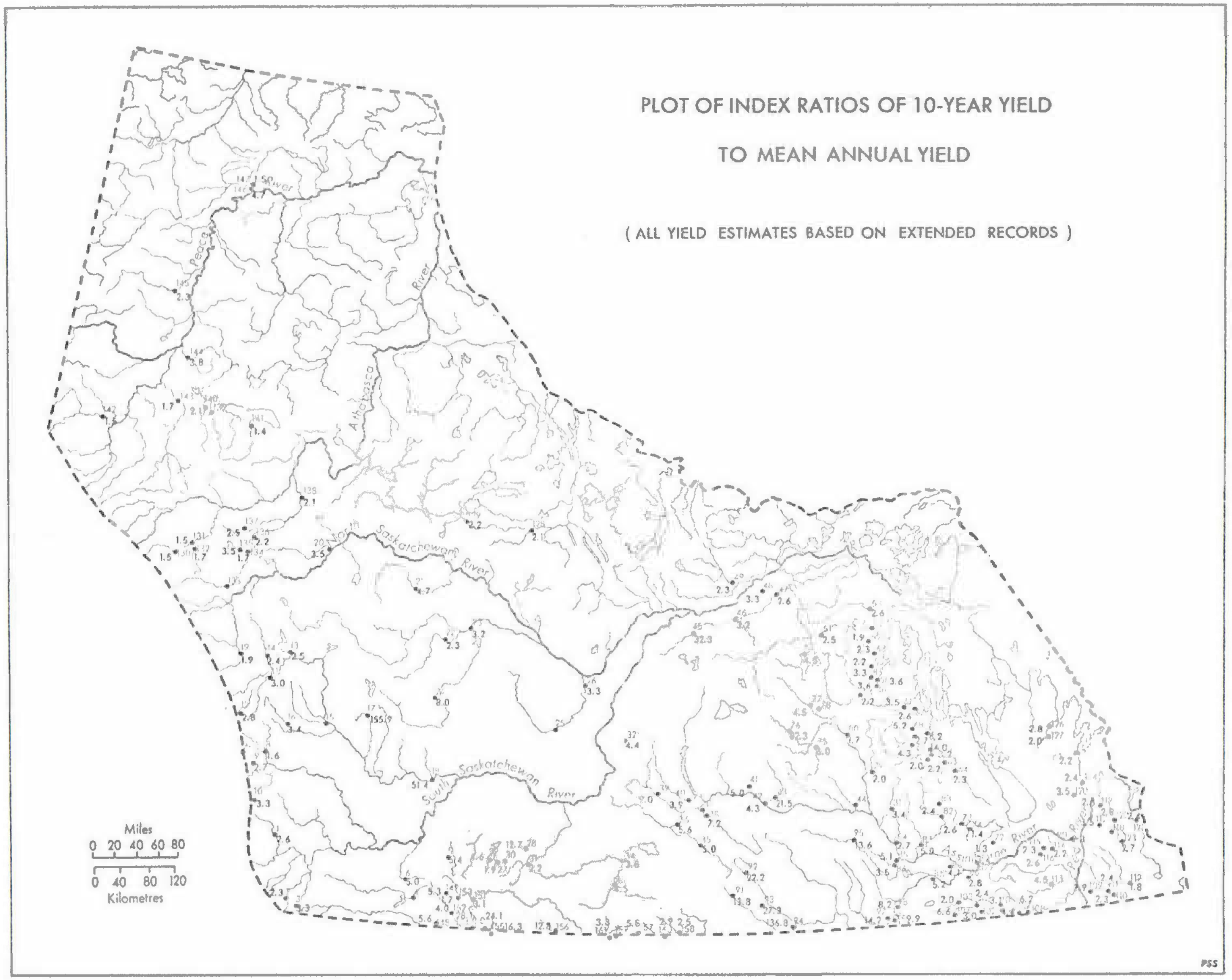

PLOT OF INDEX RATIOS OF 10-YEAR YIELD
TO MEAN ANNUAL YIELD
(ALL YIELD ESTIMATES BASED ON EXTENDED RECORDS)
Peace River
Athabasca River
North Saskatchewan River
South Saskatchewan River
Assiniboine River
Miles
0 20 40 60 80
0 40 80 120
Kilometres
PSS

Figure 4 - 2

a local scale. On the basis of these observed patterns, regional subdivision based on index-ratios was ruled out for the present research.

4.5.3 Regional Subdivision by Multivariate Grouping Techniques

In view of the failure of the residual plots and index-ratio plots to provide a basis for a meaningful subdivision of the study area, it was proposed to employ a multivariate technique to subdivide the study area according to indices of climatic and hydrologic variability. As discussed in a previous section, the large scale regional variations in streamflow patterns tend to be closely related to climatic conditions which govern the available moisture supply, while smaller scale local variations in streamflow are more closely related to variations in other physical geographic patterns. Building on this premise, it was proposed to approach the regional subdivision problem from the point of view of the large scale climatic patterns which control the available moisture supply.

The multivariate grouping technique utilized in the present study employed factor analytic methods in conjunction with an optimal grouping algorithm. A factor analysis with a principal components solution and varimax rotation was used in the identification of the most significant dimensions underlying the set of available climatic and hydrologic variables as compiled for the study basins. Factor scores were calculated for each observation on three factors. These factor scores were then employed as input to an optimal grouping computer program as developed by Semple et al. (n.d.). This grouping program has recently been employed in a study of microclimatic zonation in Northern Alberta (MacIver 1970).

The initial set of variables employed in the grouping analysis included all of the 20 climatic measures, and the index-event ratios for annual yields and annual flood flows as compiled for each of the study basins. All of the climatic measures were originally selected on the basis of their theoretical relationships to streamflow, and therefore, were considered to be potentially useful in the regional subdivision problem. The index-event ratios for the annual yields and annual flood flows are measures of the year to year variability in streamflow patterns, and therefore, were considered for inclusion in the grouping analysis of the study basins.

The first step in the grouping procedure was to execute a factor analysis resulting in the rotated factor matrix which has been reproduced as Table 4-17. Only the first four factors have been reproduced in the table. From the rotated factor matrix, the four factors may be identified in terms of their highest loadings. The first factor is an index of moisture availability. The second factor is an index of the intensity of winter temperatures. The third factor is an index of spring and summer temperature conditions; and the fourth factor is an index of the year to year variability in streamflow.

In prepartion for basin grouping, the factor analytic procedure was extended a further step and factor scores were calculated for each observation on each of the first four factors. The individual factor scores for each observation and dimension were estimated by summing the products of

TABLE 4-17

ROTATED FACTOR MATRIX FROM THE FACTOR ANALYSIS OF THE CLIMATIC VARIABLES AND HYDROLOGIC INDEX EVENT RATIOS

(N=146)

	FACTOR			
	1	2	3	4
VARIABLE				
1 LYR	-0.27546	-0.02521	0.18370	0.85435
2 LFR	-0.33616	-0.12292	0.10643	0.81333
3 MAP	0.94427	0.00780	-0.01139	-0.26472
4 MAS	0.77806	-0.25243	-0.53428	-0.02732
5 MASP	0.20104	-0.32233	-0.77741	0.20141
6 MWP	0.87893	0.10123	-0.32611	-0.03114
7 MSP	0.83273	-0.35328	0.08532	-0.12922
8 MWSP	0.96352	-0.17102	-0.11021	-0.09668
9 A10YP	0.95616	-0.03363	-0.03332	-0.19055
10 W10YP	0.84298	0.14730	-0.35348	0.03566
11 MWT	0.05767	-0.98876	-0.02008	0.05985
12 MJANT	-0.04030	0.98681	0.10524	-0.06300
13 MST	-0.17764	0.07635	0.90669	0.19970
14 MJUNT	-0.03493	0.41826	0.86094	0.11045
15 WMXT	-0.01039	-0.98157	-0.11705	0.04003
16 JAMXT	0.00781	0.97717	0.15449	-0.05215
17 PE	-0.04979	0.01648	0.93952	0.19753
18 THAET	0.87871	0.01553	0.09153	-0.33981
19 TUAET	0.84284	-0.38643	0.28569	-0.16368
20 LTHSUR	0.84937	0.10192	-0.29741	0.02062
21 TUSUR	0.90234	0.16760	-0.13122	-0.28301
22 MATR	-0.07010	0.88645	0.43708	0.04208

CUMULATIVE PROPORTION OF TOTAL VARIANCE

0.45531	0.71513	0.85424	0.91671

the normalized raw input data and factor loadings for each variable. The factor scores as calculated for each of the study basins are listed in Table 4-18.

The optimal grouping algorithm employed in the present analysis utilizes three sets of factor scores as a basis for the grouping of the observations into a set of optimal groups. This method employs the three factor scores for each observation to fix its location in three dimensional space. The distance between pairs of observations is then employed to estimate group centroids in a stepwise fashion. The optimal grouping is controlled by two criteria, the first defining the minimum explained variance to be associated with an acceptable group and the second defining the maximum total explained variance for all groups. The grouping calculations in the present study were made utilizing a computer program prepared by Semple et al. (n.d.) and modified by MacIver (1970).

The optimal grouping algorithm was applied to the factor scores of the first three dimensions of the rotated factor matrix, Table 4-17. The limiting criteria were set at 1% for the minimum variance explained by an acceptable group and at 95% for the maximum variance explained by all groups. The calculations resulted in the identification of two groups which accounted for a total of only 19% of the total variance. Although the total variance explained by these groups is low, when they were plotted on a map of the study area, two well defined areal groupings were observed.

The first group of basins included those in the eastern part of the study area and those in northern Alberta. This group included 83 basins. The centroid of the group had factor scores of 0.35, 0.65 and 0.15 for the first, second and third dimensions respectively. These factor scores may be interpreted to characterize the first group of basins as having above average moisture supply, relatively extreme winter conditions, and warmer spring and summer temperatures with higher PE.

The second group of basins included those in the southwestern section of the study area. This group included 63 basins. The centroid of this second group had factor scores of -0.45, -0.88 and -0.19 for the three dimensions. Interpreting these scores the area might be characterized as having below average moisture supply, less extreme winter conditions and slightly less extreme summer temperatures. The third factor, which relates to spring and summer temperatures, shows a much less pronounced differentiation between groups than do the first two.

The map, Figure 4-3, was prepared to illustrate the grouping of the study basins. Each hydrometric station has been plotted to indicate the group to which it has been assigned and its distance from the group centroid. This distance is the distance relative to the three dimensional plots employed in the grouping procedure. The general pattern of basin grouping as illustrated on the map is contradicted in three areas by seemingly anomalous stations. In order to smooth the tentative regional boundary each of the anomalous areas was examined in detail with the aim of explaining the assignment of basins to a particular group.

TABLE 4-18

FACTOR SCORES RESULTING FROM FACTOR ANALYSIS OF THE CLIMATIC VARIABLES AND HYDROLOGIC INDEX EVENT RATIOS

(N=146)

BASIN NO.	FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4
1	1.35450	-2.08698	-2.22653	0.59946
2	2.28512	-2.43297	-1.33724	1.35398
3	1.12879	-2.79941	-1.22787	0.63033
5	-1.79255	-1.35280	-0.23772	-0.11667
6	-0.69237	-1.61579	1.11490	0.77676
8	1.66278	-2.30929	-1.61127	-0.46227
9	1.80452	-2.08366	-1.76565	0.07424
10	2.41931	-1.99794	-2.27155	0.94244
11	0.42246	-1.04715	-0.55916	-0.66929
12	2.50641	-1.97922	-2.56039	0.30717
13	0.22754	-0.98037	-0.56463	-1.12027
14	0.76126	-1.21907	-0.68011	-0.73835
16	-0.27138	-1.72261	-0.07112	-0.70516
17	-0.55771	-0.45036	-0.54134	4.41912
18	-1.23569	-0.23765	0.12666	3.87795
19	0.51755	-1.45303	-0.53736	-0.83347
20	0.18901	-0.51214	-0.19037	-0.25669
21	-0.56799	0.62496	-0.29419	-0.37126
22	-1.42925	-0.16643	-0.28986	-1.24476
23	-1.09920	-0.30009	-0.26451	-0.86256
24	-1.64241	-0.06218	-0.59454	0.47403
25	-1.33113	0.41583	0.01304	0.97209
27	-0.67580	-1.37069	-0.88700	-1.15936
28	-0.75454	-0.75151	-0.48137	1.21397
29	-0.55932	-1.03848	1.01931	0.06637
31	-0.90202	-1.04561	-0.37910	-0.82636
32	-1.44936	0.14701	0.59349	0.09231
33	-1.35107	-0.96679	0.78776	-0.05177
34	-1.31171	-1.14344	0.66627	-0.19577
35	-0.71090	-0.14362	0.72275	-0.95968
36	-0.61459	-0.44054	0.98097	0.37739
38	-0.65235	0.28259	0.50377	0.67272
39	-0.49412	0.13551	0.26299	0.80141
40	-1.16192	0.29637	0.46541	-0.22062
41	0.54753	0.74629	-0.86071	0.90128
42	-0.17592	0.35982	0.10480	0.24994
43	0.39915	0.85756	-0.91074	2.14299
45	-0.60151	1.25896	-0.81767	2.41777
46	-0.41812	1.20003	-0.75290	-0.11091
47	-0.33491	1.29942	-0.75173	-0.39794
48	-0.32905	1.42679	-0.46827	-0.12810
49	-0.24755	1.29669	-0.73207	-0.50712
50	-0.22123	1.37784	-1.27263	0.07328
51	-0.26732	1.23297	-1.13995	-0.32774
53	0.09910	1.66676	-1.36407	-0.19330
54	-0.60022	1.13739	-0.52558	-0.73894
55	-0.00437	1.41220	-1.24519	-0.14212
56	-0.41331	1.20159	-0.63007	-0.21709
57	-0.78387	1.03283	-0.12431	-0.49741
58	-0.51651	1.15407	-0.51973	-0.05789
59	-0.02131	1.60713	-1.24497	-0.25440
60	0.00574	1.41313	-1.24699	-0.06601
61	-0.40474	0.90276	0.04354	-0.36921
62	0.87992	0.54242	0.48071	-0.11590
63	1.00312	0.49524	0.57568	-0.10574
64	0.96086	0.52632	0.49524	-0.26567
65	0.98143	0.58839	0.49403	1.20549
66	1.06224	0.51097	0.52609	0.23795
67	1.21619	0.54857	0.51289	1.33372
68	1.15455	0.53269	0.51559	0.92687
69	1.11859	0.52399	0.51962	0.65596
70	1.04594	0.50597	0.52524	0.14836
71	0.99223	0.49370	0.53092	-0.23425
72	0.50966	0.43561	0.71293	-0.81212
73	0.16726	0.39287	0.46026	-0.52461
74	0.52198	0.55992	0.47585	-0.74067
75	0.25411	0.97164	-1.00069	0.93272
76	0.38265	2.84656	-0.70712	2.26376
77	-0.45905	1.33772	-1.30329	-0.00627
79	-0.95151	0.96556	-0.02793	-1.17535
80	-0.87078	0.99559	-0.09995	-0.96604
81	-0.97551	0.77556	0.29136	-1.35701
83	-0.10285	0.24608	0.49481	-0.96785
84	-0.16127	0.67251	0.35766	-0.87549
85	1.09725	0.40866	0.15760	1.15203
86	0.76639	0.36897	0.23653	0.31663
87	0.09303	0.66191	0.46613	-0.40079
88	0.50919	0.29976	1.21066	0.45377
89	0.32662	0.31486	1.01835	-0.65090
90	1.09555	0.59236	1.23741	-0.40479
91	-0.67761	-0.32418	0.60423	0.49682
92	-0.92058	0.21694	0.55987	1.21674
93	-0.35599	-0.21666	0.86427	1.11104
94	-0.47765	-0.56109	1.54484	2.86026
95	0.56294	0.33759	-0.04729	1.94736
96	-0.55958	0.34660	0.39559	0.70969
97	0.37700	0.41378	0.51745	1.76068
98	0.28452	0.35436	0.93505	0.55789
99	0.68895	0.12565	1.11480	-0.03340
100	1.34051	-0.03254	1.17031	0.55752
101	0.29663	-0.19902	1.20229	-0.80177
102	-1.12995	-0.38796	1.45702	-0.99363
103	1.53270	-0.07048	0.91776	0.10815
104	1.63791	-0.03948	0.92225	0.77294
105	-1.05614	-0.29041	1.46426	-0.98422
106	-0.96770	-0.26183	1.51491	-0.61178
107	0.06920	0.04598	1.70567	-0.16561
109	0.57532	-0.14745	1.73677	-0.96031
110	0.54911	-0.15174	1.73291	-1.23457
111	0.54914	-0.15444	1.74273	-1.20854
112	1.11477	0.51182	0.74758	-0.82211
113	1.43062	0.11580	1.65081	0.54523
114	1.45126	0.14091	1.26125	-0.41743
115	1.53200	0.12540	1.26611	-0.32491
116	1.37643	0.17686	1.26799	-0.14351
117	0.93494	0.55998	1.22497	0.02830
119	1.09619	0.91622	0.59900	-0.31900
120	1.13115	0.92492	0.55617	-0.07431
121	1.27230	0.60787	0.63768	-0.30492
122	-0.99022	0.33547	1.25192	-1.06990
123	-0.99050	0.33441	1.25067	-1.05491
124	1.10278	0.91441	0.55404	-0.24851
125	1.07693	0.90978	0.95773	-1.43819
126	1.24575	0.96282	0.51424	0.08385
127	1.18367	0.94862	0.52040	-0.35695
128	-1.04613	0.47592	-0.81464	-1.04420
129	-1.02481	0.45737	-0.76550	-1.12941
130	0.78491	-1.70349	-0.71271	-1.57118
131	0.41447	-1.31323	-0.81276	-1.54910
132	0.59391	-0.95136	-0.91981	-1.09451
134	0.41887	-1.00227	-0.95036	-1.32482
135	0.36826	-0.72256	-0.46176	-0.70191
136	0.21576	-0.59245	-0.32647	-0.88355
137	0.16794	-0.65586	-0.17488	-0.95733
138	0.15891	-0.64876	-0.59830	-1.15995
140	0.14742	0.22433	-0.90364	-0.33484
141	-0.38056	0.17644	-1.64897	-1.13179
142	0.04034	-0.99427	-2.02535	-0.01951
143	0.31583	0.11265	-1.13792	-0.65101
144	0.07041	0.94448	-1.97393	0.72549
145	-0.01942	1.17671	-2.28661	0.33633
146	-0.91312	2.12391	-2.26008	0.25293
147	-1.20912	2.71490	-2.64737	-0.57834
148	-1.52186	-1.36291	0.53889	0.52991
149	-1.37581	-1.11643	-0.96753	0.15008
150	-2.43976	-0.73874	0.11108	1.26072
151	-1.47888	-1.12821	-0.97865	-0.50378
152	-1.43498	-1.24343	-0.31691	-0.11448
153	-1.56811	-1.16326	-0.95128	-1.16865
154	-2.00247	-0.82695	0.49572	1.88996
155	-2.47696	-0.75056	0.10986	1.03623
156	-1.06568	-1.46662	0.91357	1.79678
157	-1.18881	-0.95183	0.96893	0.27173
158	-1.02833	-0.55187	0.69281	-0.67859
160	-0.95886	-0.49527	0.68682	-0.64727
161	-1.50351	-1.11785	0.93616	-0.56593

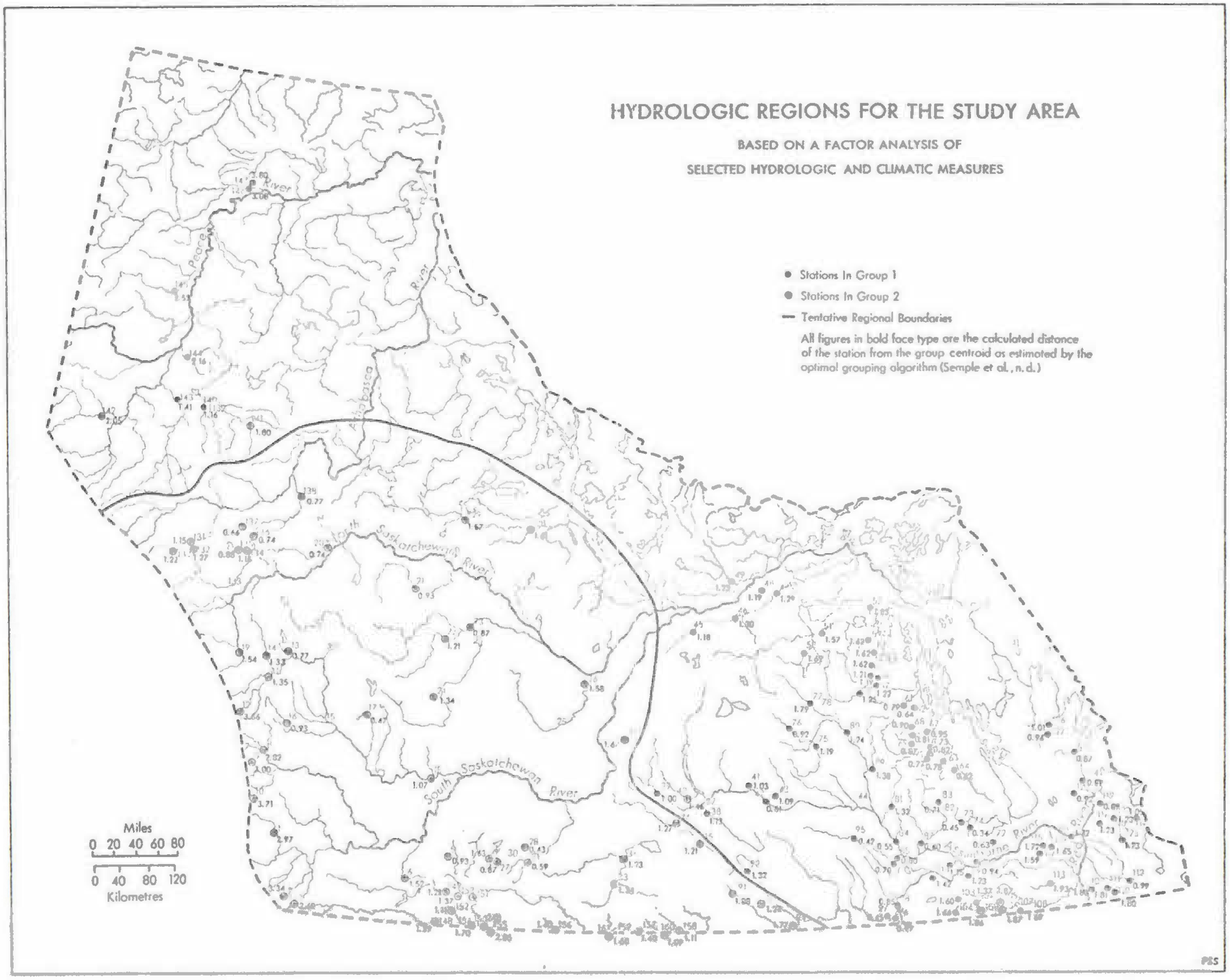

HYDROLOGIC REGIONS FOR THE STUDY AREA
BASED ON A FACTOR ANALYSIS OF
SELECTED HYDROLOGIC AND CLIMATIC MEASURES
● Stations In Group 1
● Stations In Group 2
— Tentative Regional Boundaries
All figures in bold face type are the calculated distance of the station from the group centroid as estimated by the optimal grouping algorithm (Semple et al., n.d.)
Peace River
Athabasca River
North Saskatchewan River
South Saskatchewan River
Assiniboine River
Miles
0 20 40 60 80
0 40 80 120
Kilometres
PSS

Figure 4 - 3

The first anomaly to the basin grouping is in south central Manitoba where basins 102, 105 and 106 were assigned to Group 2, while all of the surrounding basins were assigned to Group 1. An examination of the climatic data for stations in this area revealed that this anomalous grouping resulted from lower precipitation estimates for these basins. These lower estimates were the result of a particular climatological station, Hansboro, North Dakota, Number 190, being included in the Thiessen calculations for the basins. This climatological station which is located in an area of lower precipitation to the southeast of Turtle Mountain, has a mean annual precipitation of 15.6 inches compared to over 18 inches for most of the surrounding stations. It is not possible on the basis of available data to indicate the true effect of this area of lower precipitation on the basins in question. However, having explained the reasons for the anomaly, and recognizing that these three basins are not closely linked to the centroid of Group 2 as indicated by their distance statistics of over 1.8 (see Figure 4-3), it was decided to assign these basins to Group 1 for the purposes of analysis.

The second anomaly to the grouping pattern is in south central Saskatchewan where basin 40 was assigned to Group 2 while nearby basins 38 and 39 were assigned to Group 1. An examination of the basic climatic data for these stations revealed that the precipitation estimates for basin 40 were lower than for either 38 or 39. This lower precipitation can be attributed to the influence of a particular climatological station, Davidson, Saskatchewan, No. 81. This climatological station is located in a drier area to the northwest of basin 40. After examining the distance statistics from the centroids of their respective groups, 1.46 for basin 40, 1.13 for basin 38, and 1.00 for basin 39, and on the basis of the single climatological station involved, it was decided to assign basin 40 to Group 1 for the purposes of analysis.

A third anomalous area in the grouping pattern occurred in northern Alberta where basins 141 and 142 were assigned to Group 1. On examination of the climatic data for these basins it was found that basins 141 and 142 were somewhat drier than were basins 140 and 143. After considering the distance statistics, 1.80 and 2.05 for basins 141 and 142 respectively and 1.16 and 1.41 for basins 140 and 143 respectively, it was decided to assign basins 141 and 142 to Group 1 for the purposes of analysis.

Having examined the anomalies in the spatial distribution of the basin grouping, a tentative boundary was drawn for the regions of hydrologic similarity (Figure 4-3). The first group, the eastern region, included 89 study basins; and the second group, the western region, included 57 basins. Although the grouping analysis accounted for only 19% of the total variance in the factor scores for the three climatic factors, the regional grouping was utilized as a basis for further analyses of the relationships between streamflow characteristics and the various measures of physical geographic patterns. These relationships were analysed for each of the two regions employing the techniques of variable screening by factor analysis and stepwise multiple regression.

4.5.4 Analysis for the Eastern Region

The analysis of the relationships of each of the dependent variables to the various measures of climatic and other physical geographic patterns for the 89 study basins in the eastern region employed the same methods as were utilized in the full study area analysis of Section 4.4.3. The full set of independent variables was screened for multicollinearity by factor analytic techniques; and the resulting selection of independent variables was employed in the stepwise multiple regression analysis for each of the dependent variables.

The full set of 39 independent variables for each of the 89 study basins in the eastern region was factor analysed resulting in the rotated factor matrix as presented in Table 4-19. Eighteen defining variables from Table 4-19 were selected and were factor analysed a second time. The resulting rotated factor matrix is reproduced as Table 4-20. This table served as the basis for the ultimate selection of a set of 10 independent variables for the eastern region. These 10 variables were then employed in a stepwise multiple regression analysis for each of the four dependent variables.

The initial correlation matrix for all of the variables employed in the eastern region multiple regression analyses is reproduced in Table 4-21; and the regression relationships developed are summarized in Tables 4-22 to 4-25 inclusive. The results of the eastern region regression analysis for the dependent variable LMY are presented in Table 4-22. The final equation (4-9) contains six independent variables which combine to explain 73% of the total variance in the dependent variables. The signs of the regression coefficients correspond to those anticipated on the basis of physical theory; and the equation is significant at the 1% level for an analysis of variance F test. The results of the eastern region analysis for the dependent variable LM10YY are presented in Table 4-23. The final regression equation includes five independent variables which combine to explain 81% of the total variance in the dependent variable. The regression coefficients associated with the first four variables conform to physical theory; however, the positive coefficient of the sixth term, the LALCDA, requires an explanation. This coefficient indicates that in basins with a higher percentage area of lakes, the 10 year yield tends to be higher. This relationship may indicate that in high flow years some of the additional water may be derived from lake storage. The analysis of variance confirmed the significance of the regression equation (4-11) at the 1% level. The results of the eastern region regression analysis for the dependent variable LMF are displayed in Table 4-24. The resulting regression equation contains six independent variables and has an R^2 value of 0.66. The regression coefficients for the first five terms are significant at the 1% level and have signs which conform to theoretical expectations. The sixth variable, LMCS, has a coefficient which is not significant at the 5% level and has a negative sign which is contrary to physical theory. In view of the lack of significance of the regression coefficient, it is not possible to give further consideration to the role of LMCS on the basis of the present data. The analysis of variance for equation 4-11 confirmed the significance of the relationship at the 1% level. The results for the analysis of the LM10YF

TABLE 4-19

ROTATED FACTOR MATRIX RESULTING FROM THE FACTOR ANALYSIS OF ALL INDEPENDENT VARIABLES FOR THE EASTERN REGION

(N=89)

	FACTOR										
VARIABLE	1	2	3	4	5	6	7	8	9	10	11
1 LTDA	0.30554	-0.14427	0.09468	-0.01483	0.29859	0.01407	-0.87725◄	-0.06862	0.01866	-0.09358	-0.01157
2 LNCDA	-0.25243	-0.16988	0.06229	-0.04234	-0.03960	-0.18352	-0.05839	-0.06510	-0.11055	-0.91354◄	0.01544
3 LCDA	0.32484	-0.11808	0.08427	0.00544	0.30511	0.04730	-0.87584◄	-0.06015	0.02545	0.04957	-0.01584
4 LMCS	-0.06169	0.11450	-0.50766	0.01055	0.15399	-0.08473	0.35219	0.71028◄	0.10316	0.14397	-0.01493
5 LBEL	-0.10533	-0.25214	-0.40869	-0.34389	0.54613	-0.23697	-0.09808	0.49089	0.15220	-0.10026	-0.06361
6 LBR	0.10101	0.04241	-0.52169	-0.00313	0.36678	-0.17308	-0.27821	0.33493	0.16318	0.15379	-0.04911
7 LBL	0.26536	-0.12142	-0.05458	0.02325	0.27150	0.01779	-0.81627	-0.02268	0.40736	-0.02423	-0.01159
8 LBS	0.04050	-0.01068	-0.30250	0.00055	0.06725	0.01439	-0.24078	0.07333	0.90299◄	0.11762	-0.00528
9 LALTDA	0.15892	-0.16627	-0.26562	0.28266	0.24677	-0.84261◄	0.02073	0.04926	-0.01251	-0.04842	-0.01136
10 LALCDA	0.13419	-0.18552	-0.26444	0.23773	0.23060	-0.84805◄	0.06311	0.04179	-0.00704	-0.18943	-0.01481
11 LSTDA	0.95636	0.05863	0.04649	0.10495	0.11139	0.03351	-0.19449	-0.01737	0.02566	0.10464	0.00210
12 LSCDA	0.95819◄	0.04684	0.05233	0.10841	0.10367	0.02112	-0.19920	-0.02411	0.02121	0.07644	0.00712
13 LSLTDA	0.95405	-0.00681	-0.01135	0.13163	0.14143	-0.13083	-0.14587	-0.00110	0.01343	0.07497	-0.00287
14 LSLCDA	0.95551◄	-0.02116	-0.01494	0.12529	0.13003	-0.16443	-0.13233	-0.00466	0.01333	-0.00709	-0.00132
15 LDDTDA	-0.00096	0.02033	-0.95853	-0.00922	0.15418	-0.06512	0.01045	0.05561	0.06105	0.14143	-0.02061
16 LDDCDA	-0.06235	-0.02035	-0.95545◄	-0.06294	0.13114	-0.12312	0.04992	0.02571	0.06456	-0.12577	0.00309
17 LOFTDA	-0.02719	-0.02830	0.96390	0.01072	-0.14100	0.05701	0.00127	-0.05400	-0.07456	-0.15494	0.02003
18 LOFCDA	0.03150	0.01859	0.96078◄	0.05262	-0.10928	0.14116	-0.05149	-0.05260	-0.06253	0.17554	0.01436
19 LFTDA	0.44288	0.42456	-0.10378	0.29215	0.10554	-0.07473	-0.05300	0.13415	-0.10291	0.21027	0.01828
20 MAP	0.05603	0.70600	0.13721	-0.17064	-0.57676	0.12071	0.13486	-0.09109	-0.05128	-0.03596	0.13824
21 MAS	0.03311	0.90004	-0.22047	-0.06953	0.20619	-0.02689	-0.02153	0.02710	-0.00110	0.05863	-0.26341
22 MASP	0.02904	0.39440	-0.33077	0.10543	0.69986	-0.10730	-0.14781	0.09673	0.02769	0.09523	-0.42786◄
23 MWP	0.00415	0.97148◄	-0.03209	-0.04598	0.07080	0.08294	0.02732	0.03606	0.02543	0.09620	-0.03055
24 MSP	-0.22030	0.27605	0.07527	-0.24120	-0.78263	0.03595	0.22059	0.04619	-0.06415	-0.12898	0.07034
25 MWSP	-0.14581	0.79643	0.03067	-0.19172	-0.48095	0.07646	0.16642	0.05418	-0.07698	-0.07571	0.02774
26 A10YP	0.02002	0.72534	0.04648	-0.31398	-0.46266	0.18387	0.14205	0.01425	-0.01008	-0.04409	0.15774
27 W10YP	-0.09666	0.93738◄	-0.07295	0.07054	0.11878	-0.02671	0.07137	0.12311	0.05140	0.09250	-0.06790
28 MWT	-0.19648	0.08914	0.01679	-0.86555	-0.38465	0.13487	0.11999	-0.04034	-0.03627	-0.05130	-0.00541
29 MJANT	0.20587	-0.07371	0.02178	0.90324◄	0.30276	-0.10672	-0.08566	-0.00645	0.04111	0.05649	-0.00683
30 MST	-0.25565	-0.10279	0.12434	-0.33240	-0.83355	0.12186	0.15694	-0.02734	0.00818	0.02556	-0.14153
31 MJUNT	-0.15223	-0.02152	0.23823	-0.06532	-0.90452◄	0.14187	0.21550	-0.04458	-0.03984	0.02985	-0.06078
32 WMXT	-0.07490	0.03157	0.00333	-0.93856	-0.26254	0.11864	-0.01166	-0.00228	-0.00063	0.05054	0.00345
33 JAMXT	0.10311	-0.00112	0.01583	0.94129◄	0.24753	-0.08158	0.04950	-0.04137	0.00950	-0.01116	-0.00337
34 FE	-0.25776	-0.03845	0.18086	-0.30046	-0.85319◄	0.12034	0.21631	-0.02292	-0.01378	0.01421	-0.04306
35 THAET	0.02424	0.52822	0.12242	-0.21670	-0.70428	0.10550	0.12610	-0.07306	-0.04321	-0.05378	0.20953
36 TUAET	-0.12104	0.38626	0.12551	-0.45588	-0.73523	0.14040	0.18167	-0.06387	-0.03886	-0.02558	0.06016
37 LTHSUR	0.12641	0.90575	0.09445	0.03546	0.02345	0.10611	0.08643	-0.08729	-0.04644	0.06371	-0.00280
38 TUSUR	0.15797	0.80638	0.12614	0.02836	-0.40010	0.09185	0.08732	-0.09564	-0.05203	-0.03740	0.24109
39 MATR	-0.00341	-0.11238	0.22089	0.85273	-0.41269	-0.03407	0.12037	-0.00919	0.02529	0.04740	-0.02180

CUMULATIVE PROPORTION OF TOTAL VARIANCE

0.32622	0.50091	0.65627	0.75944	0.82683	0.87044	0.90714	0.93007	0.94611	0.96050	0.97088

◄ DEFINING VARIABLES SELECTED FOR FURTHER ANALYSIS

TABLE 4-20

ROTATED FACTOR MATRIX RESULTING FROM THE FACTOR ANALYSIS OF THE SELECTED DEFINING VARIABLES FOR THE EASTERN REGION

(N=89)

VARIABLE	FACTOR 1	2	3	4	5	6	7	8	9	10
1 LTDA	-0.91129	0.06563	-0.09759	-0.02708	-0.24692	-0.09134	0.09137	-0.00316	-0.12225	0.24308
2 LNCDA	-0.03982	-0.04798	-0.16785	0.05396	0.19226	0.12066	0.93577◁	-0.15804	-0.10435	-0.03716
3 LCDA	-0.91294◁	0.08453	-0.06924	-0.04766	-0.25426	-0.10420	-0.05051	0.02237	-0.10117	0.24750
4 LNCS	0.31265	-0.39401	0.16110	0.00065	0.05390	-0.11363	-0.16802	-0.14109	0.79460◁	0.15694
5 LBS	-0.15369	-0.21895	0.03644	-0.06814	-0.03383	-0.94769◁	-0.11621	-0.00990	0.06912	0.06858
6 LALTDA	-0.01127	-0.19789	-0.08290	-0.24055	-0.12139	-0.01049	0.02925	-0.91001	0.07052	0.19973
7 LALCDA	0.03118	-0.23344	-0.11005	-0.19441	-0.11441	-0.00577	0.16633	-0.90284◁	0.04536	0.17375
8 LSCDA	-0.26002	0.07369	0.02267	-0.13094	-0.93279	-0.02779	-0.14291	-0.01630	-0.03448	0.09860
9 LSLCDA	-0.20157	0.00195	-0.03487	-0.13591	-0.93206◁	-0.01343	-0.06927	-0.21029	-0.00485	0.12359
10 LDDCDA	0.06044	-0.94934◁	0.02081	0.05035	0.05377	-0.11864	0.02860	-0.17736	0.10668	0.15778
11 LOFCDA	-0.06447	0.94753	-0.01859	-0.04449	-0.02113	0.11972	-0.03097	0.19480	-0.12806	-0.13359
12 MACP	-0.17945	-0.26469	0.46678	-0.14916	-0.00605	-0.05631	-0.11261	-0.15683	0.11809	0.62228
13 MWP	0.08114	-0.00471	0.96858◁	0.05454	-0.05032	0.00181	-0.09081	0.15279	-0.00028	0.05758
14 WIOYP	0.07180	-0.01976	0.97403	-0.04485	0.06253	-0.03726	-0.07144	0.02126	0.09896	0.09035
15 MJANT	-0.11137	0.06356	-0.02891	-0.91920	-0.17054	-0.06242	-0.06791	-0.20591	0.01533	0.20203
16 MJUNT	0.28059	0.19583	-0.07368	0.14295	0.10356	0.06128	0.00319	0.21426	-0.07296	-0.88562◁
17 JAMXT	0.03532	0.03760	0.02851	-0.96029◁	-0.08555	-0.01873	0.00269	-0.17067	-0.01282	0.14870
18 PE	0.27847	0.13781	-0.07580	0.37232	0.21498	0.03614	0.02372	0.20463	-0.04589	-0.81233

CUMULATIVE PROPORTION OF TOTAL VARIANCE

1	2	3	4	5	6	7	8	9	10
0.29724	0.48767	0.63224	0.73602	0.81323	0.87489	0.91560	0.94813	0.96911	0.98486

◁ DEFINING VARIABLES SELECTED FOR INCLUSION IN THE MULTIPLE REGRESSION ANALYSES FOR THE EASTERN REGION

TABLE 4-21

CORRELATION MATRIX FOR DEPENDENT VARIABLES AND SELECTED INDEPENDENT VARIABLES FOR THE EASTERN REGION

(N=89)

VARIABLE	1	2	3	4	5	6	7
1 LMY	1.000	0.916	0.908	0.715	-0.324	0.718	-0.110
2 LM10YY		1.000	0.837	0.768	-0.284	0.829	-0.151
3 LMF			1.000	0.894	-0.239	0.708	-0.202
4 LM10YF				1.000	-0.157	0.744	-0.209
5 LNCDA					1.000	-0.073	-0.247
6 LCDA						1.000	-0.367
7 LMCS							1.000

VARIABLE	8	9	10	11	12	13	14
1 LMY	0.152	0.053	0.612	0.039	0.159	-0.390	0.125
2 LM10YY	0.203	0.103	0.605	0.001	0.095	-0.474	0.110
3 LMF	0.082	-0.029	0.422	0.072	0.145	-0.382	0.077
4 LM10YF	0.105	-0.025	0.295	0.071	0.023	-0.459	0.013
5 LNCDA	-0.232	0.281	-0.212	0.076	-0.283	0.038	-0.055
6 LCDA	0.246	0.006	0.460	-0.117	-0.109	-0.482	0.073
7 LMCS	0.237	0.242	-0.062	0.543	0.185	-0.229	0.036
8 LBS	1.000	0.070	0.101	0.323	0.034	-0.228	0.084
9 LALCDA		1.000	0.332	0.403	-0.249	-0.419	0.367
10 LSLCDA			1.000	-0.015	-0.031	-0.330	0.255
11 LDDCDA				1.000	0.006	-0.350	-0.031
12 MVP					1.000	-0.062	-0.036
13 MJUNT						1.000	-0.302
14 JAMXT							1.000

TABLE 4-22

SUMMARY OF STEPWISE MULTIPLE REGRESSION RESULTS FOR THE EASTERN REGION
ANALYSIS OF LMY AFTER INDEPENDENT VARIABLE SCREENING
(N=89)

Step Number (1)	Regression Equations[1] (2)	R (3)	R^2 (4)	F*** (5)	S.E. (6)	S.E.% of $\overline{LMY}$ (7)
	$\overline{LMY}$ = 4.18352				.76360	18.3
1	LMY = 1.604 +1.049 LCDA (.11)	.718	.516	92.8	.5342	12.8
2	LMY = 1.857 +0.810 LCDA +0.494 LSLCDA (.11) (.10	.785	.617	69.1	.4783	11.4
3	LMY = 0.661 +0.851 LCDA +0.485 LSLCDA +0.256 MWP (.10) (.10) (.07)	.819	.670	57.6	.4460	10.7
4	LMY = 1.110 +0.860 LCDA +0.437 LSLCDA +0.204 MWP -0.234* LNCDA (.10) (.10) (.07) (.08)	.833	.693	47.4	.4329	10.3
5	LMY = 1.198 +0.876 LCDA +0.425 LSLCDA +0.202 MWP -0.250 LNCDA +0.364* LDDCDA (.10) (.09) (.07) (.08) (.17)	.842	.709	40.4	.4243	10.1
6	LMY = 1.453 +0.944 LCDA +0.405 LSLCDA +0.280 MWP -0.307 LNCDA +0.529 LDDCDA -0.423* LBS (.10) (.09) (.07) (.09) (.18) (.19) (4-9)	.852	.726	36.2	.4142	9.9

[1] A t test has been employed to test the significance of each of the regression coefficients. All regression coefficients are significant at the 1% level except where noted by * or **. The * indicates the coefficient is significant at the 5% level only; and the ** indicates the coefficient is not significant at the 5% level.

*** All F statistics are significant at the 1% level.

TABLE 4-23

SUMMARY OF STEPWISE MULTIPLE REGRESSION RESULTS FOR THE EASTERN REGION
ANALYSIS OF LM10YY AFTER INDEPENDENT VARIABLE SCREENING
(N=89)

Step Number (1)	Regression Equations[1] (2)	R (3)	R^2 (4)	F*** (5)	S.E. (6)	S.E.% of $\overline{LM10YY}$ (7)
	$\overline{LM10YY}$ = 4.73699				.65213	13.8
1	LM10YY = 2.197 +1.033 LCDA (.07)	.829	.686	190.5	.3673	7.8
2	LM10YY = 2.368 +0.970 LCDA +0.336 LSLCDA (.08) (.07)	.866	.750	129.1	.3297	7.0
3	LM10YY = 1.574 +0.838 LCDA +0.331 LSLCDA +0.170 MWP (.07) (.07) (.05)	.885	.783	102.1	.3093	6.5
4	LM10YY = 1.903 +0.997 LCDA +0.295 LSLCDA +0.132 MWP -0.172 LNCDA (.07) (.07) (.05) (.06)	.894	.800	83.7	.2989	6.3
5	LM10YY = 1.740 +0.932 LCDA +0.220 LSLCDA +0.152 MWP -0.223 LNCDA +0.322* LALCDA (.07) (.07) (.05) (.07) (.14) (4-10)	.901	.811	71.3	.2918	6.2

[1] A t test has been employed to test the significance of each of the regression coefficients. All regression coefficients are significant at the 1% level except where noted by * or **. The * indicates the coefficient is significant at the 5% level only; and the ** indicates the coefficient is not significant at the 5% level.

*** All F statistics are significant at the 1% level.

TABLE 4-24

SUMMARY OF STEPWISE MULTIPLE REGRESSION RESULTS FOR THE EASTERN REGION
ANALYSIS OF LMF AFTER INDEPENDENT VARIABLE SCREENING
(N=89)

Step Number (1)	Regression Equations[1] (2)	R (3)	R^2 (4)	F*** (5)	S.E. (6)	S.E.% of $\overline{LMF}$ (7)
	$\overline{LMF}$ = 2.67960				.58804	21.9
1	LMF = 0.722 +0.795 LCDA (.09)	.708	.501	87.4	.4177	15.6
2	LMF = 0.160 +0.824 LCDA +0.189 MWP (.08) (.06)	.742	.551	52.8	.3986	14.9
3	LMF = 0.099 +0.845 LCDA +0.190 MWP +0.354 LDDCDA (.08) (.06) (.16)	.759	.576	38.5	.3896	14.5
4	LMF = 0.052 +0.911 LCDA +0.201 MWP +0.513*LDDCDA -0.431 LBS (.08) (.06) (.16) (.16)	.780	.608	32.6	.3767	14.1
5	LMF = 0.577 +0.937 LCDA +0.152 MWP +0.569 LDDCDA -0.553 LBS -0.234 LNCDA (.08) (.06) (.16) (.16) (.08)	.803	.645	30.2	.3606	13.5
6	LMF = 0.927 +0.850 LCDA +0.160 MWP +0.747 LDDCDA -0.522 LBS -0.277 LNCDA -0.194**LNCS (.09) (.05) (.19) (.16) (.08) (.13) (4-11)	.810	.556	26.1	.3573	13.3

---

[1] A t test has been employed to test the significance of each of the regression coefficients. All regression coefficients are significant at the 1% level except where noted by * or **. The * indicates the coefficient is significant at the 5% level only; and the ** indicates the coefficient is not significant at the 5 % level.

*** All F statistics are significant at the 1% level.

TABLE 4-25

SUMMARY OF STEPWISE MULTIPLE REGRESSION RESULTS FOR THE EASTERN REGION
ANALYSIS OF LM10YF AFTER INDEPENDENT VARIABLE SCREENING
(N=89)

Step Number (1)	Regression Equations[1] (2)	R (3)	R^2 (4)	F*** (5)	S.E. (6)	S.E.% of $\overline{LM10YF}$ (7)
	$\overline{LM10YF}$ = 3.23827				.54429	16.8
1	LM10YF = 1.335 +0.774 LCDA (.07)	.744	.553	107.7	.3660	11.3
2	LM10YF = 1.396 +0.794 LCDA +0.331* LDDCDA (.07) (.15)	.761	.579	59.0	.3575	11.0
3	LM10YF = 1.550 +0.843 LCDA +0.452 LDDCDA -0.330* LBS (.08) (.15) (.15)	.775	.601	42.6	.3500	10.8
4	LM10YF = 1.773 +0.847 LCDA +0.511 LDDCDA -0.423 LBS -0.172* LNCDA (.07) (.15) (.15) (.07)	.791	.626	35.2	.3406	10.5
5	LM10YF = 1.750 +0.906 LCDA +0.533 LDDCDA -0.450 LBS -0.197 LNCDA -0.118** LSLCDA (.08) (.15) (.15) (.07) (.08) (4-12)	.798	.637	29.1	.3377	10.4

---

[1] A t test has been employed to test the significance of each of the regression coefficients. All regression coefficients are significant at the 1% level except where noted by * or **. The * indicates the coefficient is significant at the 5% level only; and the ** indicates the coefficient is not significant at the 5% level.

*** All F statistics are significant at the 1% level.

are contained in Table 4-25. The final equation (4-13) contains five independent variables and has an R^2 value of 0.637. The regression coefficients for the first four terms are significant at the 1% level and conform to physical theory. The regression coefficient of the fifth term, LSLCDA, is not significant at the 5% level. The final regression equation is significant at the 1% level according to an analysis of variance.

4.5.5 Analysis for the Western Region

The methods employed in the analyses for the western region are the same as those utilized for the eastern region (Section 4.5.4) and for the full study area (Section 4.4.3). The factor analysis of the full set of 39 independent variables for each of the 57 study basins in the western region resulted in the rotated factor matrix, Table 4-26. A second factor analysis on the set of 17 defining variables resulted in the rotated factor matrix, Table 4-27. A final set of 8 independent variables was selected as the result of the screening of the independent variables. These measures were subsequently employed in a stepwise multiple regression analysis for each of the four dependent variables. The results of these analyses are summarized in Tables 4-29 to 4-32 inclusive.

Table 4-28 is the initial correlation matrix for the regression analyses for the western region. The results of the western region regression analysis for the dependent variable LMY are displayed in Table 4-29. The final equation has four independent variables which combine to account for 85% of the total variance in the dependent variable. The first three regression coefficients are significant at the 1% level while the fourth is significant at the 5% level; and the signs of the regression coefficients correspond to expectations based on physical theory. The F test of an analysis of variance indicated that the regression is significant at the 1% level. The regression results for the western region analysis of the dependent variable LM10YY are presented in Table 4-30. The final equation in this table contains three independent variables which combine to explain 80% of the variance in the de-pendent variable. All of the regression coefficients are significant at the 1% level and their signs correspond to physical theory. The F statistic confirms the significance of the regression at the 1% level. The regression results for the western region analysis of the dependent variable LMF are presented in Table 4-31. The final equation contains four independent variables, which combine to explain 77% of the total variance in the dependent variable. The first three regression coefficients are significant at the 1% level while the fourth, that associated with the LALCDA term is not significant at the 5% level. The signs of all the coefficients correspond to those expected on the basis of physical theory; and, the regression is significant at the 1% level for the F test and the standard error of the estimate is 10.3% of the mean. The results of the western region regression analysis for the dependent variable LM10YF are summarized in Table 4-32. The final equation (5-17) contains three independent variables which explain only 53% of the total variance in the dependent variable. The LDDCDA and MST terms have regression coefficients which are not significant at the 5% level. This regression analysis for the LM10YF

TABLE 4-26

ROTATED FACTOR MATRIX RESULTING FROM THE FACTOR ANALYSIS OF ALL INDEPENDENT VARIABLES FOR THE WESTERN REGION
(N=57)

VARIABLE	FACTOR 1	2	3	4	5	6	7	8	9	10
1 LTDA	-0.03416	-0.26533	0.29606	-0.87019◄	0.16275	0.02444	-0.21047	-0.04354	0.02506	0.07064
2 LNCDA	0.24438	0.18780	0.50417	-0.10774	-0.00193	-0.05626	-0.42096	-0.00998	-0.10336	0.65158◄
3 LCDA	-0.07625	-0.28221	0.19243	-0.90044◄	0.19239	0.02825	-0.13048	-0.05959	-0.01413	-0.04691
4 LMCS	-0.16725	0.06917	-0.40222	0.01703	-0.07572	0.17074	0.59926	0.16492	0.29207	-0.10197
5 LBCL	-0.36506	0.01577	-0.32956	0.12633	-0.01933	0.25950	0.76212	0.02766	0.11435	-0.14711
6 LPR	-0.40171	-0.13987	-0.19678	-0.44483	-0.03464	0.13761	0.49309	-0.05935	0.49326◄	-0.17915
7 LBL	-0.04055	-0.30452	0.20968	-0.91326	0.16639	0.02651	-0.18057	-0.01177	0.51996	0.03361
8 LPS	-0.10314	-0.26104	-0.13363	-0.91475	0.09702	0.01906	-0.00603	-0.90267◄	-0.00117	-0.00220
9 LNLTDA	0.08759	-0.26493	0.12563	-0.19991	0.91591◄	0.02304	-0.11160	-0.01944	-0.01766	-0.03453
10 LNLCDA	0.11239	-0.24903	0.16972	-0.21162	0.91243◄	0.01694	-0.16557	-0.06775	-0.01224	0.03876
11 LSTDA	-0.28606	-0.91999	0.03392	-0.18131	0.05623	0.06163	-0.13194	-0.07181	-0.00880	-0.04491
12 LSCDA	-0.25971	-0.92763◄	0.07107	-0.13476	0.05759	0.07075	-0.13951	-0.01797	-0.01272	-0.01530
13 LSLTDA	-0.22275	[illegible]	0.02663	-0.20335	0.25971	0.05598	-0.10512	-0.01068	0.01219	-0.04200
14 LSLCDA	-0.19181	-0.90725◄	0.07999	-0.20961	0.24523	0.07147	-0.17151	-0.00743	0.01614	0.00915
15 LDDTDA	-0.03554	0.01130	[illegible]	0.00093	-0.04107	-0.00608	0.33608	-0.04192	-0.06599	[illegible]
16 LDDCDA	0.11102	0.08710	-0.91993◄	0.19539	-0.15771	-0.00285	0.13914	-0.00119	0.00446	0.14974
17 LOFTDA	0.06046	-0.02399	0.92307	-0.06798	0.00793	-0.01111	-0.32459	0.02400	0.05493	0.23193
18 LOFCDA	-0.09045	-0.10257	0.92917◄	-0.20953	0.12745	-0.00294	-0.13513	0.02392	-0.03090	-0.09508
19 LFTDA	-0.63580	-0.53724	0.02152	-0.17451	0.07047	0.16788	0.10528	-0.03439	0.15317	-0.04476
20 MAP	-0.93851◄	-0.27991	0.03438	-0.05363	-0.02442	0.03282	0.11497	-0.02799	0.00712	-0.04796
21 MAS	-0.53720	-0.02510	-0.00026	0.01106	-0.00764	0.12035	0.30936	-0.01222	0.01503	0.00411
22 MASP	-0.54055	0.21770	-0.04663	0.06033	0.04407	0.13476	0.61731	-0.03875	0.06128	0.05754
23 MMP	-0.91373	-0.14276	-0.05521	-0.05490	0.03629	-0.00192	0.24701	-0.02938	0.17111	-0.01275
24 MSP	-0.91577	0.00544	0.00646	0.02597	-0.09201	0.09414	0.30216	-0.01563	-0.07099	-0.07143
25 MWSP	-0.96748	-0.05958	-0.01304	-0.01124	-0.05679	0.05544	0.28974	-0.00792	0.00142	-0.04999
26 AIDYP	-0.95768◄	-0.19436	0.01546	-0.01289	-0.02695	0.09934	0.15109	-0.02453	0.01033	[illegible]
27 WIDYP	-0.96517	-0.05075	-0.07053	-0.05755	-0.02225	-0.03142	0.29952	[illegible]	0.12102	-0.00279
28 MWT	-0.41122	0.16692	-0.23052	0.15097	-0.15151	-0.10927	0.52143◄	-0.02134	-0.00117	0.00753
29 MJANT	0.44053	-0.20723	0.24978	-0.13711	0.10419	-0.01327	-0.81212	0.00132	-0.02934	0.01630
30 MST	0.62719	0.24219	0.01980	0.04666	-0.02616	-0.53229◄	[illegible]	0.01626	-0.04433	0.02372
31 MJULT	0.72440	0.28401	0.02024	0.04579	-0.02174	[illegible]	-0.71277	[illegible]	-0.03316	0.00217
32 WIXT	-0.37598	0.12814	-0.26701	0.12570	-0.11755	-0.01659	0.95720◄	0.09959	-0.04975	-0.05710
33 JANXT	0.39149	-0.20130	0.26310	-0.11017	0.06491	-0.26955	[illegible]	-0.02136	0.00912	0.06138
34 PE	0.65617	0.35712	0.00012	0.09957	-0.02949	-0.53244◄	-0.11639	0.02248	-0.04597	0.04107
35 THRET	-0.80362	[illegible]	0.04740	-0.10353	-0.01065	0.36912	0.03529	-0.05246	0.00969	-0.03913
36 TURET	-0.91924	-0.07674	-0.00246	-0.01834	-0.03927	-0.13658	0.23734	-0.05367	0.03040	0.03444
37 LTHSUR	-0.85094	-0.02997	-0.02876	0.07143	-0.05536	0.09002	0.33513	0.01434	-0.01618	-0.04607
38 TUSUR	-0.92631	-0.34173	0.04869	-0.07043	-0.00312	0.15927	0.05767	-0.01591	-0.03482	-0.07530
39 LATR	0.68512	0.09938	0.17027	-0.04026	0.00920	-0.25213	-0.64488	0.01931	-0.03790	0.01920

CUMULATIVE PROPORTION OF TOTAL VARIANCE

1	2	3	4	5	6	7	8	9	10
0.44955	0.70128	0.79232	0.84357	0.88149	0.90836	0.93017	0.94765	0.96071	0.97131

◄ DEFINING VARIABLES SELECTED FOR FURTHER ANALYSIS

TABLE 4-27

ROTATED FACTOR MATRIX RESULTING FROM THE FACTOR ANALYSIS OF THE SELECTED DEFINING VARIABLES FOR THE WESTERN REGION

(N=57)

VARIABLE	FACTOR 1	2	3	4	5	6	7	8	9
1 LTDA	-0.20751	0.07233	0.27262	0.86403	-0.21172	0.13983	0.12880	-0.20161	0.00560
2 LNCDA	0.10704	-0.23893	0.31116	0.10843	-0.01858	-0.02400	0.82691◁	-0.35344	-0.02589
3 LCDA	-0.20941	0.10635	0.23007	0.88780◁	-0.23950	0.15907	-0.00719	-0.11563	0.01138
4 LBR	-0.06233	0.40723	-0.11957	0.37691	0.07337	0.12717	-0.27740	0.43298	0.01248
5 LBS	-0.19500	0.10925	-0.08285	0.21648	-0.12416	0.93716◁	-0.01978	0.01463	0.01580
6 LALTDA	-0.16233	0.00620	0.13653	0.17020	-0.94989	0.05755	-0.02758	-0.12482	-0.01355
7 LALCDA	-0.15716	-0.01636	0.15371	0.18282	-0.93402◁	0.08529	0.05838	-0.17600	-0.01581
8 LSCDA	-0.87500	0.34035	0.09461	0.21138	-0.12874	0.14038	-0.06584	-0.12782	0.04653
9 LSLCDA	-0.84088◁	0.29366	0.08320	0.22383	-0.31709	0.14752	-0.04035	-0.13241	0.02638
10 LDDCDA	0.06001	-0.03248	-0.94239◁	-0.18605	0.15705	0.05550	-0.09145	0.14623	-0.02841
11 LOFCDA	-0.08725	0.06591	0.93851	0.19784	-0.13391	-0.03143	0.14367	-0.14398	0.00886
12 MAP	-0.24643	0.79354	0.12177	0.07399	0.07022	0.06567	-0.07899	0.25889	0.45069
13 A10YP	-0.16467	0.80511	0.09991	0.02145	0.08076	0.05572	-0.06434	0.29743	0.45585
14 MWT	0.13867	0.22477	-0.15634	-0.17007	0.19894	0.01148	-0.11766	0.89251◁	0.08953
15 MST	0.12424	-0.95934◁	-0.01756	-0.05457	0.00979	-0.04172	0.07563	-0.15976	0.13518
16 WMXT	0.12312	0.26558	-0.18991	-0.14544	0.15814	-0.00973	-0.18752	0.88436	-0.00231
17 PE	0.22784	-0.94502	-0.02725	-0.09264	0.08522	-0.05607	0.09378	-0.06362	0.11918

CUMULATIVE PROPORTION OF TOTAL VARIANCE

1	2	3	4	5	6	7	8	9
0.32916	0.62870	0.73394	0.80592	0.87118	0.90928	0.94486	0.96804	0.98591

◁ DEFINING VARIABLES SELECTED FOR INCLUSION IN THE MULTIPLE REGRESSION ANALYSES FOR THE WESTERN REGION

TABLE 4-28

CORRELATION MATRIX FOR DEPENDENT VARIABLES AND SELECTED INDEPENDENT VARIABLES FOR THE WESTERN REGION

(N=57)

VARIABLE	1	2	3	4	5	6
1 LMY	1.000	0.933	0.941	0.675	-0.311	0.755
2 LM10YF		1.000	0.865	0.771	-0.246	0.792
3 LMF			1.000	0.749	-0.308	0.732
4 LM10YF				1.000	-0.057	0.693
5 LNCDA					1.000	0.149
6 LCDA						1.000

VARIABLE	7	8	9	10	11	12
1 LMY	0.327	0.201	0.564	-0.240	0.050	-0.552
2 LM10YF	0.387	0.248	0.566	-0.258	-0.056	-0.458
3 LMF	0.304	0.184	0.527	-0.196	0.025	-0.466
4 LM10YF	0.308	0.229	0.403	-0.146	-0.167	-0.235
5 LNCDA	-0.096	0.183	-0.099	-0.423	-0.535	0.357
6 LCDA	0.406	0.484	0.542	-0.448	-0.333	-0.175
7 LBS	1.000	0.247	0.416	0.054	-0.020	-0.185
8 LALCDA		1.000	0.511	-0.360	-0.433	0.003
9 LSLCDA			1.000	-0.251	-0.271	-0.391
10 LDDCDA				1.000	0.342	0.076
11 MWT					1.000	-0.333
12 MST						1.000

TABLE 4-29

SUMMARY OF STEPWISE MULTIPLE REGRESSION RESULTS FOR THE WESTERN REGION
ANALYSIS OF LMY AFTER INDEPENDENT VARIABLE SCREENING
(N=57)

Step Number (1)	Regression Equations[1] (2)	R (3)	R^2 (4)	F*** (5)	S.E. (6)	S.E.% of $\overline{\text{LMY}}$ (7)
	$\overline{\text{LMY}}$ = 4.08734				1.04009	25.4
1	LMY = 0.795 +1.397 LCDA (.15)	.755	.570	73.1	.6877	16.8
2	LMY = 1.125 +1.419 LCDA -0.822 LNCDA (.12) (.13)	.869	.755	83.0	.5248	12.8
3	LMY = 9.258 +1.234 LCDA -0.592 LNCDA -0.165 MST (.12) (.12) (.03)	.913	.833	88.2	.4367	10.7
4	LMY = 8.902 +1.407 LCDA -0.569 LNCDA -0.161 MST -0.532* LALCDA (.11) (.11) (.03) (.27)	.920	.847	72.0	.4220	10.3

(4-13)

---

[1] A t test has been employed to test the significance of each of the regression coefficients. All regression coefficients are significant at the 1% level except where noted by * or **. The * indicates the coefficient is significant at the 5% level only; and the ** indicates the coefficient is not significant at the 5 % level.

*** All F statistics are significant at the 1% level.

TABLE 4-30

SUMMARY OF STEPWISE MULTIPLE REGRESSION RESULTS FOR THE WESTERN REGION
ANALYSIS OF LM10YY AFTER INDEPENDENT VARIABLE SCREENING
(N=57)

Step Number (1)	Regression Equations[1] (2)	R (3)	R^2 (4)	F*** (5)	S.E. (6)	S.E. % of $\overline{\text{LM10YY}}$ (7)
	$\overline{\text{LM10YY}}$ = 4.74375				.72058	15.2
1	LM10YY = 2.352 +0.950 LCDA (.10)	.792	.627	92.6	.4439	9.4
2	LM10YY = 2.548 +1.016 LCDA -0.490 LNCDA (.08) (.09)	.874	.763	86.9	.3572	7.5
3	LM10YY = 6.469 +0.956 LCDA -0.379 LNCDA -0.079 MST (.08) (.05) (.02)	.895	.801	71.2	.3303	7.0

(4-14)

---

[1] A t test has been employed to test the significance of each of the regression coefficients. All regression coefficients are significant at the 1% level except where noted by * or **. The * indicates the coefficient is significant at the 5% level only; and the * indicates the coefficient is not significant at the 5% level.

*** All F statistics are significant at the 1% level.

TABLE 4-31

SUMMARY OF STEPWISE MULTIPLE REGRESSION RESULTS FOR THE WESTERN REGION
ANALYSIS OF LMF AFTER INDEPENDENT VARIABLE SCREENING
(N=57)

Step Number (1)	Regression Equations[1] (2)	R (3)	R^2 (4)	F*** (5)	S.E. (6)	S.E.% of $\overline{\text{LMF}}$ (7)
	$\overline{\text{LMF}}$ = 2.63495				.68975	26.2
1	LMF = 0.519 +0.840 LCDA (.11)	.732	.536	63.4	.4743	18.0
2	LMF = -0.734 +0.913 LCDA -0.536 LNCDA (.09) (.09)	.845	.713	67.2	.3761	14.3
3	LMF = 4.440 +0.856 LCDA -0.431 LNCDA -0.075 MST (.08) (.09) (.03)	.866	.750	53.1	.3543	13.4
4	LMF = 4.179 +0.939 LCDA -0.414 LNCDA -0.072 MST -0.434** LALCDA (.09) (.09) (.03) (.22)	.876	.767	42.9	.3452	13.1

(4-15)

---

[1] A t test has been employed to test the significance of each of the regression coefficients. All regression coefficients are significant at the 1% level except where noted by * or **. The * indicates the coefficient is significant at the 5% level only; and the ** indicates the coefficient is not significant at the 5% level.

*** All F statistics are significant at the 1% level.

TABLE 4-32

SUMMARY OF STEPWISE MULTIPLE REGRESSION RESULTS FOR THE WESTERN REGION
ANALYSIS OF LMIOYF AFTER INDEPENDENT VARIABLE SCREENING
(N=57)

Step Number (1)	Regression Equations[1] (2)	R (3)	R^2 (4)	F*** (5)	S.E. (6)	S.E.% of $\overline{\text{LMIOYF}}$ (7)
	$\overline{\text{LMIOYF}}$ = 3.37760				.50997	15.1
1	LMIOYF = 1.897 +0.553 LCDA (.08)	.693	.480	50.7	.3711	11.0
2	LMIOYF = 1.781 +0.665 LCDA +0.673** LDDCDA (.09) (.35)	.717	.514	28.5	.3621	10.7
3	LMIOYF = 3.302 +0.649 LCDA +0.672** LDDCDA -0.030** MST (.09) (.35) (.03)	.726	.527	19.7	.3605	10.7

(4-16)

---

[1] A t test has been employed to test the significance of each of the regression coefficients. All regression coefficients are significant at the 1% level except where noted by * or **. The * indicates the coefficient is significant at the 5% level only; and the ** indicates the coefficient is not significant at the 5% level.

*** All F statistics are significant at the 1% level.

indicates that it is not possible to develop a meaningful equation for this variable in the western region on the basis of the available independent variables.

4.6 COMPARISON OF REGRESSION RESULTS FOR THE FULL STUDY AREA AND BY REGION

The results of the multiple regression analyses after variable screening, for the full study area and for the eastern and western regions, have been summarized in Table 4-33. On the basis of the data summarized in this table, it must be concluded that the twofold subdivision of the study area by multivariate grouping has not resulted either in substantially increased values of R^2, or in a reduction in the standard errors of the estimates. From these observations, it is concluded that the regional grouping at the present scale has not been successful in improving the predictive value of the regression relationships. The relatively large errors must therefore be attributed to factors which are operative on a more local scale than the twofold regional division which has been employed in the present analyses. In an effort to further account for some of these local variations, it was proposed to attempt to further subdivide the study area.

4.7 FURTHER ATTEMPTS TO SUBDIVIDE THE STUDY AREA

Three further attempts were made to establish subdivisions of the full study area such that the predictive power of the regression relationships would be improved. The first method employed utilized the multivariate grouping technique as described in Section 4.5.3 in an attempt to further subdivide both the eastern and western regions. In each case two further groups were identified; however, when mapped, these groups did not have spatial continuity and it was not possible to establish meaningful subdivisions for either the eastern or western regions.

A second attempt to improve the regression relationships for each region involved the subjective modification of the regional boundaries. The ten study basins in Northern Alberta which were grouped in the eastern region were deleted from that region. In the western region, seven study basins in southwestern Alberta which drained areas of foothills and mountains were removed from the grouping. These modified groups were then subjected to a full analysis by factor analytic and multiple regression techniques. The resulting relationships were not appreciably improved over those of the previous analyses.

A third attempt at the establishment of more meaningful regional subdivisions involved a complete regrouping of the full set of study basins. This regrouping was based on a re-examination of the original multivariate grouping analysis, as described in Section 4.5.3. A further consideration of the original factor matrix, Table 4-17, led to a proposal to repeat the grouping analysis on the basis of the first two and fourth factors, rather than the first three factors as had been previously employed. The fourth factor represented a measure of hydrologic variability, and it was expected, might lead to a better grouping for analysis purposes. The application of the optimal grouping algorithm resulted in the definition of two groups which when plotted grouped spatially, but

TABLE 4-33 - COMPARATIVE SUMMARY OF MULTIPLE REGRESSION RESULTS AFTER VARIABLE SCREENING FOR THE FULL STUDY AREA AND FOR THE EASTERN AND WESTERN REGIONS

Dependent Variable	Region	Equat No.	No. of Steps	R^2	S.E. as % of Mean Log Units
LMY	Full study area	5-6	5	0.782	10.1
	Eastern Region	5-10	6	0.726	9.9
	Western Region	5-14	4	0.847	10.3
LM10YY	Full study area	5-7	5	0.807	6.4
	Eastern Region	5-11	5	0.811	6.2
	Western Region	5-15	3	0.801	7.0
LMF	Full study area	5-8	6	0.724	12.7
	Eastern Region	5-12	6	0.656	13.1
	Western Region	5-16	4	0.767	13.1
LM10YF	Full study area	5-9	4	0.587	10.6
	Eastern Region	5-13	5	0.637	10.4
	Western Region	5-17	3	0.527	10.7

were not very different from the original eastern and western regions. The groups resulting from this analysis were not employed in further regression analyses.

All attempts to further subdivide the study area into meaningful hydrologic regions were unsuccessful. This difficulty in regional delimitation led to a general conclusion with regard to variations in prairie hydrologic patterns. On the basis of the data employed in the present study, it would seem that local variations are dominant in the definition of prairie hydrologic patterns. It was not possible to improve the relationships on the basis of regional subdivision. Seemingly, the alternatives to this approach must involve a more detailed examination of local patterns.

4.8 TEST OF REGRESSION RELATIONSHIPS FOR THE COMPARATIVE TEST SAMPLE

Earlier in this chapter (Section 4.3.3), the selection of a random sample of 15 study basins was discussed. These basins were not included in the analyses, and therefore represent an independent sample for the testing of the regression results. Such an independent sample for testing the regression results was considered useful, in that the significance tests which are normally employed with regression analysis are based on the assumptions of random sampling and normally distributed variates. Although logarithmic transformations were applied to several of the variables (Section 4.3.2), it was not expected that the normality assumption would be totally satisfied. It was also recognized that the random sampling assumption had not been met by the input data (Section 4.3.1). Therefore as a check on the stability of the regression relationships, it was proposed to test the performance of these relationships on a sample of data for which the dependent variables had been previously measured.

Three regression relationships for each of the dependent variables were considered. These relationships resulted from the full study area analyses by multiple regression after variable screening, the western region analyses and the eastern region analyses. For each of the dependent variables the regression equations were employed to estimate the value of the dependent variables for basins in the sample. In the case of the relationships from the full study area analyses, all of the 15 study basins in the sample were considered, while for the eastern and western regions, only those basins lying within each area as delimited on Figure 4-3 were considered. The residuals for each application of the equations have been expressed as percentages of the observed values. These data are reported in Table 4-34 and provide a basis for a further consideration of the predictive value of the relationships.

The percentage residuals from the test sample applications of the regression equations are listed in Table 4-34. Each column of the table contains the results for a particular equation. The figures which have been entered in the table for each of the test basins for each equation are the percentage residuals estimated in arithmetic units and the percentage residuals estimated in logarithmic units. The latter group of figures have been enclosed in parentheses in the table. The last four rows in the table contain the means and standard deviations of the percentage residuals for each of the equations.

TABLE 4-34

PERCENTAGE RESIDUALS RESULTING FROM APPLICATION OF REGRESSION EQUATIONS TO COMPARATIVE TEST SAMPLE OF BASINS*

Basin No.	Dependent Variable LMY			Dependent Variable LMTOYY			Dependent Variable LMF			Dependent Variable LMTOYF		
	Equat 4-5 Full Study Area (10.1)**	Equat 4-9 E. Region (9.9)**	Equat 4-13 W.Region (10.3)**	Equat 4-6 Full Study Area (6.4)**	Equat 4-10 E.Region (6.2)**	Equat 4-14 W.Region (7.0)**	Equat 4-7 Full Study Area (12.7)**	Equat 4-11 E.Region (13.1)**	Equat 4-15 W.Region (13.1)**	Equat 4-8 Full Study Area (10.6)**	Equat 4-12 E.Region (10.4)**	Equat 4-16 W.Region (10.7)**
4	-50.1 (-8.2)		-24.4 (-3.3)	+ 3.2 (+0.3)		+36.8 (+3.2)	-57.6 (-14.7)		-24.5 (-4.8)	+15.9 (+2.0)		-16.2 (-2.4)
7	- 0.5 (-0.0)		-10.9 (-1.0)	-18.7 (-1.7)		-23.9 (-2.2)	+27.8 (+3.2)		+ 3.2 (+0.4)	-33.9 (-4.8)		-37.3 (-5.5)
15	+57.5 (+4.6)		+162.8 (+9.7)	+70.1 (+4.8)		+109.3 (-6.6)	+ 4.2 (-0.6)		+41.7 (+5.2)	-27.7 (-3.7)		-30.2 (-4.1)
25	-37.6 (-5.1)		-29.1 (-3.7)	+16.9 (+1.7)		+39.8 (+3.2)	-15.4 (-2.7)		-33.3 (-6.6)	-47.4 (-7.3)		-35.3 (-4.9)
30	-46.4 (-7.8)		-62.1 (-12.2)	+91.9 (+7.7)		+74.9 (+6.6)	-84.3 (-33.3)		-71.5 (-22.5)	-52.5 (-10.9)		- 8.2 (-1.3)
37	-12.4 (-1.3)	-52.9 (-7.5)		+57.6 (+4.1)	+34.2 (+2.6)		+ 8.3 (+1.3)	+39.6 (+5.5)		+61.5 (+6.3)	+116.4 (+10.1)	
44	-38.4 (-5.5)	-19.8 (-2.5)		+88.8 (+6.8)	+70.7 (+5.7)		+34.2 (+5.6)	+49.5 (+7.6)		+195.2 (+17.4)	+229.9 (+19.2)	
52	+ 7.2 (+0.5)	-11.3 (-0.9)		+ 5.5 (+0.4)	+55.3 (+3.2)		+99.1 (+8.7)	+78.4 (+7.3)		+75.9 (+6.3)	+93.0 (+7.7)	
78	-17.2 (-2.5)	+31.3 (+3.6)		-27.6 (-3.4)	-15.8 (-1.8)		+133.3 (+21.9)	+180.0 (+26.6)		+88.0 (+10.7)	+74.0 (+9.6)	
82	-47.1 (5.8)	-41.8 (-4.9)		-27.8 (-2.8)	+17.9 (+1.4)		+37.7 (+5.0)	+43.8 (+5.7)		+93.7 (+9.0)	+80.8 (+8.0)	
108	+269.8 (+20.2)	+207.5 (+17.3)		+30.3 (+2.9)	- 5.9 (-0.6)		+185.1 (+27.5)	+184.6 (+27.4)		+83.1 (+10.3)	+78.5 (+9.8)	
118	-21.6 (-2.5)	-17.3 (-1.9)		+ 3.3 (+0.3)	- 3.3 (-0.3)		-15.7 (-2.5)	-17.5 (-2.8)		-41.6 (-6.8)	-31.3 (-4.8)	
133	+ 8.9 (+0.7)		+ 9.4 (+0.7)	- 0.6 (+0.0)		-19.3 (-1.6)	-39.2 (-5.9)		-23.8 (-3.2)	-48.0 (-7.0)		-30.7 (-3.9)
139	-60.1 (-7.7)	-53.2 (-6.4)		-37.6 (-3.8)	-42.2 (-4.4)		-43.6 (-7.4)	-32.3 (-5.0)		-22.4 (-3.0)	-18.8 (-4.4)	
159	-51.3 (-7.8)		- 5.1 (-0.6)	+32.2 (+2.8)		+95.1 (+6.6)	-29.8 (-5.7)		- 5.6 (-0.9)	+63.0 (+6.5)		+19.7 (+2.4)
Mean % Arithmetic	-2.6	+5.4	+5.8	+19.2	+13.9	+44.7	+16.3	+65.8	-16.3	+26.9	+78.6	-19.7
S.D.	81.5	86.2	72.8	41.9	38.0	52.5	74.1	80.5	34.9	73.6	81.4	20.3
Mean % Log	-1.9	-0.4	-1.5	+1.3	+0.7	+1.3	+0.1	+9.0	-4.6	+1.7	+6.9	-2.8
S.D.	7.2	8.0	6.5	3.5	3.2	4.9	14.2	12.0	8.8	8.5	8.0	2.7

*Figures in the body of the table are the percentage differences between the observed and predicted magnitudes of the dependent variables. These differences have been expressed as percentages of the observed magnitudes. The percentages in parentheses are based on logarithmic units, while those without parentheses are based on arithmetic units.

** Standard Error of Estimate of the regression equation expressed as a percentage of the mean value of the dependent variables.

On the basis of the results summarized in Table 4-34, it is possible to make several observations and to draw conclusions regarding the validity of the regression equations developed in the present research. On comparison of the percentage residuals for the test sample based on logarithmic units with the standard errors of the estimates for the regression equations, it may be concluded that the regression relationships are stable. That is, the errors which resulted from the application of the equations to the independent test sample of basins were similar in magnitude to those expected on the basis of the standard errors of estimate which resulted from the analyses for the full set of 146 study basins. This stability of the relationships lends credibility to the statistical significance of the regression equations.

The percentage residuals based on arithmetic units have been included in the table to illustrate the magnitudes of residual errors which are involved in the prediction of the actual magnitudes of streamflow events. The skewed nature of the distribution of these errors is evident in the larger magnitudes associated with the positive residuals than with the negative residuals. While the percentage errors are somewhat larger than might be considered desirable for prediction purposes, the regression relationships developed in this research are useful models for the estimation of streamflow characteristics on a regional basis. The relationships are stable, conform to physical theory, and are based on readily available data.

The earlier conclusion, that the division of the study area into hydrologic regions did not result in a significant improvement in the predictive stength of the regression relationships, is confirmed on the basis of the results summarized in Table 4-34.

CHAPTER V

SUMMARY, CONCLUSIONS AND SUGGESTIONS FOR FURTHER RESEARCH

5.1 INTRODUCTION

In this final chapter, the methodology and findings of the present research are summarized, conclusions are drawn, and several suggestions for future research are introduced.

5.2 SUMMARY OF THE PRESENT RESEARCH

In recent years, the needs for planning and management of water resources have grown rapidly. Many more streamflow data are required; and it is to this need that the present study has been directed. The ultimate aim of this research has been to develop predictive relationships for the estimation of streamflow characteristics for ungauged basins in the plains area of the Canadian Prairie Provinces. A second aim of the study has been to add to our understanding of plains' hydrologic patterns through the identification of climatic and other physical geographic variables which are related to streamflow chatacteristics.

The multivariate nature of the relationships involved, led to the adoption of a systems approach to the present study. The multivariate statistical techniques of multiple correlation and regression analysis and factor analysis have been employed in the regional analyses of the hypothetical model which is of the form:

$$\text{STREAMFLOW CHARACTERISTICS} = (\text{CLIMATIC PATTERNS AND OTHER PHYSICAL GEOGRAPHIC PATTERNS}) \qquad (5\text{-}1)$$

A group of 161 study basins was selected for analysis. Each of these basins met defined criteria with regard to basin location, size of drainage area, available streamflow data, and natural flow conditions.

It was beyond the scope of the present research to analyse all possible streamflow characteristics; rather four characteristics, the mean annual yield, the mean 10 year yield, the mean annual flood, and the mean 10 year flood were selected for analysis. These characteristics were chosen because of their potential usefulness in the planning process. The actual estimation of the dependent variables for each of the study basins was based upon frequency analyses of the available annual yield and annual flood flow data series. The frequency analyses of the annual yield and annual flood flow data series for each of the study basins involved the fitting of a least squares regression line to the data series as plotted on lognormal probability paper. A group of 39 independent variables, various measures of climatic and other physical geographic patterns, was compiled for each of the study basins. Each variable was selected on the basis of two considerations. The first consideration related to its theoretical relationships to the dependent variables, and the second related to the available data sources and the problems of data compilation. The first group of independent variables included 20 measures of climatic patterns. Each of these measures was related to one of the three most important climate controlled processes in the hydrologic cycle, precipitation, snowmelt and evapotranspiration. For the purposes of the present study, the basic climatic data set was comprised of published climatic normals of temperature and precipitation based on the 30 year period 1930-1960. The second group of independent variables included 19 measures of other physical geographic patterns. Each of these measures was related to one of three hydrologically significant groups of variables which were classified as measures of drainage area, measures of basin topography, and measures of surficial geology, soils, vegetation and landuse. All of the chosen variables were measured from 1:250,000 scale topographic maps of the study basins. The full set of four dependent variables and 39 independent variables was compiled for each of the 161 study basins. These data were subsequently employed in the statistical analyses of the hypothetical model.

The statistical analyses of the hypothetical model involved two stages, the first entailing an examination of the model for the entire study area, and the second involving its examination on the scale of hydrologic regions. In the first stage of the analyses two approaches were utilized for the estimation of statistical models. The first method involved the use of stepwise multiple regression techniques in the analyses of the relationships of each of the dependent variables to the full set of 39 independent variables. The second method utilized the technique of factor analysis to screen the independent variable set for multicollinearity. This screening served as a basis for the selection of a set of independent variables for inclusion in the stepwise multiple regression analysis for each of the dependent variables. The second approach to the full study area analysis resulted in a more satisfactory set of regression models. The resulting equations were statistically significant, and the signs of the regression coefficients conformed to physical theory. Although a very slight loss in explanatory power was observed relative to the first approach, the gain in physical significance resulted in this second group of relationships being judged to be superior in the context of the present study.

All of the relationships resulting from the full study area stage of the analyses had relatively large standard errors of estimation. The magnitudes of these errors, particularly in view of the fact that they were measured in logarithmic units, limited the usefulness of the relationships for predictive purposes. In an attempt to improve the predictive power of the relationships, the second stage of the analyses was undertaken. This stage entailed the subdivision of the study area into regions of hydrologic similarity in order to account for some of the regional variations in streamflow patterns and to improve on the predictive strength of the equations resulting from the analyses. Attempts to subdivide the study area, on the basis of plots of regression residuals and on the basis of plots of hydrologic index-event ratios, were not successful as local variations dominated and obscured any possible regional divisions. The use of a multivariate optimal grouping technique based on factor scores resulting from a factor analysis of the climatic variables and the hydrologic index-ratios, provided a twofold grouping of the study basins. The grouping resulted in a well-defined spatial division of the study area. However, the regression equations which resulted from analyses on a regional basis while conforming to physical theory did not result in any appreciable improvement in the predictive potential of the relationships. Further attempts at subdividing the study area to improve the regression relationships were also unsuccessful.

The predictive performance of the regression equations was tested by their application to data for the test sample of 15 basins. On the basis of these calculations, it was concluded that the regression relationships were stable and applied equally as well to the test sample as to the original data.

5.3 CONCLUSIONS

The twofold aims of the present research were firstly, to develop predictive relationships for the estimation of streamflow characteristics in ungauged areas of the prairie, and secondly, to add to our overall knowledge of prairie hydrologic patterns through the identification of climatic and other physical geographic patterns which are closely related to streamflow characteristics. With respect to the first of these aims, a limited degree of success has been achieved in the present study. The stability of the relationships has been demonstrated by their application to the independent test sample of 15 basins. However, the standard errors of the estimates associated with the regression equations are relatively large. It is suggested that care must be taken in the use of these relationships for the prediction of streamflow characteristics for ungauged basins within the study area. With the aim of improving the predictive strength of the regression relationships, it is possible to suggest several extensions of the present research. These suggestions for further research are outlined in the following section of this chapter.

The present analyses have been successful with respect to the second aim of the study, to add to our overall knowledge of prairie hydrologic patterns through the identification of climatic and other physical geographic patterns which are closely related to streamflow characteristics. In order to illustrate this conclusion, the results of the full study area stepwise multiple regression analyses, after factor analytic screening of the independent variables, are considered.

The use of factor analytic techniques in the screening of the independent variable sets led to the development of more meaningful regression equations. In the case of the full study area analyses by these techniques (Section 4.4.3), the original set of 39 independent variables was collapsed to a group of nine variables which were relatively free of multicollinearity, and therefore approximated the assumption of independence. The stepwise multiple regression analysis for each of the dependent variables, employing this set of nine independent variables, resulted in significant regression equations in which the signs of the regression coefficients conformed to physical theory. The relationships were found to be stable when applied to an independent test sample of 15 basins.

Of particular interest, in the present study, is the consistently strong influence of the drainage area measures in the regression models. In all cases, the single most important variable is the LCDA measure. This is, of course, not unexpected in that larger contributing drainage areas are expected to produce larger streamflow events. However, a second drainage area measure the LMCDA, was also found to be significant in the regression models. This variable exhibited a negative effect on the magnitudes of streamflow events. Although this variable is a measure of the difference between the TDA and the CDA measures, its significance in the equations would suggest that it provides an index of the non-contributing area which may have been included in the CDA measure as a result of measurement errors. It is suggested that the true meaning of this variable may be that in cases where some non-contributing drainage areas have been delimited by the relatively crude measures of the present study, there may be further non-contributing areas which are in proportion to the measured NCDA percentage. At any rate, the importance of drainage area measures has been confirmed, and it would seem that a possible direction for further investigations has been established (Spence 1972).

Another observation of particular significance in the present study related to the spatial distribution of residuals from the regression analyses. When the residuals from the various multiple regression analyses were mapped, the resulting patterns were local in nature and did not reveal any large scale regional patterns. This observation leads the author to suggest that local variations in prairie hydrologic patterns are the dominant factor in limiting the predictive value of the regression relationships. The importance of these local variations has been confirmed by the failure of the regional subdivision of the second stage of the analysis to result in significant improvements in the predictive strength of the regression relationships.

In addition to establishing statistical relationships, two methodological conclusions have been reached. The first conclusion is that the lognormal distribution is the most appropriate 2-perameter distribution for use in the frequency analyses of both the annual yield and annual flood flow data series for prairie streams. The lognormal distribution was selected over the normal, Gumbel and log-Gumbel distributions on the basis of its empirical fit to the available data series. This conclusion is an empirical confirmation of the use of this distribution which previously has been widely employed in engineering hydrology for plains' streams (for examples of the use of the lognormal distribution in hydrological studies for the plains see Ansley 1959; and Neill et al., 1970).

The second methodological conclusion is that factor analytic screening of the independent variables prior to regression analysis results in more meaningful regression models. The models developed by this technique conform to physical theory while sacrificing only a small degree of explanation relative to the all variable regression approach.

5.4 SUGGESTIONS FOR FURTHER RESEARCH

In the present analyses, only a limited degree of success has been achieved in the development of statistical models for the prediction of streamflow characteristics for ungauged plains' basins. On the basis of these results it is possible to make several suggestions for the extension of the present research with the aim of improving the relationships. Of particular relevance to further research is the observed importance of local deviations as exemplified in the results of the regression analyses of the present study. It is proposed that extensions to the present research should concentrate on an examination of local scale patterns which may be related to these local variations.

The dependent variables in the present research, the selected streamflow characteristics, have been estimated by frequency analyses of the available streamflow data series. The number of streamflow data available can only be increased by continued data collection over time. Therefore with respect to the dependent variables, it is not possible to rely on further data collection at this time; but rather, efforts to improve the relationships must concentrate on making better use of the available records. In this regard, it is suggested that a review of the reliability of the available streamflow data be considered. In the present research the published streamflow data and gauging station descriptions have been accepted as a basis for basin selection. In view of the numerous local anomalies which have been noted in the analyses, it is suggested that the streamflow data for anomalous basins be reviewed with the aim of detecting any inconsistencies or human influences such as diversions or storage developments which have not been previously identified. Such unrecognized limitations in the original data set may have resulted in some of the prediction errors in the present analyses.

The first group of independent variables employed in the study, the climatic measures, were based on 30 year climatic normals of temperature and precipitation. These data were employed for ease of data compilation and on the assumption that the year to year variations in hydrologic conditions would be accounted for in the frequency analyses of the streamflow data. In view of the limited success of the present analysis, it is proposed that consideration should be given to the year to year variations in climatic patterns. Such considerations might be based on frequency analyses of several years of climatic data. This approach to developing climatic variables may be of particular relevance in the semi-arid sections of the study area in which the annual variations in water balance patterns are relatively pronounced.

The second group of independent variables, the measures of other physical geographic patterns, should also be re-examined with a view to explaining some of the local hydrologic anomalies. The drainage area measures employed in the present study are far from ideal, in that a high degree of subjectivity exists in their measurement. However, these measures have proven to be highly significant in the present study and it would seem reasonable to suggest that further refinement in the methods of drainage area delimitation might result in a reduction of the errors in the analyses. Another group of physical measures which may hold the key to some of the local anomalies includes measures of surficial geology, soils, and landuse patterns. All of these variables which are operative on a local scale have been omitted from the present study because of a lack of suitable data sources. However, any extension to the present study should include efforts to provide at least some indices of these factors which are closely related to infiltration rates and capacities.

The present study has resulted in the successful development of meaningful models of prairie hydrologic patterns. The relationships are statistically significant, stable when applied to independent data, and in agreement with physical theory. Unfortunately, the magnitudes of the standard errors of estimates associated with the regression equations are relatively large and limit the predictive usefulness of the models. Several extensions to the present research have been proposed with the aim of improving the predictive strength of these relationships. At the present time, the relationships developed in this study provide a basis for the preliminary estimation of streamflow patterns for ungauged areas. Care must be exercised in the interpretation of these estimates and it is anticipated that further research along the lines suggested above will result in more accurate and useful relationships.

LIST OF REFERENCES

Amorocho J. and W.E. Hart 1964. A critique of current methods in hydrologic systems investigation. *Transactions of the American Geophysical Union*, Vol. 45, pp. 307-321.

Ansley, R.W., 1959. Statistical Interpretation of Canadian Hydrometric Data. Unpublished M.Sc. Thesis, Department of Civil Engineering, University of Alberta, 47pp.

Benson, M.A., 1959. Channel-slope factor in flood-frequency analysis. *Journal of the Hydraulics Division, Proceedings of the American Society of Civil Engineers*, Vol. 35, Separate No. 1994, 9pp.

___________, 1960. Characteristics of frequency curves based on a theoretical 1000 year record. *in*, T. Dalrymple, Flood frequency analyses. *U.S.G.S. Water-Supply Paper*, 1543-C, pp. 51-74.

___________, 1962a. Evolution of methods for evaluating the occurrence of floods. *U.S.G.S. Water-Supply Paper*, 1580-A, 30pp.

___________, 1962b. Factors influencing the occurence of floods in a humid region of diverse terrain. *U.S.G.S. Water-Supply Paper*, 1580-B, 64pp.

___________, 1964. Factors affecting the occurrence of floods in the southwest. *U.S.G.S. Water-Supply Paper*, 1580-D, 72pp.

___________, 1965. Spurious correlations in hydraulics and hydrology. *Journal of the Hydraulics Division, Proceedings of the American Society of Civil Engineers*, Vol. 91, No. HY4, pp. 35-42.

Benson, M.A. and R.W. Carter, 1969. Areal extension of streamflow information by multiple regression techniques. *in*, *Water Balance in North America*. Proceedings of the Symposium, American Water Resources Association, A.H. Laycock. Ed., pp. 137-142.

Black, P.E., 1966. Thornthwaite's mean annual water balance. Silviculture General Utility Library Programme GU-101, State University, College of Forestry, Syracuse University, Syracuse, New York, 18pp.

Bostock, H.S., 1964. A provisional physiographic map of Canada. *Geologic Survey of Canada, Paper* 64-35, 24pp.

Canada, Department of Energy, Mines and Resources, 1965-1968a. *Surface Water Data Alberta*. Inland Waters Branch, Ottawa.

___________, 1965-1968b. *Surface Water Data Saskatchewan*. Inland Waters Branch, Ottawa.

__________, 1965-1968c. *Surface Water Data Manitoba*. Inland Waters Branch, Ottawa.

__________, 1969a. *1968 Surface Water Data Reference Index Alberta*. Inland Waters Branch, Ottawa, 53pp.

__________, 1969b. *1968 Surface Water Data Reference Index Saskatchewan*. Inland Waters Branch, Ottawa, 42pp.

__________, 1969c. *1968 Surface Water Data Reference Index Manitoba*. Inland Waters Branch, Ottawa, 42pp.

__________, 1970. Planning of a hydrometric network for the Prairie Provinces, British Columbia and the Territories. Inland Waters Branch, Report No. 5019-1-70, prepared by the Shawinigan Engineering Co. of Canada, 162pp.

Canada, Department of Northern Affairs and Natural Resources, 1930-1964. *Surface Water Data for Arctic and Western Hudson Bay Drainage*. Water Resources Papers, Nos. 68, 71, 75, 82, 84, 88, 92, 97, 101, 105, 109, 112, 117, 121, 125, 127, 132, 135, 138, 141, 145.

Canada, Department of Transport, 1968a. *Climatic Normals*. Vol. 1, Temperature, Meteorological Branch, Toronto, 66pp.

Canada, Department of Transport, 1968b. *Climatic Normals*. Vol. 2, Precipitation, Meteorological Branch, Toronto, 110pp.

Chow, V.T., 1950. Discussion of annual floods and the partial duration series. by W.B. Langbein. *Transactions American Geophysical Union*, Vol. 31, pp. 939-941.

__________, 1964a. Statistical and probability analysis of hydrologic data. *in* *Handbook of Applied Hydrology*, V.T. Chow ed., Toronto, McGraw-Hill, Section 8, Part 1.

Clibbon, P.B. and L.E. Hamelin, 1967. Landforms. *in* *Canada, A Geographical Interpretation*. J. Warkentin ed., Toronto, Methuen.

Cole, G., 1966. An application of the regional analysis of flood flows in *River Flood Hydrology*. The Institution of Civil Engineers, London, pp. 39-57.

Collier, E.P. and G.A. Nix, 1967. Flood frequency analysis for the New Brunswich-Gaspé Region. *Technical Bulletin*, No. 9, Canada, Department of Energy, Mines and Resources, Inland Waters Branch, Ottawa, 20pp.

Cooper, R.H. and R.F. Howells, 1969. Subroutine GPL - an autoplotting routine for use with the calcomp plotter. Mimeographed program documentation, Department of Civil Engineering, University of Alberta, Edmonton, 10pp.

Coulson, A., 1967a. Flood frequencies of Nova Scotia streams. Technical Bulletin, No. 4, Canada, Department of Energy, Mines and Resources, Inland Waters Branch, Ottawa, 31pp.

________, 1967b. Estimating runoff in Southern Ontario. Technical Bulletin, No. 7, Canada, Department of Energy, Mines and Resources, Inland Waters Branch, Ottawa, 19pp.

Coulson, A. and P.N. Gross, 1967. Measurement of the physical characteristics of drainage basins. Technical Bulletin, No. 5, Canada, Department of Energy, Mines and Resources, Inland Waters Branch, Ottawa, 22pp.

Dalrymple, T., 1960. Flood frequency analysis. U.S.G.S. Water-Supply Paper, 1543-A, 80pp.

Dixon, W.J., ed., 1970. BMD Biomedical Computer Programs. University of California Publications in Automatic Computation, No. 2, Los Angeles, University of California Press, 600pp.

__________, 1970a. BMD Biomedical Computer Programes X Series Supplement. University of California Publications in Automatic Computation, No. 3, Los Angeles, University of California Press, 260pp.

Durrant, E.F., 1959. Discussion of - Practical Flood Frequency Analysis, by E. Kuiper. Proceedings of Hydrology Symposium No. 1 Spillway Design Floods. National Research Council, Ottawa, pp. 96-99.

Durrant, E.F. and S.R. Blackwell, 1959. The magnitude and frequency of floods on the Canadian Prairies. in Proceedings of Hydrology Symposium No. 1 Spillway Design Floods. National Research Council, Ottawa, pp. 101-160.

Ezekiel, M. and K.A. Fox, 1961. Methods of Correlation and Regression Analysis. New York, Wiley, 548pp.

Goulding, B.L. and D.E. Low, 1960. Physical characteristics of drainage basins. Journal of the Hydraulics Division, Proceedings of the American Society of Civil Engineers, Vol. 86, No. HY3, Paper No. 2409 ,11pp.

Harman, H.H., 1967. Modern Factor Analysis. Chicago, University of Chicago Press, 474pp.

Holmes, R.M. and G.W. Robertson, 1959. A modulated soil moisture budget. Monthly Weather Review, Vol. 87, No. 3, 7pp.

Horton, R.E., 1945. Erosional development of streams and their drainage basins; hydrophysical approach to quantitative morphology. Bulletin of the Geological Society of America, Vol. 56, pp. 275-370.

Howe, G.M., H.O. Slaymaker, and D.M. Harding, 1967. Some aspects of the flood hydrology of the upper catchments of the Severn and Wye. The Institute of British Geographers Transactions and Papers, Publication 41, pp. 33-58.

Jarvis, C.S. and others, 1936. Floods in the United States magnitude and frequency. U.S.G.S. Water-Supply Paper, 771, 497pp.

Karuks, E., 1964. Prediction of daily average flood flows for Southern Ontario rivers. Ontario Joint Highway Research Programme Report, No. 21, Department of Civil Engineering, University of Toronto, 59pp.

Kennard, J., K.T. Bass and E. Sandeman, 1963. The effect of the average height of a catchment on the estimation of the intensity of flood discharge. International Association of Hydraulic Research, 10th Congress, Vol. 2. pp. 173-188.

King. L.J., 1969. Statistical Analysis in Geography Toronto, Prentice-Hall, 288pp.

Langbein, W.B., 1949. Annual floods and the partial duration flood series. Transactions American Geophysical Union, Vol. 30, pp. 879-881.

________, 1960. Hydrologic data networks and methods of extrapolating or extending available hydrologic data. in Hydrologic Networks and Methods, World Meteorological Organization, Bangkok, pp. 12-41.

Langbein, W.B. and others, 1947. Topographic characteristics of drainage basins. U.S.G.S. Water-Supply Paper, 968-C, pp. 125-155.

Laycock, A.H., 1959. Local water supplies in the prairies. Unpublished mineographed paper, presented to the Canadian Association of Geographers Meeting, Saskatoon, 12pp.

________, 1967. Water deficiency and surplus patterns in the Prairie Provinces. Prairie Provinces Water Board Report, No. 13, Hydrology Division, Prairie Farm Rehabilitation Administration, Regina, 92pp.

Leopold, L. B., 1959. Probability analysis applied to a water-supply problem. U.S.G.S. Circular, No. 410, 18pp.

Lull, H.W. and H.W. Anderson, 1967. Important watershed characteristics affecting water yield, flood peaks, and erosion and sedimentation and the basic data needed for prediction. International Conference on Water for Peace, Washington, 7pp.

Lull, H.W. and W.E. Sopper, 1966. Factors that influence streamflow in the Northeast. Water Resources Research, Vol. 2, No. 2, pp. 371-375.

MacIver, D.C., 1970. Macroclimatic Zonation in Northern Alberta. Unpublished M.Sc. Thesis, Department of Geography, University of Alberta, Edmonton, 187pp.

Morisawa, M.E., 1959a. Relation of quantitative geomorphology to streamflow in representative watersheds of the Appalachian Plateau Province. Office of Naval Research Technical Report, No. 20, 94pp.

__________, 1959b. Relation of morphometric properties to runoff in the Little Mill Creek Ohio drainage basin. Office of Naval Research Technical Report, No. 17, 8pp.

Nash, J.E. and B.L. Shaw, 1966. Flood frequency as a function of catchment characteristics. in River Flood Hydrology. The institution of Civil Engineers, London, pp. 115-136.

Neill, C.R., D.I. Bray, M.F. Schouten and J.R. Card, 1970. Selected characteristics of streamflow in Alberta. Research Council Of Alberta, Edmonton.

Penman, H. L., 1963. Vegetation and Hydrology. Commonwealth Bureau of Soils Technical Communication No. 53, Harpenden, England, 124pp.

Riggs, H.C., 1968. Frequency curves. in Techniques of Water-Resources Investigations of the United States Geological Survey, Chapter A2, U.S.G.S., 15pp.

Sanderson, M.E. and D.W. Phillips, 1967. Average annual water surplus in Canada. Climatological Study, No. 9 ,Canada, Department of Transport, Meteorological Branch, Toronto, 76pp.

Schneider, W.J., 1965. Areal variability of low flows in a basin of diverse geologic units. Water Resources Research, Vol. 1, No. 4, pp. 509-515.

Semple, R.K., E. Casetti and L.J. King, n.d.. The determination of the optimal number of groupings in classification problems. Discussion Paper, No. 10, Department of Geography, Ohio State University, 38pp.

Slaymaker, H.O. and W.W. Jeffrey, 1969. Physiography-land use interactions in watershed management. Water Balance in North America. Proceedings of the Symposium, American Water Resources Association, A.H. Laycock ed., pp. 183-192.

Slivitzky, M.S., 1966. Appraisal of methods of analysis. Proceedings of Hydrology Symposium No. 5 Statistical Methods in Hydrology National Research Council, Ottawa, pp. 183-192.

Solomon, S., 1966. Statistical association between hydrologic variables. Proceedings of Hydrology Symposium No. 5 Statistical Methods in Hydrology, National Research Council, Ottawa, pp. 55-115.

Solomon, S.E., J.D. Denouvilliez, E.J. Chart, J.A. Woolley and C. Cadou, 1968. The use of a square grid system for computer estimation of precipitation, temperature, and runoff. Water Resources Research, Vol. 4, No. 5, pp. 919-929.

Spence, E.S. 1972. The Contributing Drainage Area Variable in the Hydrology of the Canadian Plains. International Geography 1972. Eds. W.P. Adams and F.M. Helleiner, University of Toronto Press, pp. 198-200.

Stammers, W.N., 1966. The application of multivariate techniques in hydrology. Proceedings of Hydrology Symposium No. 5 Statistical Methods in Hydrology. National Research Council, Ottawa, pp. 255-267.

Stichling, W. and S.R. Blackwell, 1958. Drainage area as a hydrologic factor on the glaciated Canadian prairies. International Association of Scientific Hydrology General Assembly, Toronto, Pub. No. 45, pp. 635-376.

Strahler, A.N., 1964. Geology, quantitative geomorphology of drainage basins and channel networks. in Handbook of Applied Hydrology. V.T. Chow ed., Toronto, McGraw-Hill, Section 4, Part II, pp. 4-39 to 4-76.

Tennessee Valley Authority, 1966. Design of a hydrologic condition survey using factor analysis. International Association of Scientific Hydrology Bulletin, Annee 11, No. 2, pp. 131-176.

Thiessen, A.H., 1911. Precipitation averages for large areas. Monthly Weather Review, Vol. 39, pp. 1082-1084.

Thomas, D.M. and M.A. Benson, 1970. Generalization of Streamflow Characteristics from drainage-basin characteristics. U.S.G.S. Water-Supply Paper, 1975, 55pp.

Thronthwaite, C.W., 1948. An approach toward a rational classification of climate. Geographical Review, Vol. 38, No. 1. pp. 55-94.

Thornthwaite, C.W. and J.R. Mather, 1957. Instructions and tables for computing potential evapotranspiration and the water balance. Publications in Climatology, Vol. 10, No. 3, Drexel Institute of Technology, Laboratory of Climatology, Centerton, N.J., pp. 185-311.

United States, Department of Commerce, 1962a. Decennial census of United States climate - monthly normals of temperature, precipitation and heating degree days, Minnesota. Climatography of the United States, No. 81-17, Washington, 2pp.

__________, 1962b. Decennial census of United States climate - monthly normals of temperature, precipitation and heating degree days, Montana Climatography of the United States, No. 81-28, Washington, 2pp.

__________, 1962b. Decennial census of United States climate - monthly normals of temperature, precipitation and heating degree days, North Dakota. Climatography of the United States, No. 81-28, Washington, 2pp.

Wallis, J.R., 1965. Multivariate statistical methods in hydrology - a comparison using data of known functional relationship. Water Resources Research, Vol. 1, No. 4, pp. 447-461.

__________, 1968. Factor analysis in hydrology - an agnostic view. Water Resources Research, Vol. 4, No. 3, pp.521-527.

Ward, R.C., 1967. Principles of Hydrology. Berkshire, England, MacGraw-Hill, 403pp.

Wilson, E.M., 1969. Engineering Hydrology. London, Macmillan, 182pp.

Wolf, P.O., 1966. Comparison of methods of flood estimation. in River Flood Hydrology. The Institution of Civil Engineers, London, pp. 1-23.

Wong, S.T., 1963. A multivariate statistical model for predicting mean annual flood in New England. Annals of the Association of American Geographers, Vol. 53, pp. 298-311.

GB
1230
.P7
S64
1972

Spence, Edward S.

An analysis of the relationships of selected streamflow characteristics to physical georgraphic patterns in the plans area of the Canadian Prairie Provinces.