FORECASTING SPRING SEASON WATER YIELD USING MULTIVARIATE ANALYTIC TECHNIQUES

by

PAUL J. PILON

A thesis
presented to the School of Graduate Studies and Research
in partial fulfillment of the
requirements for the degree of
Master of Applied Science
in
Civil Engineering

Ottawa, Ontario, 1981

I hereby declare that I am the sole author of this thesis.

I authorize University of Ottawa to lend this thesis to other institutions or individuals for the purpose of scholarly research.

PAUL J. PILON

I further authorize University of Ottawa to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

PAUL J. PILON

University of Ottawa requires the signatures of all persons using or photocopying this thesis. Please sign below, and give address and date.

ABSTACT

Considerable interest is being shown in the adoption of sophisticated techniques in the design and operation of water resources systems in the light of increasing energy awareness and costs. The need for improved seasonal water yield models is accentuated by the monetary implications associated with their development.

Several studies indicate that the regression of principal components may possibly generate forecast models superior to the models derived by the traditional regression analysis. This study has investigated the effectiveness of principal-component regression in the forecasting of spring season water yield. This was achieved through the collation of regression and principal-component regression models for a multi-reservoir hydroelectric system in the Saguenay-Lac St Jean region of Quebec.

A general model building scheme was adopted to aid in model development. In addition, a detailed residual analysis (randomness, distribution), the adoption of split-sampling technique where possible, the analysis of the rationality of regression coefficients, and the analysis of the indicators of the relative importance of the independent/predictor variables were performed to aid in the model verification step of model development. Inclusion of the above analyses aided in the detection of inadequate forecast models.

Results of the study indicated that principal-component regression was indeed superior to the traditional regression analysis in the development of spring season water yield models.

ACKNOWLEDGMENT

The author wishes to acknowledge the supervisor of this research work, Dr. K. Adamowski, for his guidance and personal encouragement throughout the program.

The author also wishes to thank his wife Janice for her personal sacrifices and understanding during the preparation of this thesis.

Finally the author wishes to thank Alcan Smelters and Chemicals Ltd. for assistance offered concerning data for the study area.

Financial assistance for the program was supplied by grants from the National Research Council - Grant Number A7467.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

INTRODUCTION

1.1 MOTIVATION

Amorocho and Hart(2) purport that two distinct groups of investigators exist in the modes of approach to the basic hydrologic problem of establishing the relationships between precipitation and runoff, as shown in Figure 1 The first division of research concerning the hydrologic cycle is termed 'physical hydrology'. Its primary motivation is "the study of the physical phenomena [where] the eventual practical application of this knowledge for engineering and other purposes is recognized but not explicitly sought." The second path of investigation is termed 'systems investigation'. Its prima caussa is the "investigation of hydrologic systems for the explicit purpose of establishing quantitative relationships between precipitation and runoff, which can be used for reconstruction or prediction of flood sequences and watershed yields."

Amorocho and Hart(2) delineate the researchers of the 'system investigations' path as consisting of two principal categories called 'parametric hydrology' and 'stochastic hydrology'. They define parametric hydrology as being the "development of relationships among physical parameters in-

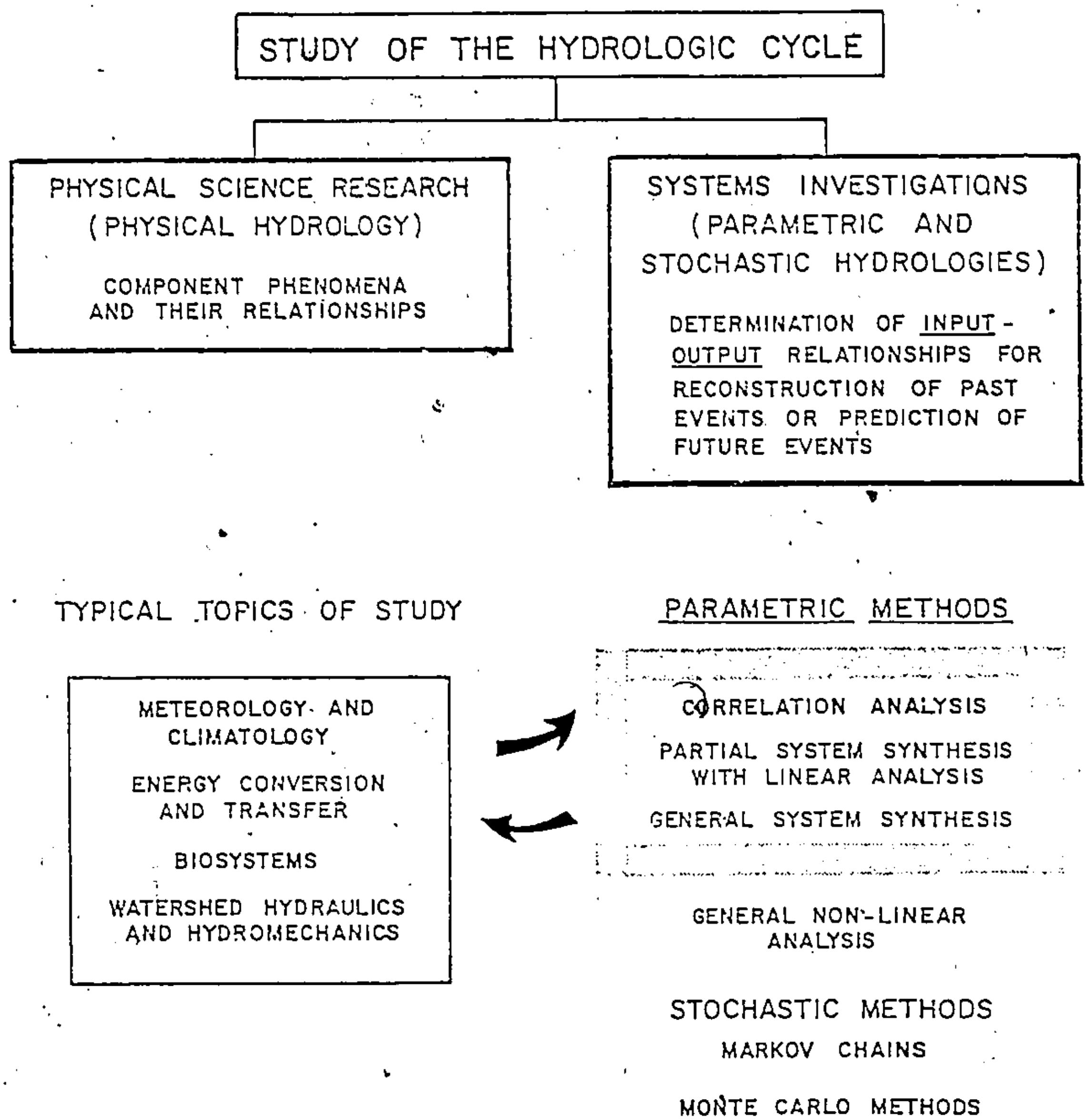


Figure 1 : Summary of hydrologic analysis methods(2).

volved in hydrologic events and the use of these relationships to generate, or synthesize, non-recorded hydrologic sequences." One of the principal methods of parametric hydrology is correlation analysis, which represents one of the basic approaches to forecasting watershed yield (3).

Considerable interest is being shown in the adoption of sophisticated techniques in the design and operation of water resources systems in the light of increasing energy awareness and costs. The need for improved seasonal water yield models is accentuated by the monetary implications associated with their development. Various sources (52,64,74) indicate realizable benefits which would result from even a one or two percent improvement in the cost effectiveness of current programs. Due to the increasing complexity and interaction of various water uses--energy generation; navigation; flood, irrigation, and pollution control--the development of reliable and accurate forecast models is prerequisite for optimum water resources management. The desire for accurate forecasts is easily demonstrated in the production of hydroelectric power, where "predicting the spring runoff...is very important to the other 41 weeks of operation" (64).

The river forecasting problem in hydrology consists of two basic types of forecasts (3). The first method is short-term forecasting which estimates discharge from a day, to a week, to a month in the future (8,17). Short-term forecasts are commonly based on the following variables: snow cover, antecedent streamflow, and observed plus predicted meteorological conditions. Several models of the snow accumulation and melting process attempt to mathematically represent each of the components of the above process(3,17,57,77). Figure 2 shows a typical flow-chart of such a model.

Anderson(3) indicates that the second basic type of forecast is the seasonal water yield. Statistical correlation-multiple regression models represent the predominant method for the forecasting of seasonal water yield. This method relates the measured snow cover, past precipitation, and other hydrometeorological variables with the seasonal water yield.

Snyder(65) points out one of the basic assumptions of statistical correlation models or multiple regression. He states that the so-called independent variables should not be correlated with each other, while noting that in hydrologic analyses the independent variables may not be uncorrelated. Koelzer and Ford(45) indicate that high correlation between independent variables may result in an apparent lack of model consistency. Sharp et al(62) also comment on the erroneous results which may be obtained when the independent variables are highly multicollinear. The existence of multicollinear independent variable in hydrology led Snyder(65) to review the multivariate technique of principal component

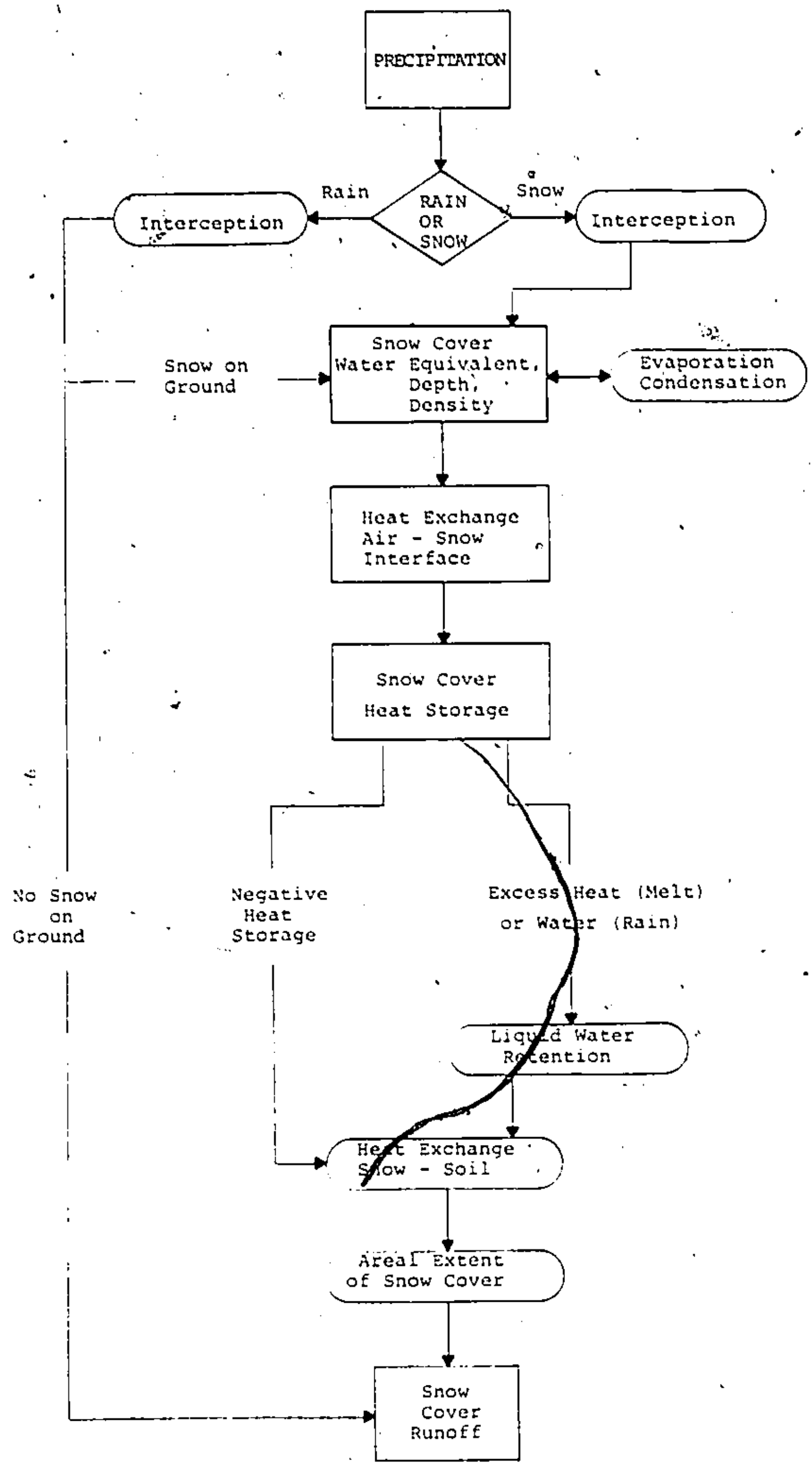


Figure 2 : Flow chart of typical simulation model(3).

analysis. This technique deals with the determination of all the truly independent components of variation in an array of variables. Snyder(65) states that component analysis has "excellent possibilities...in hydrologic studies."

The prime objective of this study is to investigate the possibilities of principal-components regression for the forecasting of spring season water yield. A collation of tentative regression and principal-components regression models for a multi-reservoir hydroelectric system in the Saguenay-Lac St-Jean region of Quebec will be viewed.

1.2 LITERATURE REVIEW

1.2.1 Regression as a Forecasting Tool

The technique of regression analysis represents one of the earliest statistical methods to be applied in water resources. Regression analysis is used as a complement to hydrologic theory to estimate hydrologic parameters and hydrologic models which could be applied in places where little information is available. Regression models are basically used to forecast future events and to establish some causal relationship among variables(78). Snyder's synthetic hydrograph technique(28) represents an early application of regression analysis, as does the use of regression analysis for the prediction of runoff volume to aid in the operation of power utilities(11,24,64).

Early investigations into the forecasting of the volume of the spring freshet are evident. Finlayson (24) examines several prominant factors concerning spring runoff in the northern part of the St. Maurice watershed in the province of Quebec. As almost half of the annual runoff to the Gouin reservoir in the St. Maurice watershed is represented by the spring freshet, he indicates that snow cover measurements should form the basis of forecast models. However, his study reveals that no "definite relationship" exists between freshet volume and the watershed's snow cover water equivalence. This relationship is not improved by the inclusion of the precipitation occurring between the date of the snow pack and the beginning of the freshet. The volume of the spring freshet does, however, show a strong relationship with the precipitation occurring during the freshet period--a factor unknown at the time of the forecast. He concludes that the water content of the snow pack does not have a dominant influence on the volume of the spring freshet, thus, lessening the possibility of accurate forecasts based on snow surveys.

The following year, at the meeting of the 1954 Eastern Snow Conference, Hopkins (33) discusses the selection of independant variables for the prediction of spring runoff. He indicates that precipitation and snow survey records constitute the bulk of accurate information collected. Other phenomena such as wind, percent of possible sunshine, temperature, and relative humidity affect the volume of the runoff during snowmelt but can not be satisfactorily estimated. He cautions that the averages of precipitation and snow surveys are only indices which may, on occasion, give biased results for a short period of time. The most imperceived factor in predicting spring runoff is identified as being the "water loss"; that is, the amount of water required to saturate the ground so that water will run off.

The realization that the inclusion of significant predictor/ independent variables aid in the generation of a more accurate regression model for the forecasting of the volume of the freshet, marks the beginning of a search for other possible variables. In addition to the snow survey and precipitation indices, Cavadias(11) attempts to introduce the variables of groundwater, wind, and temperature. Wind, as a variable, proves to be insignificant and is discarded. The temperature index during the freshet is introduced via "evapotranspiration", while the February runoff of an adjacent river is implemented as an indirect measure of the groundwater conditions. With the inclusion of both freshet temperature and precipitation, significant results concerning the prediction of the freshet volume are insinuated. Snow survey data or snow water equivalent plus the precipitation which occur between the date of the snow survey and the commencement of the freshet do not yield an accurate relationship with the volume of the freshet(11).

Cavadias(12) went on to analize the problem of forecasting the spring runoff in Quebec using a short-term approach. The introduction of several independent variables, the analysis of the distributional assumptions concerning the independent variables, and the subdivision of the runoff period into nine periods to increase the accuracy of the prediction are the main points brought forth in the paper. The inclusion of the precipitation during the period under question and the runoff from the previous period, as independant variables, implies a short-term rather than seasonal model. Cavadias notes that the introduction of precipitation as an independent variable necessitates the forecasting of this phenomena. However it is also noted that the attainable accuracy of such forecasts is not satisfactory. Therefore, a total of seven explanatory variables are introduced to the regression analysis. The first is the water equivalent of the snow cover for the 9325 km^2 (3600 mi^2) drainage area of the reservoir. This variable is based on measurements made at only two snow courses in the watershed. The second variable is the February runoff of the Batiscan River. As in his previous paper (11), February runoff is used as an index of groundwater conditions in the watershed under study. Runoff during the period prior to the predicted period represents the third independent variable. The fourth variable is defined as the cumulative degree days above $0^{\circ}C$ ($32^{\circ}F$) during the period under consideration. The precipitation forecasted at the Gouin reservoir during this period is identified as the fifth variable, while the sixth is termed "the water budget variable." It is defined as the water equivalent of snow at Gouin on March 15, plus the precipitation at Gouin from March 1 to the beginning of the period minus the run-off from March 16 to the beginning of the period. The seventh and last variable is the cumulative degree days above 0°C (32°F) at Gouin. The value of degree days is computed from March 1 to the beginning of the period to be forecasted. The regression models for the nine periods contain from two to five of the above mentioned seven independent variables, depending on the period in question. Cavadias(12) comments that a scattergram of spring runoff using the regression models does not "show any striking improvements over the results achieved in the 1955 study [11]."

Cavadias and Brisebois(15) discuss a similar study concerning the application of multiple regression analysis to the forecasting of the spring inflow of the Cabonga Reservoir on the Gatineau River in Quebec. They list five general categories of independent variables as being pertinent to the forecasting of the spring runoff of the Cabonga reservoir. They are the precipitation before or during the flood, the runoff before the flood, the degree days, the influence of the ground water conditions, and the water budget. The degree days, the influence of the groundwater conditions, and the water budget are calculated by the same method as described by Cavadias(11,12). Six regression models forecasting the volume of six different periods of the spring runoff are presented. The depicted periods are: April; May; June; April and May; May and June; and April, May, and June.

In the western United States, the Soil Conservation Service of the Department of Agriculture(66) discuss their use of regression analysis to obtain April through September runoff for the Columbia River at the International Boundary. The water content of the snow pack, precipitation, soil moisture beneath the snow pack, soil temperature, base flow of individual rivers, wind speed, and solar radiation represent the independent variables in their study. The derived regression model to forecast the April through September runoff is a linear additive model consisting of the independent variables: snow water index, the November 1 base flow index, and the average precipitation for April. The study does not define the above mentioned independent variables. It is assummed April precipitation is unknown at the time of the April to September forecast. If the proposed forecast model is to be operational, the inclusion of future phenomena as independent variables necessitates their accurate prediction. However, the study does point out that "75% to 85% of the streamflow [in the western states] is produced by the snow that accumulates in the mountainous areas"(66). This snow pack constitutes the main variable in the determination of the summer streamflow, unlike the east where the summer streamflow would be somewhat less dependent on the previous winter snow pack.

The great influence of the snow pack in the Columbia Mountain system of western North America is also evident in studies of seasonal snowmelt volumetric forecasting in British Columbia. Hunter(37) provides a general description of the seasonal streamflow forecasting techniques of the British Columbia Water Resources Service. Regression equations are used for the seasonal volumetric forecasts. Hunter classifies the independent variables in regression model construction into three main categories. The first category includes the factors having influence on the dependent variable, that is the spring and summer volume flow prior to the start of the snow accumulation period. In this category one finds late summer and autumn precipitation as a soil moisture index. The inclusion of September through October precipitation improves the regression model's reliability on many river basins in British Columbia. The second variable of the first category is termed the "carry-over groundwater storage index." This index is most commonly taken as the previous spring and summer runoff. He found this variable not to be too significant in improving the British Colombia forecast models. The third and last variable of the first category is base or late autumn flow for a specified data. November 1 flow is usually selected as the late autumn flow index. However, the inclussion of this third variable is ineffective. The second main category of variables in regression model construction, as shown by Hunter(37), include hydrometeorological variables obtained during and at the end of the snow accumulation season. The first variable of this category is snow course water equivalents measured at 170 sites throughout British Columbia. This is thought to be more significant than the second variable, that is the accumulated winter precipitation for the November to March period. The inclusion of both the snow course water equivalents and the winter precipitation into a single index improves the derived model on some basins. The third and final variable of the second main category is that of groundwater conditions. Various accumulated streamflow values between the late autumn and early spring are used as possible indicators of ground water conditions. This variable proves valuable on many basins in British Columbia when used in conjunction with snow pack and precipitation data.

The third main category includes factors which occur subsequent to the forecast date. Hunter(37) realizes that the variables in the third category do not occur prior to the forecast periods, but have an influence on the resulting runoff. It should be noted that the volumetric runoff forecasts are made for the March-September, the April-September, and the May-September periods. On low-level basins, these periods may terminate in July. The variables in this category are the precipitation which occurs during the freshet period, accumulated degree-day temperature, and evaporation pan data. Forecast models, which include precipitation that occurs during the forecast period as an independent variable, assume that average precipitation occurs during the runoff period. Various precipitation values may be inserted in the model to provide forecasting for planning purposes.

Koelzer and Ford (45) attempt to determine the influence of various climatic factors on snowmelt runoff using a Bureau of Reclamation multiple-purpose reservoir system located in northern Colorado and southern Wyoming. Analogous to Hunter (37), Koelzer and Ford group various hydrometeorological variables into three broad categories. The general categories were as follows: a) factors prior to the snow accumulation period; b) factors reflecting the influence of conditions during the snow-accumulation period; and c) factors reflecting the influence of conditions subsequent to the commencement of snowmelt runoff season. The snowmelt runoff period is assumed to be April 1 through July 31. The antecedent hydrometeorological factors prior to the snow accumulation period are the July through September precipitation and the October through December streamflow. October 1 is assumed to be the start-date of the snow accumulation period. Its factors consist of precipitation, snow course water content, and temperature. Various combinations of the precipitation occuring between October 1 and the forecast date, as well as various snow course water equivalents, are used. The following measures of temperature are also analyzed: a) September-November average temperature; b) November-December temperature excess; c) October-December temperature excess; d) January-March temperature excess; e) November-March temperature excess; f) February average temperature; g) March average temperature; and h) February-March average temperature. The temperature index represents the cumulative degree-days of daily maximum temperature over the given period and a chosen base temperature. The base temperature is selected as the point where melting would begin at the average snow elevation. The third general category of hydrometeorological factors, which are to represent the melt-season factors, include precipitation and temperature as independant variables. The periods of subsequent precipitation analyzed include: a) April-June; b) May-June; c) April-May; d) May-July; e) June-July; f) May; g) June; and h) July. The sole subsequent temperature investigated is the average of April.

Results of the study by Koezer and Ford indicate that antecedent July-September precipitation, and fall and winter streamflow conditions do not prove satisfactory in expressing the effect of past hydrometeorological conditions on subsequent runoff. The authors of the report feel that antecedent hydroclimatic factors should have been significantly correlated with the snowmelt runoff. They propose three possible reasons for the failure of their models to show significant correlation of antecedent factors with the subsequent runoff. The explanations are: a) the inaccuracies in the primary factors may have obscured the effect; b) the historic records may not have included years of sufficient extreme in the antecedent factors to yield significant results; and c) the index chosen to reflect the antecedent state may not have been appropriate. However, analysis of the effects of winter precipitation and snow water content indicate both factors should be included when forecasting. Inclusion of both factors decreases the significance of the model as these factors tend to be highly interdependent. Koelzer and Ford point out that the inclusion of colinear independent variables could possibly result in a lack of consistency in the generated model.

The analysis of antecedent variables by Koelzer and Ford indicate that snow water content should not be broken into periods. Winter temperatures are also investigated as a supplement to the snow water content and precipitation. The authors conclude that in most cases the temperature index gives unsatisfactory results. The effect of precipitation during the defined snowmelt season on runoff is also investigated. It is intuitively obvious to the researchers that the precipitation of the freshet should have a pronounced influence on the melt season runoff. It is reasoned that the inclusion of the melt season precipitation, although unknown at the time of the forecast, would generate a regression model whose individual factors would be more correctly expressed. This is due to the incorporation of the major factors influencing the melt season runoff in the regression model. The inclusion of melt season precipitation in the equation increases the reliability of the estimate of influence of the factors known at the time of the forecast. However, the inclusion of melt season precipitation is only of academic interest as the standard error of the regression model does not satisfactoraly reflect the probable forecast error, as the melt season precipitation is unknown at the time of the forecast.

Ford(26) presents an initial study of forecasting seasonal runoff from two Alaskan watersheds. The two basins are: the Eklutna basin, 43 km (30 mi) northeast of Anchorage; and the Long River drainage, about 43 km (30 mi) east-southeast of Juneau. The Eklutna basin, having an area of 308 km^2 (119 mi^2), is roughly 8 percent glaciated. Mean runoff from this basin is about 1.0 m (3.25 ft) or 305.3 Mm^3 (10.78 billion cubic feet(BCF)). The Long River drainage area is roughly 25 percent glaciated. This watershed has a surface area of 84.2 km^2 (32.5 mi^2), and reports a mean annual runoff of approximately 4.9 m (16 ft) of depth or 410.6 Mm^3 (14.5 BCF). The influence of the glaciers on runoff is reported to be greatest during warm dry years and least during cold wet years. Runoff from such glaciers accentuates the influence of temperature on runoff. Precipitation and tempera-ture data comprise the climatological factors analyzed for forecasting seasonal runoff. The exclusion of other clima-tological factors in the study is reported as being due to the existance of only a relatively short period of discharge records for the sites. Of particular importance in these investigations is the analysis of evident interdependence of the effects of the two primary causal factors--precipitation and temperature. It is reported that an abundance of accu-mulated precipitation during the winter could possibly pro-duce a proportionely high discharge during the following summer months, depending upon the temperatures which occur during the snow melt season.

The annual flow patterns of the two Alaskan watersheds are characterized by extremely high flows occuring in the summer months. Ford(26) hypothesizes that early season ru-noff is felt attributable to the melting of the previous winter's accumulated snow cover. He notes that glacial melt, augmented by precipitation, has a sustained influence on summer season runoff, especially in the late seasons. The model proposed for the Eklutna watershed forecasts the major portion of the runoff season--July through August ru-noff--in kilo-acre feet. The model is based on October through June total precipitation and the mean of June's dai-ly maximum temperature. The independent variables in the derived regression model include precipitation, and a composite independent variable of precipitation multiplied by temperature. The derived regression model for forecasting the July through September runoff for the Long River watershed include January through June total precipitation and June's mean daily maximum temperature. The independent variables in the accepted model include precipitation and a multiplicative composite variable of precipitation and temperature. Ford(26) indicates that the two derived models are more accurate than steamflow estimates made in other western states, and that useful estimates of seasonal runoff are obtainable from hydrometeorological relationships derived by regression analysis.

A similar illation is presented by Sharp et al(62) concerning the development of various regression models to estimate seasonal runoff, as well as annual and monthly runoff. These authors analyze the 2388 km^2 (922 mi^2) Delaware River basin in Kansas to provide a background for an examination of the method of multiple regression analysis. This statistical technique is adopted because of its efficacy for the evaluation of the parameters affecting the water yield of the river basin. The authors delineate three tacit assumptions in the application of multiple regression analysis to hydrologic problems. Their assumptions are: a) the independent variables are error free-- error occurs only in the dependent variable; b) the variance of the dependent variable is constant regardless of the values of the independent variables--which is called the homoscedastic variance; c) the observed values of runoff be uncorrelated random variables. A forth assumption is proffered in the application of tests of significance. That is, the population of the dependent variable be normally distributed about the regression line for any level of the independent variables. However, the only information available concerning the distribution of the population is provided solely by the sample. The authors remark that usually in hydrologic analysis the samples are simply not large enough to afford reliable information about the distribution of the dependent variable about the regression line. From this, they deduce that the fourth assumption may be suspect, and that high coefficients of correlation and determination and high t-test values may be misleading. Possible violations of the first three assumptions are illustrated by the authors. The first assumption, which states errors are only present in the dependent variable, is obviously violated to some degree by hydrologic data. In the instance of the second assumption, the authors note that the variance in runoff values may not be entirely unaffected by the values of the independent variables. Precipitation is used to illustrate the possible violation of the second assumption. Riggs(60), in discussion of the paper by Sharp et al, states that their third premise which assumes the dependent variable to be random uncorrelated events is not appreciably violated when the dependent variable is annual, seasonal, or monthly flows. He also agrees that the first assumption is obviously violated, but adds that it appears such errors are of insufficient magnitude to have appreciable effect on the results.

The monthly analysis by Sharp et al indicate that for most months, precipitation and groundwater variables are significantly related to water yield. Temperatures, however, prove to be infrequently significant when related to monthly water yield. These monthly data for the Delaware River basin are combined to perform the seasonal analysis. The seasons are defined as: a) April through June; b) July through September; c) October and November; and d) December through March. Four hydrometeorological variables prove significant in all the seasons. They are monthly precipitation, accumulated monthly precipitation in excess of 12.7 cm (5 in) per month, first antecedent month's precipitation, and groundwater in storage. Monthly temperature proves significant only in the April through June season. The authors note that the signs of derived regression coefficients are not consistent. They indicate that this frequently occurrs in multiple regression analysis with the inclusion of two highly correlated variables, such as monthly precipitation and accumulated monthly precipitation in excess of 12.7 cm (5 in) per month. They also note that the inclusion of too many variables in a multiple regression analysis may possibly lead to difficulty in the physical interpritation of the results, and is exemplary practice. Riggs(60) states that due to the many factors affecting streamflow, the inability to recognize even some of these factors, and the lack of precision in describing these factors quantitatively, dictates that streamflow should be expressed as a statistical relation. He contends that multiple regression represents a useful tool for this purpose, provided the required assumptions are met.

Golding(27) describes the prediction of seasonal runoff for the 9.4 km^2 (3.6 mi^2) Marmot Creek experimental watershed located 40 km (25 mi) southeast of Banff, Alberta. The two snowmelt runoff periods are defined as May 1 to June 30 and May 1 to July 31. These runoffs are correlated with snow accumulation which is measured as: a) March snowpack as measured at each snow course; b) maximum pack at each snow course; and c) mean sub-basin water equivalent from the March grid sampling. May 1 to June 30 runoff has higher correlation with March snow-course measurements, than with either the maximum snow-course measurement or the mean snow water equivalent. The highest correlation with snow water equivalence is obtained for the May 1 to June 30 runoff. Golding feels that the poorer correlation for the period May 1 to July 31--although snowmelt is said to continue into July--is due to such factors as rainfall and evapotranspiration during the extra month. It is noteworthy that the correlation of the runoff period--May 1 to June 30--with March's snow course measurements yielded a "median correlation coefficient of 0.89 with 23 of the 32 correlations significant at the 95% level of probability."

The current water supply forecasting program of the California Department of Water Resources(58) exemplifies the importance of snow cover monitoring and subsequent runoff prediction. Fango et al(58) demonstrate a procedure for the updating of water supply forecasts during the period of snow melt utilizing snow covered area as a parameter. The two depicted watersheds in their study are the Kings and Kern. They report that the Kings watershed covers an area of approximately 4000 km^2 (1544 mi^2), and has an average annual runoff of 1.93 Gm^3 (68.26 BCF). This represents 19 inches (480 mm) of runoff, 74% of which occurs during the spring snowmelt period--defined April to July. The Kern watershed covers 5372 km^2 (2074 mi^2) and has an average annual runoff of 773 Mm^3 (27.3 BCF). This represents 0.145 m (5.7 in) of runoff, about 67% of which occurs during the April to July snowmelt period. Pertinent hydrometeorological information such as water equivalence, precipitation, and runoff records are developed into basin indices for the prediction of April to July runoff through the application of regression analysis. The indices used in the development of the regression models are: a) the snow pack index which is based upon the observed water equivalent at approximately 20 snow courses in each basin as of April 1; b) the April to June precipitation index, based on approximately six stations, as an indication of basinwide seasonal wetness; c) the April to June precipitation index based on observed precipitation occuring during the forecast period--observed precipitation data replace average precipitation figures as the snowmelt season progresses; d) the October to March runoff index which is felt relates both to basin wetness and volume of water not stored in the basin as a result of early season runoff; and e) the previous year's runoff index which is the volume of the antecedent April to July runoff, and is thought to represent carry-over from the previous runoff season. Standard stepwise regression techniques are employed to determine the order of entry of the predictors. Several alternative orders and combinations of the predictants are considered in an attempt to reduce the number of significant variables. The selection of the best model is based on minimizing the average and the standard deviation of the differences between the forecast and the actual runoff. The best model developed for the Kings watershed includes the predictor variables: a) the April 1 snowpack index; b) the October to March precipitation index; c) the previous year's April to July runoff; and d) the April to June precipitation index. The best model developed by the California Department of Water Resources for the Kern watershed include: a) the April 1 high elevation snowpack index multiplied by the

October to March precipitation index; b) the previous year's April to July runoff; c) the April to June precipitation index; d) the April 1 low elevation snowpack index; and e) the May 1 snowpack index which is unknown at the time of the forecast. Only 20 and 18 years of data are used in the calibration of the regression models for the Kings and Kern River watersheds, respectively. The coefficients of determination are 97.1 percent for the Kings and and 97.9 percent for the Kerns. The derived F-test values for the models are 161.3 and 156.0, repectively. Such high F-values acquired during model calibration are usually indicative of overfitting. That is, evidently a limited sample exists and the calibrated models contain several independent variables.

Simple regression analysis is implemented by Baker (6) to predict spring runoff for the 3315 km^2 (1280 mi^2) Cottonwood River watershed in south-western Minnesota. The objective of his study is to show that winter precipitation could be used as a predictive tool in spring runoff forecasting. He reasons that for regions where soils freeze, measurement of winter precipitation should provide an excellent means of predicting early spring runoff. He states that the proposed method takes advantage of the relationships between precipitation, soil temperature, and early spring runoff. He defines winter precipitation as the precipitation which occurs when the soil is frozen. The soil is considered frozen when the daily soil temperature at 0.051 m (2 in) remains equal to or lower than 0°C (32°F). It remains frozen until the minimum daily spring temperature at 0.305 m (12 in) rises above 0°C (32°F). Precipitation and soil temperature are measured at one site within the watershed. This precipitation constitutes the spring snowmelt runoff. The runoff period is defined as commencing on the first day in March or April when the river discharge equals twice the average daily February discharge, and terminating on April 30 of each year. From the seven years of data used in this study, it can be seen that the soil was considered unfrozen only once prior to the commencement of the spring flood. In 1962 the spring runoff period commenced March 26, however the winter precipitation was computed through to April 24-- the date the soil was considered thawed. A simple correlation coefficient of 0.99 is obtained between the winter precipitation and the runoff for the seven years of data. A comment by Riggs(60), concerning the regression model proposed by Sharp et al(62), is equally applicable for the proposed models of Baker(6). Riggs comments that "the computed reliability of the resulting regression equation is higher than the true reliability because of this inclusion of the same data on both sides of the equation."

Material covered thus far in this chapter has delt exclusively with the development of seasonal regression models. It should be mentioned that regression analysis is also implemented in the development of annual(21,53), as well as short term(1,34,81) models. Although the methodologies tend to be consistent with seasonal studies, the development and introduction of antecedent hydrometeorological variables for model construction are not analogous with the seasonal models, and thus, will not be further analyzed.

Thus far one has witnessed a review of regression analysis employed in seasonal water yield forecasting. The works cited range from early endevours to recent attempts. Table 1 presents a synopsis of the papers reviewed in this section. From the review of the literature concerning the application of regression analysis it is evident that no apparent logical procedure in model building is followed.

Therefore, the following three basic steps in mathematical modeling(50) will be introduced: 1) identification; 2) calibration; and 3) verification. It is evident that there is a need for the development of such a logical procedure when employing the technique of regression analysis in seasonal water yield forecasting. As well, the majority of papers failed to mention the basic premises in the development and application of their models. Practitioners of regression analysis, who did make some basic premises, found their premises to differ substantially with those of their colleagues. In addition to the need for the development of a logical procedure in regression analysis, one realizes that basic premises must as well be included and met in the development of a sound comprehensive procedure for regression model construction.

Table 1 : Synopsis of Seasonal Water Yield Forecasting

AUTHOR	DEPENDENT VARIABLE(S)	INDEPENDENT VARIABLES - CONSIDERED	INDEPENDENT VARIABLES - ACCEPTED	REMARKS
Finlayson(24) (Quebec)	spring flood volume	watershed's snow water equivalent, prec. between equivalent and start of freshet.	none	precipitation occuring during the freshet period showed a strong relationship with the spring flood volume.
Hopkins(33) (Maine)	spring flood volume	precipitation, snow survey, wind, % of possible sunshine, temperature, relative humidity	none	search for possible independent variables.
Cavadias(11) (Quebec)	spring flood volume	snow survey, precipitation, ground water conditions, wind, temperature	snow water equivalent, mean Feb streamflow as measure of groundwater conditions, precipitation, evapotranspiration.	included freshet prec. which is unknown at time of forecast.
Cavadias(12) (Quebec)	spring flood volume subdivision of flood period into nine periods	water equivalent of snow, Febuary runoff - indicator of groundwater conditions, runoff during the period previous to the one considered, cumulative degree days above 32°F during the period under consideration. precipitation during the forecast period, water budget variable, cumulative degree days above 32°F from 1st March to beginning of period.	depends on period	coefficients of the prediction equations were not supplied.
Cavadias and Brisebois(15) (Quebec)	April runoff, May runoff, June runoff, April and May runoff, May and June runoff, April May and June runoff.	precipitation before or during the flood, runoff before the flood, the degree days, influence of the groundwater conditions, and the water budget.	March degree days, water budget at March 31 March precipitation, April degree days, prec. - runoff (Jan and Feb), April runoff, May degree days, groundwater conditions, March runoff, March precipitation, April precipitation.	—
SC.S.(66) (Western US)	April - September runoff	water content of snow, precipitation, soil moisture, soil temperature, base flows, wind speed, solar radiation.	water content of snow, November 1 base flow index, April average precipitation	April avge precipitation required to predict April Sept runoff.
Hunter(37) (B.C.)	seasonal runoff for: March to September April to September May to September	summer and autumn precipitation, Sept through Oct precipitation, previous spring and summer runoff, base or late autumn flow, water content of snow, winter precipitation, winter streamflow, accumulated precipitation	not generally indicated	only those variables used in the prediction of the seasonal volumes were listed
Koelzer and Ford(45) (Colorado)	April 1 - July 31 runoff	July-Sept precipitation; Oct-Dec streamflow; precipitation periods - Oct-March, Oct-Dec, Oct-Jan; snow water content - April 1, Feb 1, with Feb to March increase in water content; temperature - Sept-Nov avge, Nov-Dec excess, Oct-Dec excess, Jan-March excess, Nov-March excess, Feb avge, March avge, Feb-March avge.	Oct to forecast date avge precipitation, water content of snow courses on forecast date, April 1 through June precipitation, Nov-Dec temperature excess, Oct-Dec precipitation.	considered independent variables occuring during the forecast period not listed. Excess implies the cumulative degree days of daily max temp. over a given period and above a given base temp.
Ford(26) (Alaska)	July-August runoff (1) July-September runoff (2)	Oct to June total precipitation, June mean of daily max temperatures, Jan-June precipitation	Oct to June total precipitation, June mean of daily maximum temperatures, Jan to June precipitation.	(1) represents the Eklutna Lake inflow, (2) represents the Long River inflow.
Sharp, et al(62) (Kansas)	annual, monthly and seasonal runoff	annual precipitation, monthly precipitation, accumulated monthly precipitation in excess of 5 inches per month, ground water in storage, monthly temperature, % of watershed in row crops/ terraced/ or pasture	several	multiple correlation and regression analysis was examined in light of evaluating parameters affecting water yield.

Table 1 : (continued)

AUTHOR	DEPENDENT VARIABLE(S)	INDEPENDENT VARIABLES - CONSIDERED	INDEPENDENT VARIABLE(S) - ACCEPTED	REMARKS
Golding(27) (Alberta)	May 1 to June 30 runoff May 1 to July 31 runoff	March snowpack, maximum snowpack, mean sub-basin water equivalent from the March grid sampling	March snowpack estimates	May 1 - July 31 results were "poorer" than those of May 1 - June 30.
Mango(38) (California)	April-July runoff of Kings River, April-July runoff of Kern	April 1 snowpack index, Oct-March precipitation, April-June precipitation, Oct-March runoff, previous year runoff	April 1 snowpack index, Oct-March precipitation, previous year April-July runoff, April-June precipitation, April 1 high elevation snowpack, April 1 low elevation snowpack, May 1 snowpack.	coefficients of the prediction equations were not supplied. Equations were based on 20 and 18 years of data respectively
Baker(6) (Minnesota)	spring runoff	"cold period" precipitation	"cold period" precipitation	"cold period" infers when the soil is frozen at specified depths. Coefficients were supplied for the ten watersheds studied.

Multivariate techniques have been proposed in the literature for the forecasting of spring water yield(13,52,65). One such technique is the traditional regression analysis of certain hydrometeorological variables with the spring water yield. This technique has been presented thus far in the literature review, and will be adopted for use in this study. A second proposed multivariate technique is principal-components regression. That is, a regression of the components of certain hydrometeorological variables with the spring water yield is, as well, performed. A review of the literature concerning the development and implementation in water resources of principal-component regression analysis is presented in the following section.

1.2.2 Principal-Components Regression

Factor analysis, developed mainly by psychologists, is considered a branch of the statistical sciences(31,46). The origins of factor analysis is generally ascribed to Charles Spearman(67), who developed a psychological theory involving a single general factor and a number of specific factors in 1904. However, a crucial article by Karl Pearson in 1901 brought forth "the method of principal axes"(56). Approximately 30 years later, Hotelling(35) provided the full development of this method, which he terms "principal-components analysis." These investigators prima caussa is the attainment of scientific parsimony or economy of description by the resolution of the set of independent variables, linearly, in terms of a smaller number of categories or factors. It is evident that a given matrix of correlations of the original variables can be factored in an infinite number of ways. However, a distinction between two objectives can be made immediately. They are the extraction of maximum variance or the 'best' reproduction of the observed correlations. The method proposed by Pearson in 1901 and later developed by Hotelling in 1933 meets the first objective.

This method of principal components, or component analysis, involves "the rotation of the coordinate axes to a new frame of reference in the total variance space--an orthogonal transformation wherein each of the n original variables is describable in terms of the n new principal components" (31). Analysis of the correlation matrix with ones in the diagonal leads to principal components. The analysis of the correlation with communalities--numbers other than ones in the diagonal--leads to principal factors. Component analysis has the property that each component, accounts, in turn, for a maximum amount of variance of the variables. That is, the first principal component is the linear combination of the original variables which contributes a maximum to their total variance, while the second principal component contributes a maximum to the residual variance, and so on until the total variance is analyzed. These derived components have the characteristic that they are mutually uncorrelated.

Several excellent books illustrate the technique of principal-component analysis(4,10,29,30,31,35,43,44,81). Kendall(43) exemplifies regression of a dependent variable based upon a principal component solution of the independent variables. Orthogonalization of the regression situation is attempted due to three problems existing in regression analysis. These problems are: a) the determination of the variables to be included in the model; b) the determination of the variables to be excluded from the model; and c) the alleviation of multicollinearity in the independent variables. [The utilization of principal components with regression analysis had not been previously indicated in the literature(43).] As well, Kendall does not dwell on the use of component analysis in standard regression theory, but indicates that there seems "to be a fruitful field of inquiry here."

Snyder(65) discusses the possibilities of regression on principal components of the independent variables in hydrologic studies. The inclusion of principal components is proposed for certain hydrologic applications where multiple regression of the original independent variables may produce unsatisfactory results. In the derivation of regression models, Snyder points out two of the premises of multiple regression analysis. The first assumption states that the independent variables are fixed variates and are measured without error. The second basic assumption states that the "so-called independent variables are not correlated with each other." Snyder discusses several examples of the second premise not being fulfilled in the literature. He then introduces the multivariate technique of principal components as a method to eliminate the violation of the second premise. An example problem is then analyzed by a) multiple regression of the original multicollinear independent variables, and by b) multiple regression of the principal components of the independent variables. The model based on the first technique yields an unsatisfactory solution due to the multicollinearity of the independent variables. In comparison, the second technique employing principal components gives the more satisfactory solution in terms of the rationality of the model coefficients. However, the gain in rationality of the coefficients is accompanied by a decrease in the multiple correlation coefficient. In effect, Snyder concludes that the technique of component analysis has "excellent possibilities...in hydrologic studies", and recommends their use for those problems for which multiple regression produces unsatisfactory results.

The paper by Snyder represents one of the first practical applications in the field of hydrology of the technique of principal components. He endevours to acquaint hydrologists with a previously developed alternative to regression analysis. Cavadias(13) and Pice(59), as well, discuss the possibilities of multivariate analysis and, in particular, the method of principal components with emphasis on practical applications. A detailed description of the mathematics involved in the method of component analysis is presented by Cavadias(13). He proffers principal components as a "possible improvement of the multiple regression approach that permits a more rational elimination of some independent variables in the case of collinearities."

Wallis(79) discusses and compares the effectiveness of six methods of analysis by solving by each method a problem using identical data of known functional relationship. Regression analysis and regression of principal components are two of the methods investigated. He contends, as does Snyder, that in numerous hydrologic studies two of the underlying assumptions of regression analysis are often violated. However, Wallis's second premise is that the residual errors of the model are independent and normally distibuted. If this second premise is not true, then the conventional F and t significance tests, associated with regression analysis, is placed in jeopardy. In addition to these two premises, he states that in the presence of multicollinear independent variables, the coefficients of the regression model become unstable. This results in the inability to understand their underlying functional relationship. Wallis lists three advantages of prediction equations based on regression of principal components over those based on multiple regression of the independent variables. They are: a) the coefficients of the model tend to be stable even if intercorrelations are high; b) the method does not capitalize on errors in the dependent variable, therefore it tends to yield more reliable results than regression when the prediction equations are used with different populations; and c) the rank of the independent variable's correlation matrix is determined by the number of positive eigenvalues.

Marsden and Davis(48) illustrate a practical application of principal-components regression. They develop a model to forecast summer water supply, for the Colombia River and its tributaries in Washington state, based on regression on principal components of winter mountain precipitation data. This forecast model supersedes the conventional model based on multiple regression analysis of the original measurements. Depending on the tributory, the number of available independent variables range from 17 to 21. The independent variables are comprised of several precipitation indeces from four of five weather stations for three seasons--fall, winter, and spring. The data also includes four or five snow survey locations, depending on the tributary under study. Each individual station is considered as an independent variable to the principal component analysis. The authors state that the resultant forecast models based on principal-components regression prove more stable than the conventional regression models. They conclude that the increased stability is due to the ability of the regression on principal components method to include all of the independent variables, rather than just a subset of them.

DeCoursey and Deal(20) present a study of some multivariate analytic techniques in hydrology. Regression analysis, discriminant analysis, and canonical correlation are illustrated using 37 hydrologic and 12 watershed characteristics of 90 runoff measuring stations in Oklahoma. To illustrate their first topic, conventional multiple regression, stepwise regression, and regression on principal components are compared in the development of mean annual flow prediction equations. They note that the most stable prediction equation is probably deterministic or rational in design. Conceptual hydrology models, such as the SSARR(77) and Stanford watershed models(17), represent the application of such predictive equations. However, the implementation of these hydrologic models is usually limited by the available data, and the justification for the time and expense required for model calibration. Thus, simpler multiple regression models provide an economic alternative. DeCoursey and Deal analyze several variations of multiple regression technique. They first generate a prediction equation based on 'conventional' multiple regression. This technique includes all the available independent variables in the calibrated model. They realize that historically this method predominantly yields poor predictive models due to variable interaction causing unrealistic coefficients. Their generated model does not disprove this realization. The second employed technique, generates a prediction model based on a stepwise variable selection. This method calculates the potential increase in the regression sum of squares for each variable. The variable that contributes the most to the regression sum of squares is selected. A regression model is then formulated which includes this variable and all variables previously selected in this manner. This iterative procedure is repeated for all remaining variables until the contribution of the most significant remaining variable does not prove significant as indicated by an F test. A prediction equation is then formulated based on the described stepwise procedure. Proceeding the development of the stepwise regression model, DeCoursey and Deal discuss, but do not apply, two other approaches to the stepwise technique. The first alternative entails variable addition in the same manner as previously discussed; however, when each new variable is performed on the significance of all the previously selected variables in the presence of the most recently selected variable. The second variation of the chosen stepwise procedure assumes that all variables are in the equation initially. Using the F test, the most insignificant variables are systematically removed. The authors note that these stepwise procedures have the advantage of retaining most of the correlation that exists between the dependent and independent variables with a minimum number of variables being retained in the prediction model.

Two different prediction models based on regression of principal components are then proposed by DeCoursey and Deal. These two models differ in the alternative criteria utilized for the deletion of principal components. The first model, based on the components having the largest eigenvalues, has the anticipated sign for most of the coefficients of the variables as compared with the conventional regression model. The second model, based on the components having the highest correlation with the dependent variable, has, as well, developed some illogical signs for the coefficients of the variables. The authors note, however, that for both cases these illogical signs are likely to develop because of the large number of variables that have been included in the analysis, some of which have no relation to the dependent variable. As well, the authors note that a large portion of the variance in the data matrix may not be associated with the dependent variable. This is evidenced in the second model which experiences a lower standard error of estimate and a higher coefficient of determination than the first model. The regression model, based on the previously described stepwise procedure, has a standard error of estimate lower than either models based on regression on principal components, and a coefficient of determination less than the second regression on principal components model. The authors contend, however, that the model based on the stepwise procedure does not reflect a great deal of stability when extrapolated to other data sets due to variable interaction. They claim that the models developed by multivariate procedures should show more stability than the common regression models when applied to new data sets. This results from the principal-components-regression models' calibration, which employs only statistically significant orthogonal components.

In general, DeCoursey and Deal conclude that they are dealing with a nonrational model, and that any one of the previously described methods of model construction would probably be as good as the next. The authors maintain, however, that the multivariate techniques have slight advantages over the conventional regression techniques. They indicate that the only disadvantages of the multivariate techniques is there are no known methods to calculate standard errors of the coefficients or to place confidence limits on the mean and individual observations.

DeCoursey and Deal propose two differnt prediction models based on regression of principal components. These two models differ in the alternative criteria utilized for deleting components. These alternative criteria represent a dichotomy. Massey(49) lists these two opposing criteria as: a) the deletion of components that are relatively unimportant as predictors of the original independent variables in the problem, that is the components having the smallest eigenvalues should be elliminated; and b) the deletion of components that are relatively unimportant as predictors of the dependent variable in the problem, that is the components having the smallest correlation with the dependent variable should be elliminated. He states that usually in purely exploratory studies the emphasis is on finding the correlates of the dependent variable rather than testing its relation to any particular structural concepts. However, his experience dictates that the components with large eigenvalues are the most likely to yield natural interpretations. Wallis(80) feels that a major disadvantage of the principal comonent technique lies in the inability to physically interpret components. However, Matalas and Reiher(51) state that the "principal comonents are mathematical devices and have no practical physical interpretation." Hotelling(36), as well, when approaching the criteria problem, points out that in general there is no reason why components that are important as far as the independent variables of a problem are concerned will be highly correlated with the dependent variable in a regression. Thus, the two alternative criteria are likely to lead to different results. Furthermore, Massey(49) comments that "it is easily shown that y [the dependent variable] need not be highly correlated with components having large eigenvalues in order for the exploratory powers of the complete principal components regression to be high."

Similarly, Darlington(18) notes that factors which account for very little variance in the predictor variables are highly useful in predicting the criterion provided the predictor variables are highly reliable. He proceeds to give an example of the above point, noting that when the variables being factored contain substantial error variance, error tends to be concentrated in the factors which account for the least variance. However, measurement errors on hydrologic variables are reported to be much smaller than those in typical psychometric studies(80). As well, Fiering(23), when using principal component analysis to transform a set of streamflow observations into an orthogonal set to provide a means of generating synthetic streamflow data, comments that all factors having positive eigenvalues are significant and should be included in the principal-components regression. Thus, it is shown that when one is concerned with the modeling of the dependent variable, that the second criteria should yield superior results.

Zuzel and Cox(82) attempt to use exploratory factor analysis and regression analysis to determine the effectiveness of wind, air temperature, vapor pressure, and net radiation in predicting daily snowmelt. The authors state that principal component analysis and regression analysis are used to determine which meteorological parameters are most efficient in evaluating snowmelt. Actually a rotation(39-42) of the principal components is employed in the determination of the effectiveness of the parameters, and not the original principal components.

The introduction to this section, on the regression of principal components, indicates that two distinct objectives exist when factoring a given matrix of correlations of the original variables. They are to either extract the maximum variance or to 'best' reproduce the observed correlations. The papers reviewed thus far utilize the method which meets the first objective, while Zuzel and Cox utilize a procedure to meet the second objective. Thus the authors do not attempt a regression on principal components, but rather utilize a multivariate technique in an attempt to determine the underlying structure of the phenomenae. To quantify the results indicated by their factor analysis, stepwise linear regressions are performed on all combinations of the meteorological variables--where melt is the dependent variable. The authors conclude that if only one meteorological variable is available for snowmelt prediction, then average temperature is the best predictor. However, snowmelt prediction can be significantly improved by using vapor pressure, net radiation, and wind rather than the temerature variable. They note that the partial correlation and the beta coefficients illustrate the unimportance of temperature when all the variables are included. A similar analysis to determine the effectiveness of wind, vapor pressure, and net radiation in predicting daily temperature may further illustrate the authors' a priori knowledge.

A report by Fogarasi and Mokievsky-Zubok(25) describes the first of two stages of their application of multivariate analysis on glacier-climatological data. The first stage includes the transformation of the multicollinear climatological variables into new orthogonal independent variables--principal components. The authors describe the characteristics and the interpretations of these components. Recommendations are given for the second stage of the analysis which is an application of multiple regression on principal components aimed at the estimation of daily glacier runoff. A total of 547 complete daily sets of data are available, each containg eleven predictor variables--precipitation, mean temperature, maximum temperature, minimum temperature, calculated mean temperature, melting degree day temperature, relative humidity, daily total sunshine, global radiation, mean daily cloudiness, and daily total wind run. From the variance-covariance matrix, the authors eliminate the mean, maximum, minimum, and calculated mean temperatures from further analysis. They feel these variables and melting degree day temperature are "similar variables but with different names." The relative humidity is considered unrepresentative and is thus not included in the analysis. Daily total sunshine is also omitted, but for an unexplained reason. The authors perform a principal component analysis on the remaining five independent variables. They then interpret the derived principal components, where the principale of interpolation is variables having high correations or loadings on a component help to identify the component. They utilize the following rules of thumb for component consideration: a) the components associated with the eigenvalues less than or equal to one; and b) each component which account for at least 5 percent of the total variance. The authors conclude by commenting that normalities of the variables is not an absolute necessity for principal components analysis, but when inferences are to be made which require normally distributed variables, then the assumption of normality cannot be relaxed. As well, the authors note that in multiple regression studies the least squares theory does not require normality, however, to make inferences based on the estimates requires normally distributed variables. The authors test the distributional assumption of normality of the dependent and independent variables, and conclude that at the 0.10 percent level of significance, only 'melting degree day' is normally distributed. The authors then suggest, that proceeding with normally distributed variables, daily snowmelt can be estimated by the multiple regression of the normalized runoff values on the independent principal components.

McCuen et al (52) introduce a study which: a) discusses the model selection criteria that should be used in selecting a water yield model from alternative models; b) summarizes the advantages and disadvantages of multiple regression, stepwise regression, principal-component regression, polynomial analysis using principal components, and the consrained pattern search analyses as methods of model calibration; and c) compares these methods using hydrologic data from the Upper Sevier River basin in Utah. Although McCuen et al utilize the term "selection" to infer some criterion of goodness of fit in model acceptance or rejection, it is commonly referred to in mathematical modeling as the "verification" step(50). While the correlation coefficient and the standard error of the estimate are frequently used when comparing models of seasonal water yield, McCuen et al contend that the following criteria may be more important in selecting one model from among several alternatives: a) the rationality of the regression coefficients; b) the randomness and distribution of the residual errors; and c) the correctness of the indicators of the relative importance of the predictor variables. These criteria are used to compare seasonal water yield models calibrated using the above mentioned techniques. The prediction equations use the April through July streamflow volume as the dependent variable. The independent variables available to the investigators include: a) the April 1 snow water equivalent of each of the basin's five operating snow courses; b) an arithmetic average of three of the five of a) above; c) one snow course's March 1 snow water equivalent; d) the October baseflow; and e) the December baseflow. From their paper, it appears that various independent variables are introduced to the different calibration techniques. The principal-components regression model contains only a) and d) of the above independent variables. The authors give no explanation in general as to why the models are calibrated using different groups and manipulations of the available independent variables. They conclude by noting that the alternative methods for model calibration are shown to be superior to the frequently employed multiple regression technique. They mention that principal-components regression represents one such alternative method, and should provide more reliable models due to its ability to control the effects of multicollinearity in model construction. The authors state that while statistical criteria are valuable, the reliability of empirically developed models shall improve only if hydrologic and rationality criteria are incorporated into the model evaluation phase of the analysis. Their study shows that hydrologic criteria proves more important than the statistical criteria for model evaluation. They feel their improvemnt to the model calibration and evaluation methodology classically employed in water resources, should lead to more reliable estimates during periods of extreme observations. And, in general, if a modeling methodology includes the calibration of such alternative models and the proposed evaluation criteria, improved water supply forecasts should be realizable. In summary, the method of principal components attains scientific parsimony or economy of description of the original set of independent variables. Component analysis has the property that each component, accounts, in turn, for a maximum amount of variance of the original set of variables. It is evident that two methods exist for determining the importance of each component. The first places emphasis on the relative importance of components as predictors of the original set of variables in the analysis. The second approach places relative importance on the components as predictors of the dependent variable. In general, there is no reason why components that are important, as far as the independent variables of a problem are concerned, should be highly correlated with the dependent variable in a regression. Thus, the second approach should be adopted when considering regression on principal components, provided the variables being factored do not contain substantial error variance. It should be remembered that principal components are mathematical functions which have no practical physical interpolation.

From this review of the literature concerning the application of regression on principal components, it is evident that no apparent logical procedure in model building has been followed. The three basic steps in mathematical modeling, as mentioned in the previous section, have not been adhered to by the authors. However, one paper(52) does yield insight to the model verification step by denoting importance to: a) the rationality of the regression coefficients; b) the randomness and the distribution of the residual errors; and c) the analysis of the indicators of the relative importance of the predictor variables. There is, as well, a need for the development of a logical model building procedure when employing the technique of regression on principal components.

1.3 OBJECTIVES

The literature review has shown that considerable interest is being displayed in the adoption of sophisticated techniques in the design and operation of water resources systems, resulting from increased energy awareness and costs. It is well realized that the development of reliable and accurate flow forecast models is prerequisite for such optimum water resources management.

Several studies have suggested that principal-component regression has excellent possibilities in hydrologic studies, but no attempts have been reported in applying it to forecasting. The prime objective of this study is to investigate the effectiveness of principal-component regression in the forecasting of spring water yield. This will be achieved through the development of regression and principal-component regression models for a multi-reservoir hydroelectric system in the Saquenay-Lac St-Jean region of Quebec.

Thus, it is proposed that two general classes of linear additive models will be analyzed, composed of: a) independent/causal variables, and b) components of the independent variables.

A general categorization of possible independent variables will be performed. This implies that presently employed independent variables in operation forecasts will be expanded to include more comprehensive hydrometeorological data.

In addition, a general model building scheme will be adopted to aid in model development. As part of the model building scheme, model verification will include:

1. a detailed residual analysis (randomness, distribution),
2. the adoption of split-sampling technique where possible,
3. the analysis of the rationality of regression coefficients, and
4. the analysis of the indicators of the relative importance of the independent/causal variables.

Chapter 2

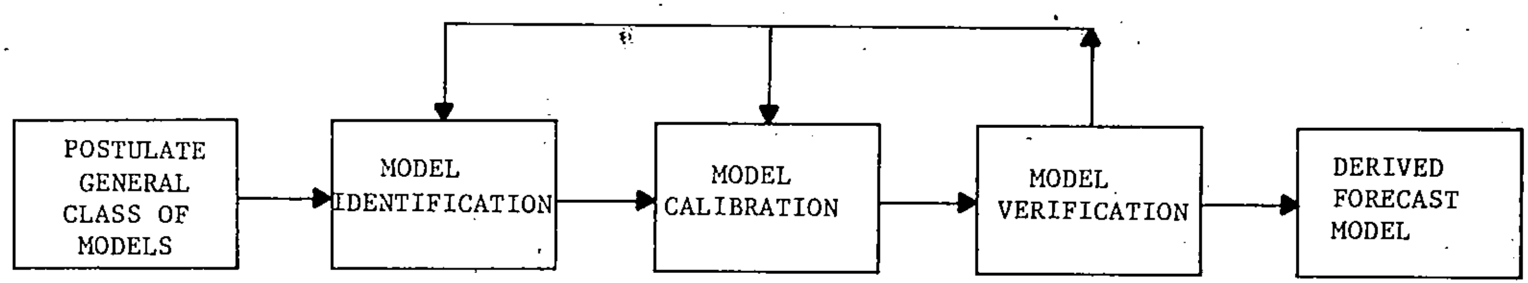
THEORETICAL DEVELOPMENTS

2.1 GENERAL DESCRIPTION

Regression analysis is used to develop two sets of prediction models used in the forecasting of spring season water yield. The two types of models are: a) regression of the original independent variables; and b) regression of the components of the independent variables. Relatioships between the dependent variable--the spring season water yield---and those independent variables for which data are available and which rational reasoning indicates are responsible for this flow, are developed. The independent variables are composed of hydrometeorological data which may be classified into three categories. The first category includes data prior to the start of the snow accumulation period that judgement might indicate as having an influence on the spring season water yield. The second category are those factors reflecting the influence of conditions during the period the snow pack is accumulating, and that affect the dependent variable. The third and final category are those hydrometeorological variables which exert their influence during the runoff period.

Analyses of the developed regression models are used to test the hypothesis that models based on regression on principal components are more stable and yield better predictions than models derived by regression of the original independent variables. The regression models are developed employing the logical procedure for model construction, as indicated in Figure 3

Figure 3: Stages to the Iterative Approach to Model Building



2.2 MODEL BUILDING SCHEME

The general model building scheme depicted in Figure 3 is used in the development of regression models. The identification step identifies a suitable class of models for the purposes at hand. As any type of class may be proposed, an appropriate form for the model is usually selected based upon experience, past information, and convenience. Once an appropriate parsimonious subclass of models has been proposed, estimates of the parameters of the tentatively entertained model are derived in the calibration phase of the model building scheme. At this step, efficient estimates of the parameters are obtained following a prescribed procedure. Once the model parameters have been estimated the model is said to have been calibrated. This fitted model is then subjected to diagnostic checks and tests of goodness-of-fit to determine the adequacy of the model and the efficiency of which the fitting process has made use of the data. If any inadequacy is found, the iterative cycle of identification, estimation, and verification is repeated until a suitable representation is found. Once a suitable model passes the verification step and is deemed adequate, the regression model is ready for its intended use(8).

2.2.1 Model Identification

The first basic step in mathematical modeling is that of model identification. Identification implies that the modeler has at his grasp a set of models that cover all aspects of the physical system of interest and has sufficient information to select the model that applies(50). It is assumed that seasonal water yield can be represented by a linear additive model of the form

$$Y = \beta_o + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p \qquad (2.1)$$

where Y is the dependent variable defined as the seasonal water yield, $X_1, X_2, \ldots, X_p$ can be either original independent variables or principal components of the independent variables, and $\beta_1, \beta_2, \ldots, \beta_p$ are unknown parameters. Two types of models are analyzed. The first is a linear additive model of the independent variables, while the second is a linear additive model of the components of the independent variables.

It is realized that the identification step of model building solely indicates the kind of representational model which is worthy of further investigation for a given set of data. The tentatively proposed form of the model is then introduced to the calibration step of the model building scheme. However, before the parameter estimation or model calibration is reviewed, it is first necessary to briefly illustrate how principal components are determined for a given set of data.

2.2.1.1 Principal Components

In multivariate analysis when p variables are identified, they usually are intercorrelated. That is, some of the information contained in one variable is contained, as well, in the p-1 remaining variables. Principal components analysis transforms these p original variables into p uncorrelated or orthogonal components. This is achieved through a linear transformation of the type

$$\zeta_i = \sum_{j=1}^{p} a_{ij} X_j \qquad ;j = 1,2,\ldots,p \tag{2.2}$$

where one has p variates $X_1 \ldots X_p$, each observed on n individuals. The observations may be arrayed in a matrix:

$$\underline{X} = \begin{vmatrix} X_{11} & X_{12} & \cdots & X_{1n} \\ \vdots & \vdots & & \vdots \\ X_{p1} & X_{p2} & \cdots & X_{pn} \end{vmatrix}.$$

where the X_{ij} is the j-th observation on the i-th variate. In matrix notation, such a linear transformations of the original variables may be written

$$\underline{Z} = \underline{X}\ \underline{A} \tag{2.3}$$

where $\underline{X}$ is as previously defined, $\underline{Z}$ is an nxp matrix of n values for each of p components, and $\underline{A}$ is a pxp matrix of coefficients defining the linear transformation.

As the original motive for the development of principal component analysis is the attainment of scientific parsimony, the matrix $\underline{Z}$ is constructed such that each component,

ζ_j, explains the maximum amount of the variance of the original variates $\underline{X}$ left unexplained by the first j-1 components. (The originators of this technique had hoped that the first q components would explain most of the system variance.) If one can express the data, $\underline{X}$, in terms of fewer than p of the ζ's, then a reduction in the dimensions of the problem is effected. The first step in the development of a prediction model based on principal components is the derivation of these principal components of the independent variables. The derived components are then introduced to the regression analysis as 'independent' variables.

The original independent variables, $X_1 \ldots X_p$, usually have differing units, thus the principal components are derived from standardized data, such that

$$x_{ij} = (X_{ij} - \bar{X}_j)/S_j$$

where X_{ij} is the i-th observation on the j-th variable, $\bar{X}_j$ is the mean of the j-th variable, S_j is the standard deviation of the j-th variable, and x_{ij} is the standardized i-th observation of the j-th variable. Once standardized data is obtained, $\underline{X}$, the next step in the derivation of the principal components of a group of independent variables is the computation of the correlation matrix, $\underline{R}$, where

$$R = \begin{vmatrix} r_{11} & r_{12} \cdots & r_{1p} \\ \vdots & \vdots & \vdots \\ r_{p1} & r_{p2} \cdots & r_{pp} \end{vmatrix}$$

The above correlation matrix has p characteristic roots or eigenvalues: $\lambda_1 \ldots \lambda_p$. They are determined as the roots of the following determinant equation of degree p:

$$\begin{vmatrix} (r_{11}-\lambda) & r_{12} \cdots & r_{1p} \\ \vdots & \vdots & \vdots \\ r_{p1} & \cdots & (r_{pp}-\lambda) \end{vmatrix} = 0 \qquad (2.5)$$

All the roots of equation 2.5 are real and positive (43,44). Choosing the largest root, λ_1, of equation 2.5 yields the component having the maximum variance. The p values of $a_{11}, a_{12}, \ldots, a_{1p}$ may be found by solving the linear system

$$\begin{aligned} (r_{11}-\lambda)a_{11} + r_{12}a_{12} + \ldots + r_{1p}a_{1p} &= 0 \\ \vdots \qquad\qquad & \vdots \\ r_{p1}a_{11} + r_{p2}a_{12} + \ldots + (r_{pp}-\lambda) &= 0 \end{aligned} \qquad (2.6)$$

where

$$a_{11}^2 + a_{12}^2 + \ldots + a_{1p}^2 = 1$$

The solution of equation (2.6) yields the first component, given by

$$\zeta_1 = a_{11}x_1 + \ldots + a_{1p}x_p = \sum_{j=1}^{p} a_{1j}x_j \qquad (2.2)$$

The remaining principal components, $\zeta_2, \zeta_3, \ldots, \zeta_p$, are derived in a similar manner utilizing the characteristic roots, $\lambda_2, \lambda_3 \ldots \lambda_p$ (13,30,43,44). And as previously mentioned, these derived components may be introduced to the regression analysis as 'independent' variables.

2.2.1.2 Model Calibration

A tentative form of the regression model is obtained from the identification step of the model building scheme. The two seperate groups of models formulated from this first step, are both postulated from the general class of models as being linear additive. Equation 2.1 indicates the general form of the model.

The first tentative group of models is based on the original independent variables, while the second group consists of a linear additive model based on the principal components of the independent variables. Having proposed the tentative form of the models to be analyzed, one may proceed with the calibration step of the mathematical modeling. Calibration infers the efficient estimation of the parameters of the tentative models. If the first tentative form of the regression model includes k independent variables, then 2^k regression models can be calibrated. As k becomes relatively large, it is evident that the number of models to be calibrated becomes prohibitive. To avoid this dilemma, the calibration and verification steps of the mathematical modeling form an iterative procedure to acquire a suitable regression model from the tentative group of possible models. Figure 3 shows the iterative process incurred in mathematical modeling.

The calibration procedure employed is that of least squares. That is, the objective function of the calibration is to minimize the sum of squares of deviations from the proposed models. However, there is no unique statistical procedure for selecting the suitable regression equation. Several procedures are available which use a few model verification techniques such as goodness-of-fit statistics and tests of hypothesis to acquire a suitable regression model. Draper and Smith(22) discuss several procedures for the determination of a suitable model. They include: a) all possible regressions; b) backward elimination; c) forward elimination; d) stepwise regression; and e) two variations of the four previous methods. These authors recommend the use of stepwise regression over the previously mentioned alternative methods. They note that the stepwise procedure does not necessarily select the absolute best model, but usually selects an acceptable one. The selection procedure employed in this thesis is the stepwise method.

2.2.1.3 Stepwise Regression

The procedure of all possible regressions involves the fitting of every possible regression equation and may become cumbersome when the number of independent variables considered is large(22,30). If k independent variables are contained in the tentatively proposed model, then 2^k equations will be generated and analyzed to determine the best model. The best model is selected according to the criterion of the verification step of the mathematical modeling procedure.

As k becomes large it is obvious that considerable computation and analyses are required to derive a suitable model from this procedure. Draper and Smith(22) note that "in general the analysis of all regressions is quite unwarranted." They also state that "the amount of computer time used is wasteful and the sheer physical effort of examining all the computer printouts is enormous when more than a few variables are being examined. Some sort of selection procedure which shortens this task is preferable."

One technique which is widely used and does shorten the selection procedure is that of stepwise regression. Draper and Smith(22) believe this technique to be the best of the previously mentioned variable selection procedures, and recommends its use.

The stepwise procedure is very similar to the forward selection procedure in that it consists of building the regression model one variable at a time, by adding at each step the variable that explains the largest amount of the remaining unexplained variation. The forward selection procedure determines the order of insertion by using the partial correlation coefficient as a measure of the variables not yet in the equation. Variables are inserted in turn until the model is deemed unsatisfactory. However, the forward selection procedure makes no effort to explore the effect that the introduction of a new variable may have on the role played by a variable which entered at an earlier step. This apparent weakness is overcome by the adoption of the stepwise procedure. The partial F criterion for each variable in the regression at every step of the calculation is evaluated. This provides a measure of contribution made by each variable as though it had been the most recent variable entered, regardless of the step at which the variable is introduced to the model. That is, the first variable to enter the model is the variable displaying the highest simple correlation with the dependent variable. The second variable selected is the one which explains the greatest portion of the variance of the dependent variable not accounted for by the first variable. The first variable is, in turn, tested for significance. If the variable is deemed insignificant, it is discarded. This iterative procedure is continued until a stable suitable model is derived. This implies that all the independent variables selected are significant, and those which are not selected are insignificant in terms of explaining the variance of the dependent variable.

One notes that with the identification of a tentative linear additive regression model consisting of k independent varaibles, that 2^k regression models may be calibrated by the method of least squares. A great reduction in the computation required to derive a suitable model is realized with the adoption of a cursory verification procedure. Using the technique of stepwise regression for the selection and rejection of independent variables, subjects models to goodness-of-fit tests and tests of hypotheses. These tests actually form part of the verification step in the analyses of the proposed models. Tests employed in the stepwise procedure, as well as several other tests and techniques that are employed in the verification of the proposed model, are discussed latter in this thesis.

2.2.2 Model Verification

The adequacy of a proposed calibrated model is determined in the verification step of the mathematical modeling. This phase of the modeling procedure attempts to determine in what manner a proposed calibrated model is inadequate, thus possibly leading to appropriate modifications. The employement of a reasonably comprehensive verification procedure in mathematical modeling should enable the modeler to maintain less doubts concerning the implementation of a derived model. However, Box and Jenkins(8) note that judgement should be used in the verification step as "statistical tests can discredit models which could nevertheless be entirely adequate for the purpose at hand."

2.2.2.1 Assumptions of the Regression Model

Thus far there has been no mention of distributional assumptions in the development of the regression models. Basic assumptions will now be put forth for the regression model:

$$Y_i = \beta_o + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p + \varepsilon_i \quad ; i = 1,2,\ldots,n \qquad (2.1)$$

The first assumption is that the error ϵ_i is a random variable with a mean of zero and a constant variance. This may be written as

$$E(\varepsilon_i) = 0 \quad , \quad V(\varepsilon_i) = \sigma^2$$

The second assumption states that ϵ_i and ϵ_j are not correlated for i not equal to j, so that

$$COV(\varepsilon_i, \varepsilon_j) = 0$$

Thus Y_i and Y_j for i not equal to j, are, as well, not correlated. The final assumption is that the error, ϵ_i, is a normally distributed random variable with a mean of zero and variance constant. That is:

$$\varepsilon_i \sim N(0,\sigma^2)$$

This final assumption implies that ϵ_i is independent, as well as not being serially correlated (22,38).

2.2.2.2 The Standard Error of the Regression Model

The proposed regression model, as depicted by equation 2.1, is:

$$Y = \beta_o + \beta_1 X_1 + \beta_p X_p + \varepsilon_i \qquad (2.1.b)$$

The standard error of the regression equation, σ, may be estimated by S, where

$$S^2 = \Sigma \varepsilon_i^2 / (n-p) = \Sigma (Y_i - \hat{Y}_i)^2 / (n-p) \qquad (2.7)$$

and Y_i is the observed value of the dependent variable, $\hat{Y}_i$ is the estimated value of the dependent variable, n is the number of observations, and p is the number of independent variables in the model.

2.2.2.3 The Multiple Coefficient of Determination

An expression for the multiple coefficient of determination, R^2, may be represented by the following formula:

$$R^2 = 1 - (n-p) S^2 / ((n-1) S_Y^2) \qquad (2.8)$$

where S_Y^2 is the variance of the dependent variable; and n, p, and S^2 are as previously defined. The multiple correlation coefficient is defined as the positive square root of the multiple coefficient of determination. The multiple coefficient of determination represents the variation in Y explained by the combined linear influence of the independent variables divided by the total variation in Y. The value of R^2 may vary from 0 to 1 and is analogous to the simple coefficient of determination in the bivariate model.

2.2.2.4 Confidence Intervals for the Standard Error of the Estimate

Having placed certain assumptions on the development of the regression models, one may proceed to place confidence intervals on the standard error of the estimate, and perform various tests of hypotheses. Confidence intervals may be placed on the standard error of the estimate, σ, by noting that $(n-p)S^2/\sigma^2$ has a chi-square distribution. Thus, the lower and upper confidence limits on σ^2 are(30), respectively

$$L = \frac{(n-p)S^2}{\chi^2_{1-\alpha/2,n-p}}$$

$$U = \frac{(n-p)S^2}{\chi^2_{\alpha/2,n-p}} \qquad (2.9)$$

2.2.2.5 Inferences on the Regression Model's Coefficients

The assumptions which are crucial for the employment of the calibration technique of least squares are: a) the expected mean value of the residuals is zero; b) the variance of the residuals is constant; c) the independent variables are fixed numbers; and d) the number of observations exceeds the number of independent variables. However, to derive significance tests and place confidence intervals on the estimated parameters of the least squares calibration, $\hat{\beta}_i$, one must assume that in addition to the above assumptions, the residuals are independently and normally distributed(38).

Inferences now may be made concerning the variances of the parameters $\beta_0 \ldots \beta_p$ of the model, where $\widehat{\beta}_0 \ldots \widehat{\beta}_p$ depicts the estimated parameters of the calibration procedure. If $\underline{X}$ is an nxp matrix of independent variables, and $\underline{X}^T$ represents the transpose of $\underline{X}$, then one may proceed to compute $\underline{C}=\underline{X}^T\underline{X}$ and $\underline{C}^{-1}=(\underline{X}^T\underline{X})^{-1}$. The variance of the estimated parameter $\widehat{\beta}_i$ may be found from

$$\mathrm{VAR}(\hat{\beta}_i) = \sigma^2_{\hat{\beta}_i} = C^{-1}_{ii}\sigma^2 \qquad (2.10)$$

where σ^2 may be estimated by S^2, as previously defined. As the estimated parameter $\widehat{\beta}_i$ is independently and normally distributed with mean β_i and variance $C^{-1}_{ii}\sigma^2$, one may compute the confidence intervals on the estimated parameter (22,30) as:

$$L_{\beta_i} = \hat{\beta}_i - t_{1-\alpha/2,n-p} S_{\hat{\beta}_i}$$

$$U_{\beta_i} = \hat{\beta}_i + t_{1-\alpha/2,n-p} S_{\hat{\beta}_i} \qquad (2.11)$$

were L_{β_i} represents the lower bound and U_{β_i} is the upper bound.

Having computed the confidence interval of a given parameter $\widehat{\beta}_i$, one may proceed to test the null hypothseis that $\widehat{\beta}_i$ is equal to a hypothetical value of β_i, represented by β_{10}, against the alternative hypothesis that $\widehat{\beta}_i$ is not equal to β_{10}. The hypothetical value usually assigned to β_{10} is zero. Thus a test of $H_o: \widehat{\beta}_i=0$ versus $H_1: \widehat{\beta}_i \neq 0$ is in reality a testing of the hypothesis that the i-th independent variable is not contributing significantly to explaining the variation of the dependent variable(30). To test the hypothesis H_o: $\widehat{\beta}_i=0$, one need simply compute the test statistic

$$t = \frac{\hat{\beta}_i - \beta_{10}}{S_{\hat{\beta}_i}} \qquad (2.12)$$

where $\widehat{\beta}_i$, β_{10}, and $S_{\widehat{\beta}_i}$ are as previously defined. The hypothesis H_o is rejected if the absolute value of the computed t is greater than a tabulated value of t at the $(1-\alpha/2)$ level of significance having (n-p) degrees of freedom(22,30). One should remember, however, that if a t-test is used to reject a coefficient, it is done on the basis of a residual variance calculated by its inclusion. Kendall(44) notes that this test becomes more questionable if several variables are rejected at the same step.

The goodness-of-fit of the regression model in general may be assessed through the implementation of a commonly used hypothesis testing procedure(68). That is, the procedure is to test the null hypothesis that the multiple correlation is zero in the population from which the sample is drawn. The overall null hypothesis H_o:R=0, is equivalent to $H_o: \beta_1 = \beta_2 = \ldots = \beta_p = 0$ versus H_1: at least one or more of the β's is not zero. If the null hypothesis is rejected, then one may conclude that one or more of the estimated parameters is different than zero. The test statistic employed for this general test(68) is

$$F = \frac{R^2/(p-1)}{(1-R^2)(n-p)} = \frac{SS_{REG}/(p-1)}{SS_{RES}/(n-p)} \quad (2.13)$$

where SS_{REG} is the sum of squares explained by the entire regression model, SS_{RES} is the residual or unexplained sum of squares, and n and p are as previously defined. One rejects H_0 if F exceeds the tabular $F_{1-\alpha;p-1;n-p}$(30).

One may note that the above F test does not indicate the precise estimated parameters which are not zero. The previously described t-test may be employed in the case where only very few parameters may be shown to be zero. However, the $\hat{\beta}_i$'s are not independent and thus the t test looses credibility as the number of parameters equal to zero are tested simultaneously. In such cases, the hypothesis that k of the independent variables are not contributing significantly to the explanation of the linear variation in the dependent variable, may be tested. The hypothesis that the k last variables are not contributing significantly, may be represented as $H_0: \beta_{p-k+1} = \beta_{p-k+2} = \ldots = \beta_p = 0$ versus H_1: at least one of these β's is not zero. The test statistic employed in this reduced model test is given(30) by

$$F = \frac{(Q_2-Q_3)/k}{Q_1/(n-p)} \quad (2.14)$$

where Q_1 is defined as the residual sum of squares on the full model with n-p degrees of freedom, Q_2 is the sum of squares due to regression on the full model with p-1 degrees of freedom, Q_3 is the sum of squares due to regression on the reduced model with p-k-1 degrees of freedom, and n, p, and k are as previously defined. H_0 is rejected if the computed F exceeds the tabulated $F_{1-\alpha;k;n-p}$. The rejection of the null hypothesis implies that at least one of the k variables is explaining a significant portion of the variation of the dependent variable.

2.2.2.6 Confidence Intervals on the Regression Model

To place confidence intervals on a point Y_h on the regression line, one must first compute the variance of Y_h, such that

$$Y_h = \underline{X}_h \underline{\beta} \tag{2.15}$$

where $\underline{X}_h$ is the point in p dimensional space (a 1 times p matrix), and $\underline{\hat{\beta}}$ is a p times 1 vector of estimates of the regression parameters $\underline{\beta}$. The variance is given by Haan(30) and Draper and Smith(22) to be

$$VAR(\hat{Y}_h) = \sigma^2 \underline{X}_h (\underline{X}^T \underline{X})^{-1} \underline{X}_h^T \tag{2.16}$$

where $\hat{Y}_h$ is an estimate of Y_h, and σ^2 may be estimated by S^2 which is as previously defined. One may then proceed to compute the confidence limits of the regression model by use of the following equation

$$\text{LOWER} = \underline{X}_h\hat{\beta} - t_{1-\alpha/2,n-p}(\text{VAR}(\hat{Y}_h))^{1/2}$$

$$\text{UPPER} = \underline{X}_h\hat{\beta} + t_{1-\alpha/2,n-p}(\text{VAR}(\hat{Y}_h))^{1/2} \qquad (2.17)$$

where all the terms are as previously defined. If one desires to estimate the confidence intervals of a predicted value of $\underline{Y}_h$, one need simply replace the $\text{var}(\hat{Y}_h)$ term in equation (2.17) with the variance of an individual predicted value of $\underline{Y}_h$ given by $\sigma^2(1+\underline{X}_h(X^T\underline{X})^{-1}X_h^T)$ (22,30).

2.2.2.7 Rationality of the Estimated Model's Parameters

Thus far only statistical goodness-of-fit criteria have been introduced in the verification of proposed models. McCuen et al (52) propose that current modeling procedures, which exclusively employ goodness-of-fit criteria in model verification, can yield more reliable models if certain alternative criteria are adopted. They suggest that the methodology of model verification should include the determination of the rationality of the model's estimated parameters, as well af an examination of the model's residuals to determine model inadequacy.

The adoption of the rationality of the parameters as a verification criteria attempts to incorporate hydrologic rationality into the development of seasonal water yield models. The rationality of the parameters may be examined in the light of its sign and magnitude. The standardized partial regression coefficient is employed to determine the relative importance of the independent variables in relation to all the independent variables which are included in the regression model. The basic formula which is used for the computation of the partial regression coefficient or partial correlation coefficient (68) is

$$r_{ij \cdot k} = \frac{r_{ij} - (r_{ik} r_{jk})}{(1 - r_{ik}^2)^{1/2} (1 - r_{jk}^2)^{1/2}} \qquad (2.18)$$

where k is the control variable or variable whose effect on the dependent variable has been linearly removed, r is the simple correlation, while i and j represent the dependent variable and independent variables. This formula may be extended to include more than one control variable. A modification is required when more than two independent variables are included in the model. The extension of this formula to include n+1 control variables is achieved by replacing the simple correlation coefficients on the right side of equation (2.18) with the n-th order partial coefficients (68). One should note that there is no effect on the derived partial correlation coefficients by the order in which control variables are added to the computation. The rationality of the parameters is determined by the relative magnitude of each parameter's partial with respect to the intuitive importance of the dependent variable in the rational determination of the dependent variable.

Hawley et al (32) state that the rationality of the sign of the models' parameters may be determined simply by noting the sign of the simple correlation of the respective independent variables with the dependent variable. If the sign of the simple correlation coefficient and that of the model's parameter in question differ, then that parameter is deemed irrational. The authors state that such irrationality is due to the "statistical manipulations in the regression technique."

2.2.2.8 Analysis of Residuals

It is proposed that the analysis of the regression model's residuals be adopted in the verification of the calibrated model. Residuals are usually employed exclusively in summary statistics concerning the goodness-of-fit of the model. The analysis of the average magnitude of these residuals is used as a basis of the presented summary statistics, such as the standard error of the estimate and multiple coefficient of determination. However, a direct examination of the model's residuals may indicate a possible violation of the underlying assumptions concerning the errors which are put forth in the adoption of the regression analysis technique (5,22,68). These underlying assumptions are that the errors a) be independent, b) have zero mean, c) have a constant variance, and d) follow a normal distribution. The logic underlying the examination of the residuals is if the proposed model is correct, then the model's residuals should not violate the above underlying assumptions.

Visual examination of the residuals is made possible by the use of various plotting schemes. Draper and Smith(22) advocate the use of visual examination in lieu of numeric examination, as visual plots tend to be more informative. They note that any violations of underlying assumptions of the residuals that require corrective action will, almost without a doubt, be revealed by the residual plots.

The visual examination of residuals is concerned predominantly with the visisble patterns of the residuals when arrayed in a scattergram. It is suggested that the residuals be plotted in every logical manner possible for a given situation. Various methods of plotting the residuals are a) overall, b) in time sequence when the order is known, c) against the fitted or predicted values of the dependent variable, and d) against each independent variable(30,68). The overall plotting implies that the residuals are plotted in one dimension. This plot enables one to visualize how the residuals are distributed about the mean--zero. Such an analysis may enable one to conclude that the residuals appear to be of a normal distribution with a mean of zero. Having assumed that the residuals are independently and normally distributed with a mean of zero and a constant variance, σ, one may verify that the unit normal deviate form of the residual, e_i, appears to follow these assumptions. The unit normal deviate of e_i is defined(22) as e_i/S_1, where

$$S_1 = \left(\sum_{i=1}^{n} \frac{(e_i - \bar{e})^2}{(n - p)} \right) = \sum_{i=1}^{n} \frac{e_i^2}{(n - p)} \quad (2.19)$$

It follows that one would expect approximately 95 percent of the e_i/S_1 to lie within the 95 percent limits of the $t_{(n-P)}$ distribution.

One of the most important features, when making a direct examination of a scattergram, is the overall patern of the points. Figure 4 indicates four general patterns which could possibly emerge from the plotting of the residuals against a variable. The patterns of Figure 4 (b), (c), and (d) show an abnormality of the model that requires corrective action. The pattern of Figure 4 (b) indicates that the variance is not constant, while pattern (c) is indicative that a linear term is missing. The pattern represented by (d) indicates that a quadratic, and possibly linear terms, should be included in the model. The constant band pattern of (a) indicates that the model does not appear to have abnormalities (22,68). If the model exhibits a pattern similar to that of Figure 4, then one could conclude that the assumptions concerning the model's residuals do not appear in jeopardy. That is, on the basis of the data analyzed, there is no reason to indicate that the underlying assumptions have been violated, thus negating the need for corrective action.

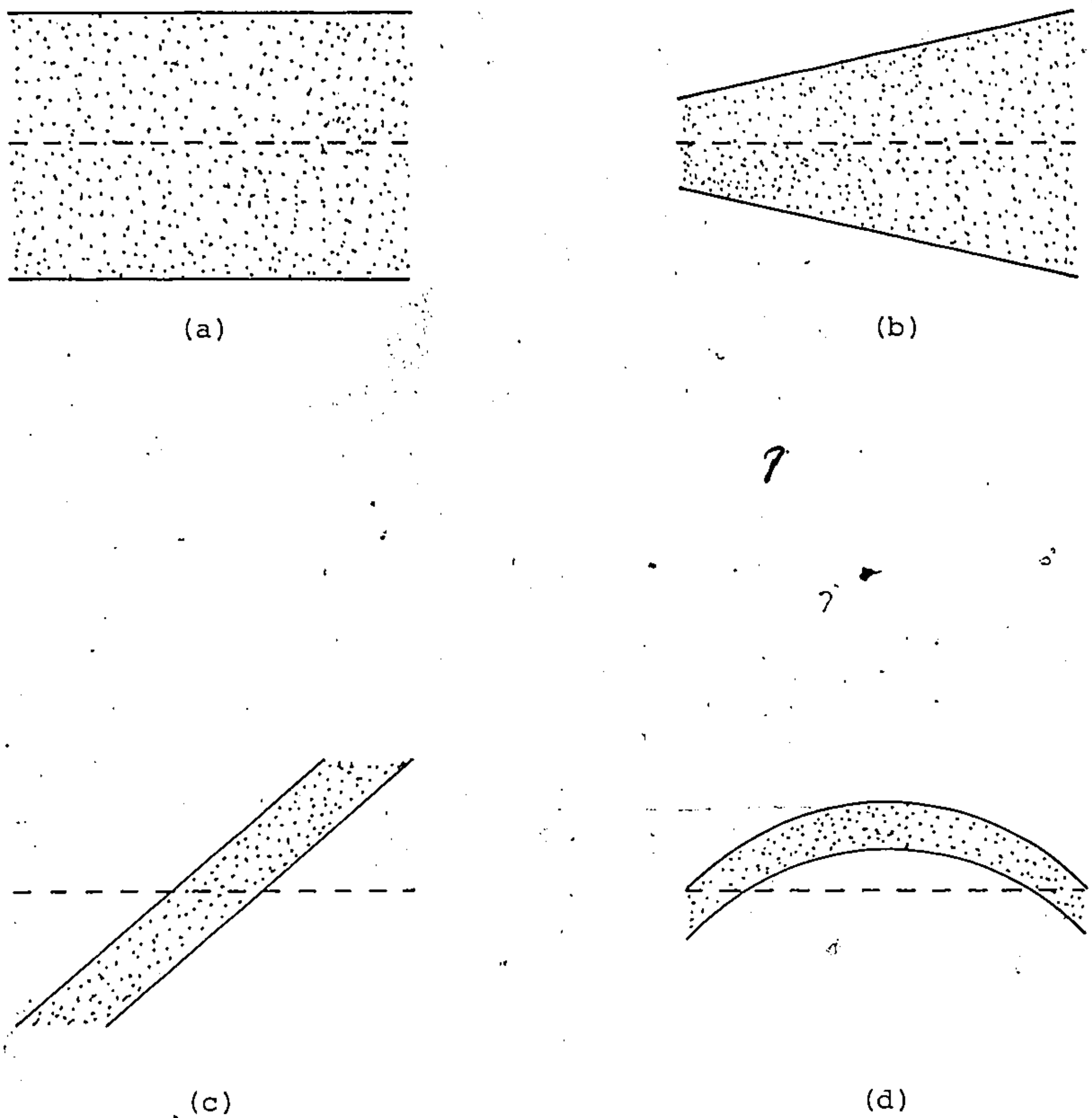


Figure 4 : Four general patterns which occur when plotting residuals(22,68).

As well, one may investigate the goodness-of-fit of the residuals to follow a normal distribution by the adoption of the nonparametric Kolmogorov-Smirnov test(30,61,81). The Kolmogorov-Smirnov statistic, D, is computed as the maximum deviation of the sample's distribution function F(x) from the specified hypothetical continuous distribution function P(x), such that

$$D = \max \left| F(x) - P(x) \right| \tag{2.20}$$

The goodness-of-fit is determined by the comparison of the computed D with a tabulated value D_t for a given confidence level. If the computed D is greater than D_t, then one may state that the observed frequency distribution departs significantly from a hypothesized frequency distribution. The sole assumption of this technique is that the data are continuously distributed. As well, Yevjevich(81) notes that the absolute maximum difference, D, is likely to be smaller than the true D of the population, as the parameters of F(x) are estimated from the sample data. He indicates this bias to be unknown and suggests the adoption of a critical D "somewhat smaller" than that suggested by the tabulated critical values of the Kolmogorov-Smirnov statistic. However, Lilliefors(47) lists a table of critical values of D which take into account that certain parameters of the distribution have been estimated from the sample.

Further nonparametric tests may be employed in the determination of the randomness of the residuals of the regression model. The first proposed test analyzes the residuals for tendancies to cluster in order of magnitudes. This is performed by classifying the observations into a few broad categories. The number of runs in each category is counted, where a run is defined as a group of consecutive observations of any category. The population mean of the residuals is given by

$$M_1 = \frac{n(n + 1) - \sum_{i=1}^{m} n_i^2}{n} \tag{2.21}$$

where n is the total number of observations, n_i is the number of observations in the i-th category, and m is the chosen number of categories. The population's standard deviation, σ_1, is given by the equation

$$\sigma_1 = \left[\sum_{i=1}^{m} n_i^2 \left(\sum_{i=1}^{m} n_i^2 + n(n + 1)\right) - 2n \sum_{i=1}^{m} n_i^3 - n^3 \right] - (n^2(n - 1)) \tag{2.22}$$

The standard normal variate is given by

$$Z = \frac{(G_1 + 0.5 - M_1)}{\sigma_1} \tag{2.23}$$

where G_1 is the number of runs.

A second test which may be employed, shows if the directions of the movements of the residuals tend to cluster. In this test the number of runs up and down, G_2, are counted. The population mean is given by the formula

$$M_2 = \frac{2n - 1}{3} \qquad (2.24)$$

and the standard deviation, σ_2, is represented by

$$\sigma_2 = \left(\frac{16n - 29}{90}\right)^{1/2} \qquad (2.25)$$

The standard normal variate on which tests may be performed is

$$Z_2 = (G_2 + 0.5 - M_2)/\sigma_2 \qquad (2.26)$$

where all the terms are as previously defined.

A third test of randomness checks for predominance of upward or downward trends in the direction of the residuals. This test is performed by counting the number of positive and negative residuals. The number of residuals being positive or negative should not be significantly different in a random sample. The population mean is given by

$$M_3 = (n - 1)/2 \qquad (2.27)$$

while the standard deviation, σ_3, is given by

$$\sigma_3 = ((n + 1)/12)^{1/2} \qquad (2.28)$$

The standard normal variate may be obtained from the formula

$$Z_3 = [M_3 - (S_2 + 0.5)]/\sigma_3 \qquad (2.29)$$

where S_2 is the smaller number of either positive or negative residuals, and the remainder of the terms are as previously defined. Cavadias and Brennan(14) note that the sampling distributions for a relatively large number of observations for the three previously mentioned randomness tests is well approximated by a normal distribution. The normal probabilities corresponding to the deviations of the number in the run from the expected value indicate that the probability of this value occurring in a series of random, independent observations.

2.2.2.9 Split Sampling

Split sampling is often considered to be one of the better procedures for the verification of proposed models. Goodness-of-fit statistics, based on the data employed in the calibration of the regression model, give an indication of the model's precision during the calibration, but not for the model's precision during actual prediction. Computation of the goodness-of-fit statistics and the analysis of residuals from the regression model based on data not included in the calibration of the model, provides an additional device for the verification of predictive models(30). Shrader et al(63) state that a data base independent of the data employed in the model's calibration should be employed in the verification of regression models. They note that in general, split sampling represents the sole method of properly testing a model.

Shrader et al(63) advocate the use of a random selection process to develop subsamples to be used in the calibration and verification of regression models. However, they note that for small sample sizes random selection may yield subsamples whose characteristics are not similar to those of the entire sample. For the split sampling technique to be valid, it is required that the derived subsamples exhibit somewhat similar characteristics. Thus, systematic random sampling is proposed in the case of small samples. They note that

> a valid subsample may be selected by identifying one or more important characteristics, seperating the data base into groups that are based on similar values of these characteristics, and randomly selecting an equal number of observations from each group. This systematic-random selection process will ensure that the sample is random and that the model will be tested using a subsample that is representative of the data used in calibration.

Chapter 3

EXPERIMENTAL DESIGN

3.1 THE PHYSICAL SYSTEM

The Saquenay-Lac St-Jean watershed lies mainly in the Grenville subprovince of the Canadian Shield. This region is characterized by rolling mountains having low relief, usually lying between 300 m (984 ft) to 1000 m (3280 ft) above mean sea level. In the northwestern region of the watershed, one finds the divide between the Superior and Grenville subprovinces of the Canadian Shield. The Superior is, as well, characterized by low relief lands containing much muskeg.

The waterways of the Saquenay-Lac St-Jean region are extensively developed for the production of hydroelectric power. Alcan Smelters and Chemicals Limited operates a multi-reservoir system consisting of six powerhouses having a total installed turbine capacity of 2687 Mw. There are three major reservoirs with a total live storage capacity of 11.61 Gm^3 (409.9 BCF). Figure 5 shows the hydroelectric system existing in the region, while Table 2 lists the powerhouse data of the structures shown in Figure 5 Table 3 lists the live storage capacities of the three major reservoirs located in the total watershed. The total watershed covers an area of 73,100 km^2 (28,225 mi^2), ranging from approximately 47°30' to 52°15' latitude north and 70°15' to 74°30' longitude west. The location and shape of the watershed is shown in Figure 6.

The creation of an accurate and reliable forecasting scheme is prerequisite for optimum water resources management in hydroelectric production(76). The prediction of water yield for the "very important" spring period greatly affects the operation of the system for the remaining portion of the year(64). In the operation of the system, the spring planning period is defined as commencing April 1 and ending June 15(73). It is, thus, desirable to obtain an accurate and reliable forecast for the spring season water yield as defined by the above dates for the operation of the system.

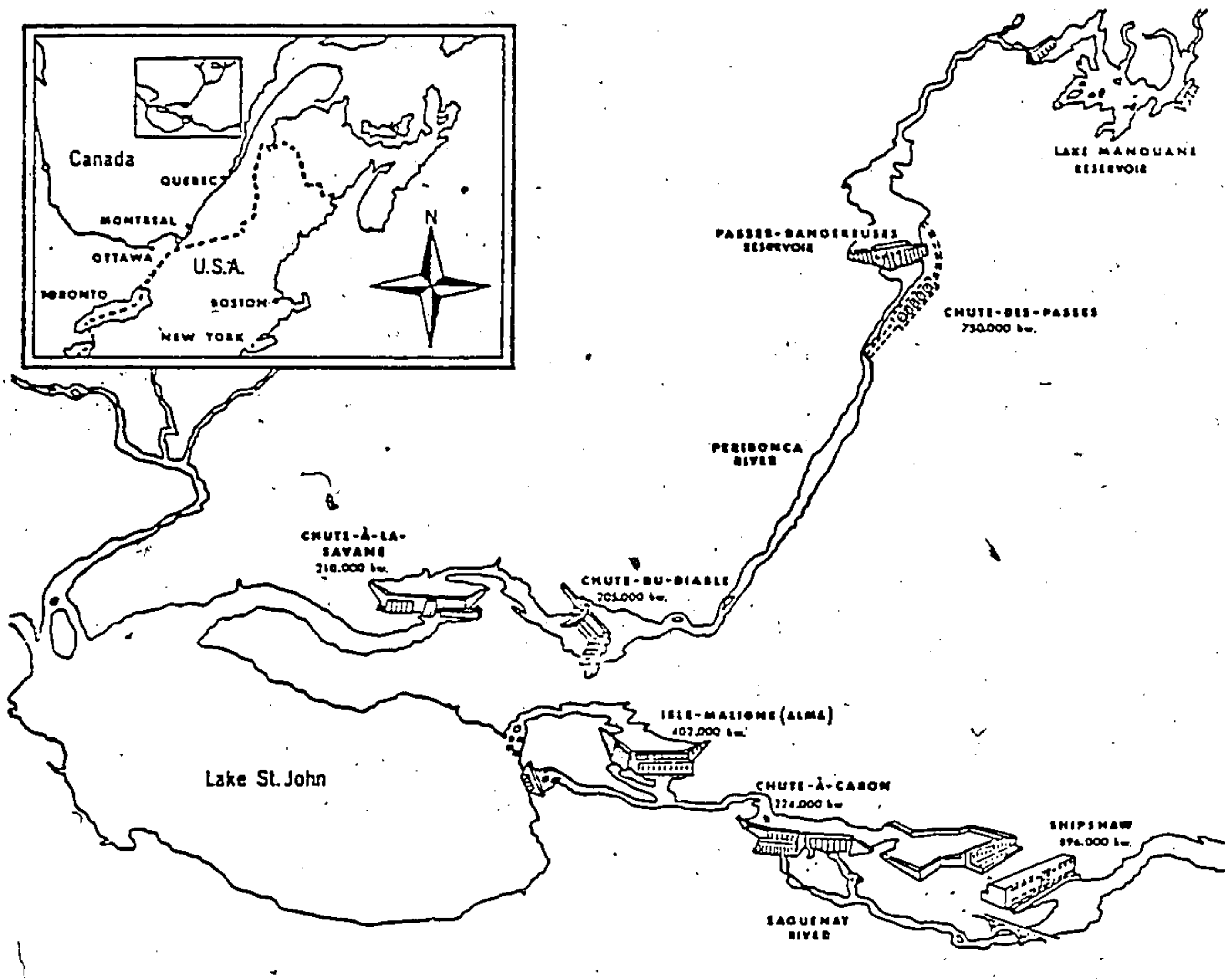


Figure 5 : The Alcan Saquenay-Lac St-Jean hydroelectric system(76).

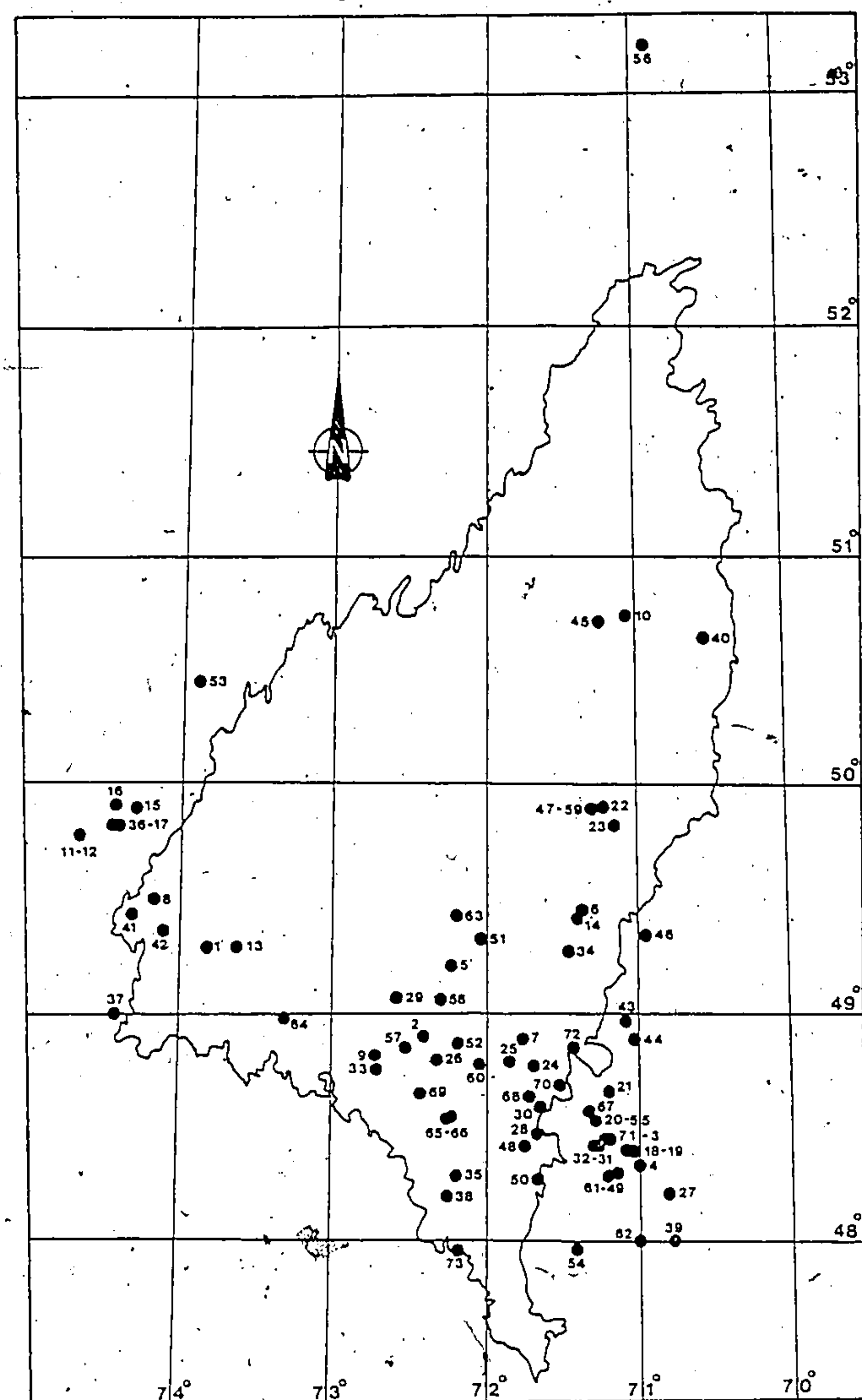


Figure 6 : Map of total Lac St-Jean watershed indicating the 73 meteorological stations as listed in Table 4.

Table 2 : Powerhouse Data

Powerhouse	Year completed	Average net head Meters	Feet	Installed turbine capacity MW	Number of turbines
Chute-des-Passes	1960	143.3 -195.1	470 -640	750.	5
Chute-du-Diable	1951	33.5	110	205.	5
Chute-a-la-Savane	1953	33.5	110	210.	5
Isle Maligne	1925	33.5	110	402.	12
Shipshaw	1943	64.0	210	896.	12
Chute-a-Caron	1930	48.8	160	224.	4
TOTAL				2 687.	43

Table 3 : Reservoir Data

Reservoir	Live Storage Capacity meters^3 x 10^9	Live Storage Capacity feet^3 x 10^9
Lac Manouan	2.55	90.
Passes Dangereuses	5.10	180.
Lac St-Jean	3.96	140.
TOTAL	11.61	410.

3.1.1 Alcan's Current Forecast Methodology

The forecasting of future water yield is based on antecedent hydrometeorological conditions employing the technique of stepwise regression. The forecasting is facilitated by use of an interactive program. This interactive nature of the forecasting system purportedly "makes the continual updating of the forecast equations much faster and more convenient than it otherwise would be"(71). It is realized that the factors to be considered for the forecasting of water yield are a) the dependent variable, b) the independent variables or input to the forecast, and c) the forecasting technique. For the Alcan system, the dependent variables considered include uncontrolled inflows to the system conditioned on an initial date and a final date, and uncontrolled inflow conditioned on an initial inflow rate and final date(70,71,73). However, as the spring period generation planning studies span April 1 to June 15, an important definition of spring season water yield represents the total volume of uncontrolled inflow observed during this period(70,71,73). Thompstone et al(73) explains that "a generation plan can be put into effect on April 1 when March runoff and results of the end of March snow survey are known." Forecasts of the spring freshet volume for this generation plan are based on March's snow survey and uncontrolled inflow data. The watershed has a mean spring flood volume for 1955 through 1976 of 18.24 Gm^3 (644 BCF). The standard error of estimate for the forecast model based on the end of March's snow cover and March's inflow is 3.23 Gm^3 (114 BCF) (70,73). The actual models relating the independent variables with the spring freshet is not shown. However, Thompstone(73) does state that "the best forecast found so far for the spring runoff was:

$$\hat{Y} = -115.2 + 44.8(X_1) + 11.3(X_2)$$

where $\hat{Y}$= volume of spring runoff, 1 April- 14 June, in BCF, X =end of February snow cover in inches of water equivalent [and] X =volume of March runoff in BCF.". The standard error of estimate for this model is 3.17 Gm^3 (112 BCF). It is reported that both proposed models' coefficients, as well as the multiple correlation coefficients, could not be rejected at the 95 percent level of significance (70,73). The models reflect 1955 through 1976 data. Model verification and diagnostic checking include only the significance testing of the models' coefficients and the multiple correlation coefficients.

3.1.2 Development of a New Forecast Methodology

Adoption of a logical procedure for model construction is imperative for the further developments of forecast models. The iterative approach to model construction, as described in Chapter 2, will be employed in this study. The general-model building scheme includes a) identification, b) calibration, and c) verification. It is assumed that the seasonal water yield can be represented by a linear additive model. Two general classes of such models are proposed for analyses. The first is a linear additive model of the independent variables, while the second is a linear additive model of the components of the independent variables. The technique of least squares is employed in the estimation of the model parameters in the calibration stage of the model building process. Proposed fitted models are subjected to diagnostic checks and tests of goodness-of-fit to determine the adequacy of each model and the efficiency of which the fitting process makes use of the data. Distributional assumptions regarding the residuals of the calibrated models are tested, when possible. As well, split sampling techniques may be employed to aid in establishing model sufficiency. If any inadequacy is found, the iterative cycle of identification, estimation, and verification is repeated until a suitable representation is found.

The factors previously expounded upon for the forecasting of water yield for the watershed in question are a) the dependent variable to be forecasted, b) the independent variable or input to the forecast, and c) the forecasting technique (70,71,73). Numerous definitions for the dependent variable are presently employed (70,71,73). However, the sole definition of the dependent variable to be adopted in this study is defined as the uncontrolled inflow volume of the Saquenay-Lac St-Jean watershed occurring from April 1 to June 15, inclusive. It is realized that this period represents a convenient description of the spring flood season, and an important generation planning time period for the operation of the hydroelectric system. Previously documented forecast models are based exclusively on snow cover in equivalent inches of water and March's uncontrolled inflows(70,73). Additional independent variables are investigated, herein, for possible inclusion in the forecast models. The independent variables are composed of hydrometeorological data which may be classified into three categories. The first category includes data prior to the start of the snow accumulation period that judgement might indicate as having an influence on the spring season water yield. The second category are those factors reflecting the influence of conditions during the period which the snowpack is accumulating, and that affect the dependent variable. The third and final category are those variables which exert their influence during the runoff period. It is obvious that models to be employed in forecasts must be based on antecedent conditions. However, the development of additional models which include causation factors occurring during the forecast period provide empirical relatioships illustrating accurately their quantitative effects. Such models should prove benificial to the operational management of hydrosystems.

In general, the new methodology includes:

1. the adoption of a general model building scheme;
2. the general categorization of possible independent variables, which implies that presently employed independent variables in operational forecasts are to be expanded to include more comprehensive hydrometeorological data;
3. the inclusion of detailed residual analysis and split sampling technique as part of the model building scheme; and
4. the proposal of two general classes of linear additive models to include a) independent variables, and b) components of the independent variables.

3.2 THE HYDROMETEOROLOGICAL DATA

The independent variables to be introduced to the proposed forecasting technique may be roughly classified into three categories, as proposed by Koelzer and Ford(45). The first category is composed of antecedent factors reflecting watershed and hydrometeorological conditions prior to the snow accumulation period. The second, is constituted to reflect the watershed and hydrometeorological conditions during the period of time when the snow pack is primarily accumulating. The third, and final category, is structured to reflect conditions subsequent to the second category. That is, the third grouping reflects the conditions during the snowmelt and runoff periods--the spring season runoff period. Undoubtedly, the lines of demarcation establishing these three categories in nature cannot be clearly established, as their dates change from year to year. The snow accumulation period is assumed to start on December 1, based on the general knowledge that the snow is usually retained for winter storage in the snowpack. The end of the snow accumulation period is March 31, as the spring flood period is defined as occurring from April 1 to June 15, inclusive. Thus the general time-structuring of the three categories are a) June 16 to November 30, b) December 1 to March 31, and c) April 1 to June 15.

The hydrometeorological variables acquired for inclusion into the three categories are: a) total monthly precipitation; b) average monthly temperature; c) monthly uncontrolled inflow; and d) end of January, February, and March average snow cover in water equivalence. Mean monthly total precipitation and mean monthly temperature data for the watershed are computed by the arithmetic averaging of meteorological stations' data listed in the Federal Government publication 'Monthly Record'. The meteorological stations used in the computation of the meteorological monthly means for the watershed are listed in Table 4, while Figure 6 shows the location of the meteorological stations listed in Table 4. Appendices A.1 and A.2 contain the mean monthly temperature data from January 1934 through July 1977, respectively. Appendix A.3 lists the average monthly uncontrolled inflow for the watershed from 1913 to 1977, inclusive. Average watershed snow cover for the snow surveys from 1955 to 1979, inclusive, are contained in Appendix A.4. The average watershed snow cover is obtained by the arithmetic averaging of all reporting snow courses operated by Alcan Smelters and Chemicals Limited in the Saquenay-Lac St-Jean region(72).

TABLE 4: METEOROLOGICAL STATIONS USED IN THE COMPUTATION OF MEAN MONTHLY TEMPERATURES AND MEAN MONTHLY TOTAL PRECIPITATION DATA FOR THE TOTAL WATERSHED

NO.	METEOROLOGICAL STATION	MINISTERE DES RICHESSES NATURELLES DU QUEBEC ALPHA NUMERIC IDENTIFICATION		GEOGRAPHICAL COORDINATES LATITUDE NORTH (DEG.-MIN.)		LONGITUDE WEST (DEG.-MIN.)		MONTHLY RECORD TOTAL PRECIPITATION DATA FROM	TO	MEAN DAILY TEMPERATURE DATA FROM	TO
1	AIGREMONT	706	0070	49	18	73	51	1973	1977	1973	1977
2	ALBANEL	706	0080	48	53	72	27	1934	1977	1934	1977
3	ARVIDA	706	0320	48	26	71	10	1934	1977	1934	1977
4	BAGOTVILLE A	706	0400	48	20	71	00	1942	1977	1942	1977
5	BARRIERE 3 RAT RIVER ROAD	706	0479	49	13	72	15	1963	1969	-	-
6	BARRIERE ETIENNICHE	706	0M7A	49	27	71	23	1973	1975	-	-
7	BARRIERE MILOT	706	ODPC	48	54	71	46	1973	1977	-	-
8	BARRIERE NORD CHIBOUGAMAU	706	0474	49	30	74	10	1963	1973	-	-
9	BARRIERE SUD LA DORE	706	047H	48	49	72	44	1963	1975	1973	1975
10	BONNARD	706	0825	50	44	71	03	1962	1977	1962	1977
11	CHAPAIS	709	1295	49	47	74	52	1962	1971	1962	1971
12	CHAPAIS 2	709	1305	49	47	74	52	1963	1977	1963	1977
13	CHEMIN CHIBOUGAMAU (M 75)	706	1368	49	17	73	38	1965	1967	-	-
14	CHEMIN CHUTE-DES-PASSES (M 47)	706	1372	49	25	71	24	1967	1975	1967	1974
15	CHIBOUGAMAU	709	1400	49	54	74	18	1936	1951	1936	1951
16	CHIBOUGAMAU	709	1400	49	54	74	27	1960	1975	1960	1975
17	CHIBOUGAMAU A	709	1401	49	49	74	25	1971	1977	1971	1977
18	CHICOUTIMI	706	1440	48	25	71	05	1934	1977	1934	1977
19	CHICOUTIMI U	706	1442	48	25	71	03	1976	1977	1976	1977
20	CHUTE-A-MURDOCK	706	1480	48	31	71	15	1934	1957	1934	1958
21	CHUTE-AUX-GALETS	706	1520	48	39	71	12	1934	1962	1934	1963
22	CHUTE-DES-PASSES	706	1541	49	54	71	15	1957	1976	1960	1977
23	CHUTE-DES-PASSES 2	706	1542	49	49	71	09	1973	1977	1973	1975
24	CHUTE-DU-DIABLE	706	1560	48	45	71	42	1951	1976	1951	1977
25	CHUTE SAVANE	706	1475	48	47	71	51	1951	1974	1951	1974
26	DOLBEAU	706	2020	48	48	72	20	1934	1938	1934	1938
27	FERLAND	706	2368	48	12	70	50	1971	1977	1971	1977
28	HEBERTVILLE	706	3040	48	27	71	40	1951	1958	1951	1958
29	HEMON	706	3090	49	04	72	36	1963	1977	1963	1977
30	ISLE-MALIGNE	706	3320	48	35	71	38	1934	1977	1934	1977
31	JONQUIERE	706	3370	48	24	71	16	1963	1975	1963	1975
32	KENOGAMI	706	3400	48	25	71	15	1934	1972	1934	1972
33	LA DORE	706	CP09	48	46	72	43	1975	1977	1975	1977
34	LAC ALEX	706	CEE5	49	15	71	28	1965	1966	1965	1966
35	LAC BOUCHETTE	706	3560	48	16	72	12	1949	1977	1949	1977
36	LAC CACHE	709	3585	49	49	74	26	1954	1973	1951	1973
37	LAC COOPER	709	1920	49	00	74	25	1952	1962	1952	1961

TABLE 4 (CONTINUED)

NO.	METEOROLOGICAL STATION	MINISTERE DES RICHESSES NATURELLES DU QUEBEC ALPHA NUMERIC IDENTIFICATION		GEOGRAPHICAL COORDINATES LATITUDE NORTH (DEG.-MIN.)		LONGITUDE WEST (DEG.-MIN.)		MONTHLY RECORD TOTAL PRECIPITATION DATA FROM	TO	MEAN DAILY TEMPERATURE DATA FROM	TO
38	LAC DES COMMISSAIRES	706	3610	48	12	72	15	1954	1975	1967	1975
39	LAC HA!HA!	706	3647	48	00	70	47	1964	1970	1964	1970
40	LAC MANOUAN	706	4620	50	38	70	32	1942	1961	1942	1961
41	LAC NICAUBA	709	3680	49	25	74	20	1952	1960	1952	1960
42	LAC NICAUBA S	706	CFHO	49	22	74	07	1973	1974	1973	1974
43	LAC ONATCHIWAY	706	3683	48	59	71	04	1934	1967	1934	1966
44	LAC ONATCHIWAY	706	3683	48	54	71	02	1973	1977	-	-
45	LAC ONISTAGAN	706	EGCG	50	43	71	17	1944	1945	1944	1945
46	LAC PAMOUSCACHIOU	706	3684	49	17	70	59	1973	1976	-	-
47	LAC PERIBONCA DAM*	-	-	49	53	71	16	1942	1943	1942	1943
48	LAC STE-CROIX	706	3690	48	25	71	45	1958	1977	1958	1977
49	LATERRIERE	706	4180	48	18	71	08	1963	1977	1963	1977
50	MESY	706	4890	48	16	71	41	1963	1977	1963	1977
51	MISTASSIBI	706	4993	49	19	72	02	1973	1975	1973	1975
52	MISTASSINI	706	4998	48	52	72	12	1963	1977	1963	1977
53	MISTASSINI POST	709	5000	50	25	73	53	1934	1977	1934	1977
54	MONT APICA	706	5100	47	58	71	25	1963	1977	1963	1977
55	MURDOCK WILSON	706	5383	48	31	71	15	1959	1961	1959	1961
56	NICHEQUON	709	5480	53	12	70	54	1942	1977	1942	1977
57	NORMANDIN	706	5640	48	51	72	32	1936	1977	1936	1977
58	NOTRE-DAME-DE-LORETTE	706	5667	49	04	72	19	1977	1977	1977	1977
59	PASSES DANGEREUSES DAM	706	5840	49	53	71	16	1942	1961	1942	1961
60	PERIBONCA	706	5960	48	46	72	04	1934	1977	1934	1977
61	PORTAGE-DES-ROCHES	706	6080	48	18	71	13	1934	1977	1934	1977
62	RIVIERE A MARS	706	6550	48	00	71	00	1963	1963	1963	1963
63	RIVIERE AUX RATS	706	6583	49	25	72	12	1963	1974	1963	1974
64	RIVIERE CHICOUBICHE	706	FFAO	48	59	73	18	1973	1973	1973	1973
65	ROBERVAL A	706	6685	48	31	72	16	1939	1977	1957	1977
66	ROBERVAL NORD	706	6688	48	32	72	14	1934	1966	1934	1967
67	ST-AMBROISE	706	6820	48	33	71	20	1954	1977	1954	1977
68	ST-COEUR-DE-MARIE	706	7060	48	38	71	43	1957	1977	1957	1977
69	ST-FELICIEN	706	7240	48	39	72	27	1938	1953	1938	1953
70	ST-LEON-DE-LABRECQUE	706	7460	48	40	71	31	1963	1977	1963	1977
71	SHIPSHAW	706	8160	48	27	71	13	1943	1977	1943	1977
72	TCHITAGAMA	706	CP07	48	51	71	28	1934	1935	1934	1935
73	VAN BRUYSSEL	707	HF96	47	57	72	10	1971	1972	1971	1972

* LAC PERIBONCA DAM is believed to have been located in the construction camp of PASSES DANGEREUSES DAM (59).

3.3 NUMERICAL PROCEDURE

Two general classes of linear additive models are proposed for analyses. The first is a linear additive model of the independent variable, while the second is a linear additive model of the components of the independent variables. The independent variables, to be introduced to the model building scheme, are classified into three general categories. In each category, various hydrometeorological variables may be compounded to reflect desired watershed conditions for the period in question. Thus, two factors become apparent. The first is the determination of the best form of the independent variables for introduction to the model building scheme, while the second concerns the form of the model.

Figure 7 shows a schematic diagram of the computer programs which are used in the numerical computations. The mainline computer program accesses the hydrometeorological data bank and selects the desired variables, and their form. The program then proceeds to the iterative stages of model calibration and verification. During this stage, multiple stepwise regression (75) is utilized to calibrate models from the proposed two general classes of models. An analysis of the residuals of the calibrated models is then performed. This analysis includes several plots and nonparametric tests to aid in acceptance or detection of model inadequacy. If a calibrated model's residuals show no apparent inconsistancies, then the model is subjected to a split sampling analysis. This aids, as well, in the determination of coefficient stability. In addition, one may compute the coefficient of determination and the standard error of the estimate for the split sample. Relative changes in these values from those obtained in original model calibration may seriously implicate calibrated models for use in forecasting. Proceeding this step, if diagnostic checking has not rejected the proposed model, then it may be deemed acceptable for implementation.

Appendix B lists and documents the computer programs depicted in Figure 7.

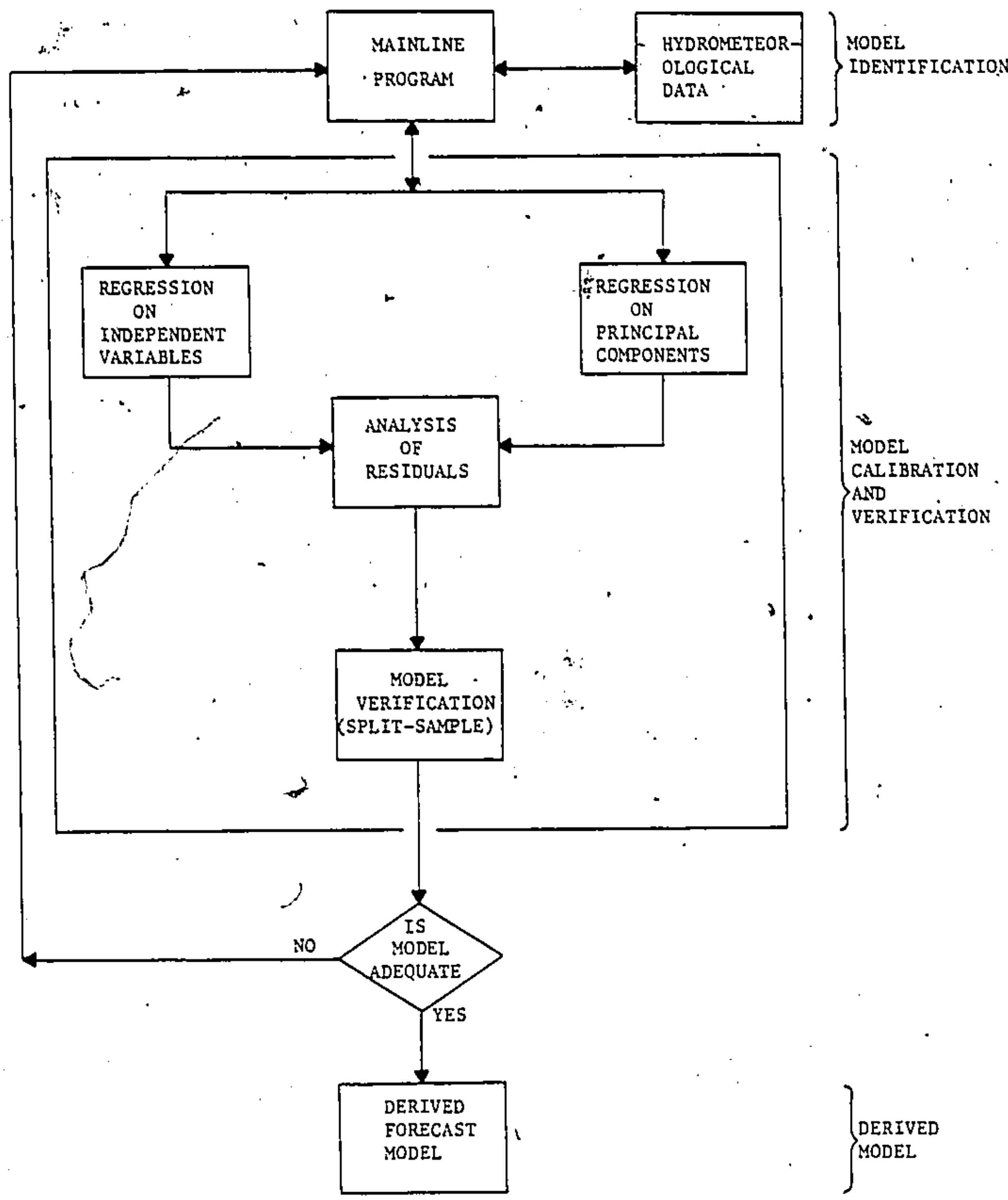


Figure 7 : Schematic of stages in the iterative approach to building forecast models.

Chapter 4

RESULTS AND DISCUSSION

4.1 THE INDEPENDENT VARIABLES

Hydrometeorological data prior to the start of the snow accumulation period, during the snow accumulation period, and subsequent to the snow accumulation period are used in the development of forecast models. Table 5 lists the hydrometeorological variables which may be allocated to desired categories to reflect watershed conditions.

4.1.1 Variables Antecedent to Snow Accumulation Period

Hydroclimatic factors prior to the snow accumulation period may have influence on the proceeding volume of the spring flood. From the hydrometeorological variables listed in Table 5, possible hydroclimatic factors based on total monthly precipitation, average monthly temperature, and monthly uncontrolled inflows may be proposed. The precipitation and temperature of July through September may be described as representing summer effects on the watershed(16,45), while the uncontrolled inflows of October and November may reflect fall conditions.

Table 6 lists the simple correlation coefficients of each proposed climatological factor with the spring flood volume.

Table 5 :Available Hydrometeorological Variables

Variable Number	Variable Description	
1	Avg. Monthly Temp. (F.)	- July
2		August
3		Sept.
4		October
5		Novem.
6		Decem.
7		January
8		February
9		March
10		April
11		May
12	Total Monthly Prec. (inches)	- July
13		August
14		Sept.
15		October
16		Novem.
17		Decem.
18		January
19		February
20		March
21		April
22		May
23	Uncontrolled Inflow (c.f.s.)	- July
24		August
25		Sept.
26		October
27		Novem.
28		Decem.
29		January
30		February
31		March
32	Snow Cover (equiv. cm of water)	- January
33		February
34		March

The null hypothesis that there is no association between the dependent and independent variables cannot be rejected at the 5 percent level of significance. Thus, the proposed hydrometeorological variables occurring prior to the snow accumulation period do not appear to have a significant effect on the volume of the spring flood. It is, as well, possible that these factors are simply not of the correct form for correlation with the spring flood volume. It is interesting to note that summer temperature and precipitation prove to be the most significant factors of this period.

Table 6 :Correlation of Factors Prior to the Snow Accumulation Period with the Spring Season Flood Volume (1955-1976)

Variable		Correlation Coefficient
Precipitation	July	0.194
	August	0.087
	July and August	0.265
	Sept.	-0.085
	October	0.020
	Nov.	0.077
	Oct. and Nov.	0.064
	July through Nov.	0.145
Temperature	July	0.376
	August	0.120
	July and August	0.300
	September	-0.059
	Aug. through Sept.	0.011
	October	-0.114
	November	-0.196
	Oct. and Nov.	-0.192
	July through Nov.	0.016
Uncontrolled Inflows	- October	-0.175
	- November	-0.094
	- Oct. and Nov.	-0.170

4.1.2 Variables During the Snow Accumulation Period

Certain hydrometeorological factors which occur during the snow accumulation period are known to have a pronounced effect on the spring freshet. One of these factors is the precipitation occurring between December and March. Precipitation which has occurred during the snow accumulation period may be introduced in an alternate form--snow cover in water equivalence. Watershed snow cover is available since 1955 for the end of January, February, and March. However, watershed monthly precipitation has been acquired to 1934 to aid in model verification by means of split sampling.

Winter baseflow is used to represent the state of the watershed's groundwater(11,12,15). Winter months' uncontrolled inflows are used for this purpose. Winter temperatures are, as well, analyzed to determine the effect of winter severity on the volume of the spring freshet.

Table 7 lists the simple correlation coefficients of the spring flood volumes from 1955 to 1976, inclusive, with the various proposed factors which occur during the snow accumulation period. The null hypothesis that there is no association between the spring flood volume and certain factors which occur during the snow accumulation period can be rejected at the 5 percent level of significance. It is evident from the simple correlation coefficients of Table 7 that winter precipitation in the form of total precipitation from December through March, and both the end of February and March watershed snow cover are related to the volume of the spring flood. February and March's uncontrolled inflows, used as indicators of the watershed's groundwater conditions, prove to be significantly correlated with the spring flood volume. However, the form with which winter temperature is introduced, does not appear to be significantly correlated with the spring flood volume.

Table 7 : Correlation of Factors which occur during the Snow Accumulation Period with the Spring Season Flood Volume (1955-1976)

Variable Description		Correlation Coefficient
Precip.	December	0.227
	January	0.403
	February	0.185
	March	0.233
	Dec. and Jan.	0.418
	Feb. and Mar.	0.320
	Dec. through Feb.	0.526*
	Dec. through Mar.	0.536*
Temper.	December	-0.232
	January	-0.015
	Dec. and Jan.	-0.161
	February	0.197
	March	0.142
	Feb. and Mar.	0.217
	Dec. through Feb.	-0.026
	Dec. through Mar.	-0.040
Uncont. Inflows	December	-0.091
	January	-0.020
	February	0.438*
	Jan. and Feb.	0.157
	March	0.488*
	Jan. through Mar.	0.294
	Feb. and Mar.	0.523*
Snow Cover	January	0.418
	February	0.535*
	March	0.445*

* :denotes significant correlation at the 5 percent level to reject the null hypothesis that no association exists between the dependent and independent variables.

4.1.3 Variables Subsequent to the Snow Accumulation Period

Hydroclimatic factors associated with this category are reported to have the greatest effect on the volume and distribution of the spring freshet in the province of Quebec(11,12,15). The basic factors in this group are represented by precipitation and temperature. These factors are introduced as either the combination of April and May or each of the two months seperately. Table 8 lists the simple correlation coefficients obtained for each of the above factors. The null hypothesis that there is no association between the spring flood volume and precipitation which occurs during the flood period may be rejected at the 5 percent level of significance. However, the analogous null hypothesis concerning the average temperature which occurs during the flood period can not be rejected at the 5 percent level of significance. Temperature, in the form introduced for analysis, does not appear to influence the quantity of water received during the April 1 to June 15 period. It is interesting to note that the coefficient of determination for the April and May precipitation is 0.48, while the coefficient for the precipitation which occurs during the snow accumulation period is only 0.29. This supports the previously reported hypothesis that factors subsequent to the snow accumulation period, which are usually unknown at the time operational forecasts are performed, have the greatest influence on the volume of the spring freshet.

Table 8 : Correlation of Factors which occur Subsequent to the Snow Accumulation Period with the Spring Season Flood Volume (1955-1976)

Variable Description	Correlation Coefficient
Precipitation April	0.526
Precipitation May	0.562
Precipitation Apr. through May	0.692
Temperature April	-0.108
Temperature May	0.139
Temperature Apr. through May	0.053

4.2 TENTATIVE REGRESSION MODELS

Two seperate groups of linear additive models have been postulated. Hydrometeorological variables which have the greatest effect on the volume of the spring freshet are drawn from each of the defined time periods. Summer temperature and precipitation are chosen to represent factors antecedent to the snow accumulation period which possibly influence the spring flood volume. During the snow accumulation period, snow water equivalence at the end of February is used to represent the state of the snowpack over the watershed. Total precipitation from December to March, inclusive, is used to replace the snowpack's water equivalence in certain models. February's and March's mean temperature is used to reflect winter severity, while December's mean temperature is included so as to reflect climatic conditions at the start of the snow accumulation period. Precipitation which occurs during the freshet period is introduced as either the total of April and May, or as each month seperately.

4.2.1 Analysis of Models for April 1 Forecasts

As the April 1 forecast represents an important estimate for production planning studies of the hydroelectric system, therefore special emphasis is given to its accuracy. The April 1 predictions described in this section are based on hydrometeorological variables known at the time of the forecast. A total of 15 models are developed based on the previously described regression of the original independent variables, and regression of the components of the independent variables. The equations are given in Table 9.

The correlation coefficients for the 15 models range from 0.536 to 0.850, representing 28.7 and 72.3 percent of the total variation, respectively. The difference between the highest and lowest represents 43.6 percent of the variation, and is significant. The standard errors of the estimate range from 2.30 Gm^3 (81.17 BCF) to 3.50 Gm^3 (123.73 BCF). As the standard deviation of the criterion variable is given as 4.06 Gm^3 (143.2 BCF), the difference in the standard error of estimate is significant.

The standardized partial regression coefficients or beta weights may be used to assess both the rationality of the models and the relative importance of the predictor variables. The coefficients of the models listed in Table 9 are considered rational, as their signs match the signs of the corresponding predictor-criterion correlation coefficients. If the predictor variable correlates positively with the criterion, then the regression coefficient should also be positive. Irrationality is thought to be a result of statistical manipulations in the regression technique. The relative importance of the predictor variables is indicated by the magnitudes of the beta weights, with the larger magni-

Table 9 : Tentative models based on factors known prior to the freshet period

Model	Variables										Statistics				
	Precipitation (Jul + Aug)	Temperature (Jul + Aug)	Temperature (Dec)	Precipitature (Dec to Mar)	Inflow (Feb-Mar)	Temperature (Feb-Mar)	Inflow (March)	Snow Cover (February)	Constant	Degrees of Freedom (ν_1, ν_2)	F	SEE (BCP)	r	r (1934-1954)	
1		14.30(2.80) (.415)	-8.68(2.25) (-.321)	29.66(1.93) (.315)	12.66(3.10) (.523)				-1910.40	(4,18)	8.056	94.789	.801	.600	Regression on independent variables
2			-8.82(1.96) (-.326)	36.76(2.08) (.390)	16.58(2.04) (.383)				- 70.64	(3,19)	5.968	110.579	.697	.674	
3				35.03(1.86) (.372)	15.03(1.74) (.347)				- 79.16	(2,20)	6.155	118.182	.617	.679	
4				50.55(2.91) (.536)					171.34	(1,21)	8.478	123.727	.536	.725	
5		13.92(3.10) (.403)	-7.23(2.12) (-.268)		26.72(4.70) (.618)			15.29(3.15) (.401)	-2020.22	(4,18)	11.623	83.650	.849	-	
6					20.34(2.97) (.470)			18.44(3.05) (.483)	- 272.68	(2,20)	10.172	105.760	.710	-	
7							30.77(2.55) (.422)	18.19(2.87) (.476)	- 122.72	(2,20)	8.556	110.271	.679	- *	
8	17.35	13.79	-8.13	39.27	15.01	4.14			-1986.48	(4,18)	9.500	89.766	.824	.654	Regression on principal components
9			-8.80	37.68	16.15				- 68.23	(2,20)	9.420	107.789	.696	.678	
10				33.87	15.56				- 82.12	(1,21)	12.916	115.348	.617	.674	
11				31.94			24.70		32.20	(1,21)	11.170	118.436	.589	.594	
12	8.09	13.41	-5.22		27.56	2.15		13.27	-2071.23	(3,19)	16.497	81.170	.850	-	
13			-3.56		15.89			22.04	- 203.18	(2,20)	11.338	102.830	.729	-	
14							32.75	17.15	- 126.32	(1,21)	17.876	107.739	.678	-	
15					20.63			18.19	- 274.84	(1,21)	21.355	103.219	.710	-	

* : Model employed in current forecasting methodology.

tudes representing the more important variables. Table 9 lists the beta weights associated with the regression models based on the independent variables. Model 1 indicates that summer temperature is more important in the explanation of the spring flood volume than antecedent winter precipitation, while model 5 indicates summer temperature to be of the same importance as watershed snow cover. Tables 6 and 7 list the simple predictor-criterion correlation coefficients for these variables. From these tables, one may note that the summer temperature--July and August--accounts for only 9.0 percent of the variation of the spring flood. However, the winter precipitation of December through March and the watershed snow cover at the end of February account for 28.7 and 28.6 percent of the variation of the spring flood, respectively. One may propose that once the effects of winter precipitation, antecedent groundwater conditions, and fall conditions have been eliminated from the spring flood volume, summer conditions have an important role in the formulation of the proceeding flood volume. In contrast, one may propose that such a cause-effect relationship does not exist, and what has occurred is due solely to spurious correlation. This proposal is supported by the correlation coefficient derived for the calibrated models of Table 9 from 1934 to 1954. One may note that model 1 generates the lowest correlation coefficient, while the most parsimonious model generates the highest statistic.

The model described as "the best forecast found so far for the spring runoff" and is used in operational practice(77) is represented by model 7. It is evident that an improvement may be achieved by representing winter groundwater conditions by the average uncontrolled inflows of February and March. Model 6 and 15 represent improved versions of the models currently employed in the forecast methodology. The coefficients of model 15 are obtained by principal-components regression. The standard error of the estimate of model 15 is 6.4 percent less than that of model 7. Further improvement is obtained through the inclusion of December's average temperature with the previously stated variables. A reduction of the standard error of estimate of 6.8 percent is obtained through principal-components regression--model 13. Appendix C shows the plot of residuals for the tentative models 1 through 15. Various residual plots, including models 6 and 15, possibly indicate that the variance is not constant. However, the nonparametric tests described in Chapter 2, did not indicate significant tendacies in the residuals of the models to cluster. Scattergrams of models 1 through 15 are contained in Appendix D.

Various factors may be studied in the selection of prediction models for water supply forecasting. The simple correlation coefficient, the standard error of the estimate, the rationality of the coefficients, and the relative importance of predictor variables have been considered in model selection. If summer conditions, when used in conjunction with other antecedent causation factors, are accepted as influencing the volume of the spring flood, then models 1,5,8, and 12 are clearly the best models. However, in addition to the previously mentioned factors, one may analyze the fulfillment of the underlying assumptions of each regression model. Table 10 lists values which may be used for comparison with the Kolmogorov-Smirnov statistic(47). From the maximum absolute deviations of Table 10, one may reject the hypothesis that the residuals of models 1 and 5 are drawn from a normal population at the 10 percent level of significance. Thus, models 8 and 12 remain as the best models to accurately forecast spring flood volume based on antecedent hydrometeorological conditions. The difference between these two models is in the introduction of winter precipitation. Model 8 employs total precipitation from December to March, inclusive, while model 12 uses the end of February's watershed snow cover. Both models are derived from regression of the principal components of the independent variables. Models 8 and 12 explain 67.9 and 72.3 percent of the variation of the criterion variable, respectively. The standard errors of the models are 2.54 Gm^3 (89.77 BCF) and 2.30 Gm^3 (81.17 BCF), respectively. As the reported forecast technique has a standard error of 3.12 Gm^3 (110.3 BCF), both models represent a dramatic increase in model accuracy.

Table 10 : The maximum absolute deviation of the cumulative normal distribution function with the sample cumulative distribution function for tentative models 1 through 15.

Model (from Table 9)	Maximum Absolute Deviation
1	0.1875*
2	0.0873
3	0.1092
4	0.0884
5	0.1715*
6	0.0876
7	0.0991
8	0.1310
9	0.0792
10	0.1128
11	0.1119
12	0.1212
13	0.1123
14	0.1121
15	0.0839

* : one may reject the hypothesis that the residuals are from a normal population at the 10 percent level of significance(75).

4.2.2 Supplementary Forecast Models

Models presented in the preceeding section extract the maximum amount of information from hydrometeorological factors known at the date of the forecast--April 1. It is well known that one of the main factors influencing the volume of the spring freshet is the precipitation which occurs during the spring runoff period. It is also well known that this factor cannot yet be accurately forecasted(54). However, models which include the precipitation which occurs during the forecast period offer insight into the cause-effect relationship occuring during the period. This insight can prove valuable to operational personnel, as an accurate numeric relationship including spring precipitation is evolved. Probabilities may then be associated with precipitation levels, thus allowing estimates of forecast volumes to be generated. This permits one to associate spring volumes with certain probabilities of future precipitation. This enables the operational personnel to determine the effect of the precipitation on the magnitude of the forthcoming spring flood event.

The models descibed in this section are based on hydrometeorological variables known at the time of the forecast, as well as precipitation which occurs during the freshet period. A total of 20 models are developed based on the described regression of independent variables and principle-components regression. The equations of the calibrated models are listed in Table 11.

The correlation coefficients for the twenty models range from 0.813 to 0.934, representing 66.1 and 87.2 percent of the total variation, respectively. The difference between the highest and lowest represents 21.1 percent of the variation, and is significant. The standard errors of the estimate range from 1.65 Gm^3 (58.29 BCF) to 2.52 Gm^3 (88.96 BCF). As the standard deviation of the criterion variable is given as 4.06 Gm^3 (143.22 BCF), the difference in the standard error of the estimate is significant. In comparison, the standard error of the estimate for the models based solely on antecedent conditions range from 2.30 Gm^3 (81.17 BCF) to 3.50 Gm^3 (123.73 BCF), while their correlation coefficients range from 0.536 to 0.850. It is, thus, evident that inclusion of data whichs occurs during the freshet period does not always infer a higher accuracy than models based solely on antecedent conditions. However, in general, a significant decrease in the standard error of the estimate is achieved through the inclusion of spring precipitation data.

The standardized partial regression coefficients or beta weights may be used to assess both the rationality of the models and the relative importance of the predictor variables. Model 25,28, and 29 contain irrational coefficients, as their signs do not match the signs of the corresponding predictor-criterion correlation coefficients. If the pred-

Table 11 : Tentative models based on factors prior to and during the freshet period

Model	Prec. (Jul + Aug)	Temp. (Jul + Aug)	Temp. (Dec)	Prec. (Dec-Mar)	Inflow (Feb-Mar)	Snow Cover (Feb)	Prec. (Apr)	Prec. (May)	Prec. (Apr + May)	Constant	Temp. (Feb-Mar)	Degrees of Freedom	F	SEE (BCF)	r	r (1934-1954)	
	Variables											Statistics					
16			-8.58(3.06) (-.318)	33.71(3.06) (.358)	12.44(2.16 (.287)		77.73(4.08) (.432)	40.42(2.83) (.318)		- 198.25		(5,17)	15.565	68.984	.906	.653	Regression on ind var
17			-8.75(3.04) (-.323)	33.33(2.94) (.354)	11.03(2.07) (.255)				54.83(5.31) (.580)	-954.82		(4,18)	17.955	70.881	.894	.688	
18				42.24(3.35) (.448)			77.47(3.21) (.430)	47.42(2.71) (.372)		- 34.17		(3,19)	13.178	87.800	.822	.669	
19				40.79(3.28) (.433)					58.59(4.70) (.620)	- 17.38		(2,20)	19.513	87.436	.813	.707	
20			-7.01(2.94) (.254)		18.65(4.64) (.431)	15.37(4.46) (.403)	81.89(5.08) (.454)	29.93(2.41) (.235)		- 352.60		(5,17)	23.164	58.288	.934	-	
21			-7.34(2.77) (-.222)		16.71(3.82) (.386)	14.06(3.72) (.368)			50.53(5.28) (.535)	- 271.61		(4,18)	22.304	64.876	.912	-	
22						16.90(3.23) (.443)	79.51(3.74) (.442)	44.89(2.50) (.352)		19.78		(3,19)	12.673	88.960	.817	-	
23					15.02(2.98) (.347)	15.11(3.45) (.396)			50.12(4.51) (.501)	- 302.24		(3,19)	20.094	75.440	.872	-	
24	0.32	4.56	-7.77	28.47	17.03		41.86	39.74		- 930.13	3.01	(4,18)	19.198	68.997	.900	.681	Regression on principal components
25	-0.51	11.31	-4.40	27.08	14.50				42.47	-1585.05	2.40	(4,18)	18.703	69.729	.898	.656	
26				22.18	12.37		65.76	52.90		- 165.62		(2,20)	23.534	82.026	.838	.672	
27				22.22	14.21			52.87		- 188.63		(2,20)	21.781	84.259	.828	.679	
28	11.62	3.15	0.59		10.71	10.88	38.55	43.87		- 807.69	2.47	(2,20)	23.654	81.880	.838	-	
29	-1.16	+9.66	-6.93		17.61	15.97			32.66	-1474.14	2.56	(4,18)	25.433	61.390	.922	-	
30					22.21	15.08	60.37	27.29		- 445.58		(2,20)	34.697	71.049	.881	-	
31					18.58	14.11			44.98	- 351.73		(1,21)	63.726	72.980	.867	-	
32			-8.77	26.22	13.15		64.83	52.92		- 155.69		(3,19)	26.266	67.928	.898	.655	
33			-10.15	31.69	15.97				43.38	- 204.59		(2,20)	35.052	70.769	.882	.679	
34			-5.00		15.48	19.36	77.22	28.79		- 355.67		(3,19)	36.980	58.931	.924	-	
35			-4.47		13.94	19.27			44.37	- 299.99		(2,20)	40.460	66.369	.912	-	

ictor variable correlates positively with the criterion, then the regression coefficients should, as well, be positive. From the magnitude of the beta weights listed in Table 11, it is evident that spring precipitation is the major factor influencing the volume of the spring flood. When the spring precipitation is entered monthly rather than as a cumulative total, the April precipitation proves more important than the May precipitation.

As previously mentioned, various factors may be studied in the selection of models for water supply forecasting. Models 20 and 34 represent the best models in terms of the highest multiple correlation coefficient and the lowest standard error of the estimate. Model 34 represents a more parsimonious description than model 20--due to principal-components regression. Spring precipitation is entered disaggregated, while winter precipitation is entered in the form of the end of February's watershed snow cover. Models 20 and 34 explain 87.2 and 85.4 percent of the variation of the spring flood volume, respectively. If spring precipitation is entered as one variable rather than two seperate entities, then models 21 and 35 are formulated. Models 21 and 35 are analogous with models 20 and 34 except for the inclusion of spring precipitation. Both of the models 21 and 35 explain 83.2 percent of the variation of the spring flood volume. Their standard errors of the estimate are 1.84 Gm^3 (64.88 BCF) and 1.89 Gm^3 (66.87 BCF), respectively. It is evident that models 21 and 35 represent a decrease in model accuracy. In addition to model accuracy, one may analyze the fulfillment of the underlying assumptions of the regression model. Table 12 lists values which may be used for comparison with the Kolmogorov-Smirnov statistic(47). From the values listed in Table 12, one cannot reject the hypothesis that the residuals of models 16 through 35 are drawn from a normal population at the 10 percent level of significance. Appendix E shows the plot of residuals for the tentative models 16 through 35. The residual plots do not indicate model inadequacy. As well, the nonparametric tests described in Chapter 2 do not indicate anomalies in the residuals. Scattergrams of models 16 through 35 are shown in Appendix F.

Models 20,21,34, and 35 represent the best models from Table 11 in terms of the various factors developed for model selection. Models 20 and 21 are based on regression of the independent variables, while models 34 and 35 are derived by principal-components regression. Models 34 and 35 represent a more parsimonious description than models 20 and 21.

Table 12 : The maximum absolute deviation of the cumulative normal distribution function with the sample cumulative distribution function for tentative models 16 through 35.

Model (from Table 11)	Maximum Absolute Deviation
16	0.0795
17	0.0822
18	0.1123
19	0.0693
20	0.0719
21	0.1245
22	0.0667
23	0.0712
24	0.0693
25	0.1496
26	0.1154
27	0.0695
28	0.0841
29	0.0806
30	0.1417
31	0.1482
32	0.0645
33	0.0733
34	0.0878
35	0.1536

Chapter 5

CONCLUSIONS AND RECOMMENDATIONS

5.1 CONCLUSIONS

From the simple correlation analysis of each general category of independent variables, one may state that for the hydrometeorological data of the Saquenay-Lac St-Jean region of Quebec:

1. summer temperature and precipitation prove to be the most significant factors antecedent to the snow accumulation period,
2. winter precipitation in the form of total precipitation from December through March, and both the end of February and March watershed snow cover are related to the volume of the freshet,
3. February's and March's uncontrolled inflows, used as indicators of the watershed's groundwater condition, prove to be significantly correlated with the spring flood volume,
4. the temperature of December proves to be the most significant temperature factor during the snow accumulation period,
5. temperature subsequent to the snow accumulation period does not appear to influence the volume of the spring freshet, and

6. precipitation subsequent to the snow accumulation period is related to the volume of the spring freshet.

It was found that the coefficient of determination for the April and May precipitation with the volume of the freshet is 0.48, while the coefficient for the precipitation which occurs during the snow accumulation period is only 0.29. This supports the hypothesis that factors subsequent to the snow accumulation period, which are usually unknown at the time operational forecasts are performed, have the greatest influence on the volume of the spring freshet.

For the case study, a collation of tentative regression and principal-component regression models was performed. The review of the literature indicated that the current forecasting methodology for the study area was derived using traditional regression techniques. The standard error of estimate and correlation coefficient for the reported forecast model are 3.12 Gm^3 (110.27 BCF) and 0.679, respectively. Through the general categorization of possible independent variables and the expansion of the hydrometeorological data base, an improved forecast model was derived using traditional regression analysis. The new model has a standard error of estimate and correlation coefficient of 3.00 Gm^3 (105.76 BCF) and .710, respectively. A principal-component regression of factors prior to the spring flood yielded a forecast model having a standard error of estimate of 2.30

Gm^3 (81.17 BCF) and a correlation coefficient of 0.850. It is evident that the forecast model developed by the regression of principal components is superior to the models derived by the traditional regression techniques.

From models derived on hydrometeorological data known at the time of the forecast and precipitation which occurs during the freshet period, it was evident that the inclusion of data which occurs during the freshet period does not always infer a higher accuracy than models based solely on antecedent conditions. It was also evident that spring precipitation is the major factor influencing the volume of the spring flood. When the spring precipitation is entered monthly, rather than as a cumulative total , the April precipitation proves more important than the May precipitation. In addition, the supplementary forecast models may be used to update (restart) forecasts as one proceeds into the forecast period.

The proposed fitted models were subjected to diagnostic checks and tests of "goodness-of-fit" to determine the adequacy of each model and the efficiency of which the fitting process made use of the data. Residual analysis to determine their randomness and normality, the analysis of the rationality of the regression coefficients, and the analysis of the indicators of the relative importance of the independent/predictor variables exposed 7 of 35 forecast models to be inadequate for forecast purposes.

5.2 RECOMMENDATIONS

Although the technique described in this thesis is mathematical in nature, engineering judgement concerning the processes of the hydrologic cycle must still form an integral part in the development of forecast models.

The study has shown that principal-component regression is a formidable tool for the water resources practitioner, thus the use of principal-component regression for the forecasting of spring water yield is advocated.

The determination that the beginning portion of freshet precipitation has a greater influence on spring water yield than the latter portion should be investigated for various regions.

REFERENCES

1 Adamcyk, P. J., Jolly, J. P., Solomn, S. I. Use of Regression Equations and Hydrologic Models for Flood Forecasting. Proceedings of the Annual Meeting of the Eastern Snow Conference, 1976.

2 Amorocho, J. and Hart, W. E. A Critique of Current Methods in Hydrologic Systems Investigation. Transactions, American Geophsical Union, Vol 45, No 2, 1964, pp. 307-321.

3 Anderson, Eric Techniques for Predicting Snow Cover Runoff. The Role of Snow and Ice Symposium, IAHS-AISH Publication No 107, Vol 2, Proceedings of the Banff Symposium, 1972, pp. 840-863.

4 Anderson, T. W. An Introduction to Multivariate Statistical Analysis. John Wiley and Sons Inc., New York, 1958.

5 Anscombe, F. J. and Tukey, John W. The Examination and Analysis of Residuals. Techonometrics, Vol 5, No 2, 1963.

6 Baker, Donald G. Prediction of Spring Runoff. Water Resources Research, Vol 4, 1972, pp. 966-972.

7 Bartlett, M. S. Tests of Significance in Factor Analysis. British Journal of Psychology (Statistical Section), Vol III, 1950.

8 Box, George E. P. and Jenkins, Gwilyn M. Time Series Analysis: Forecasting and Control (Revised Edition). Holden-Day Inc., Toronto, 1976.

9 Cambell, David R. The Use and Value of Snow Survey Data in the Operation of Hydroelectric Facilities. Proceedings of the Eastern Snow Conference, 1963, pp. 192-209.

10 Cattell, Raymond B. Handbook of Multivariate Experimental Psychology. Rand McNally and Company, Chicago, 1966.

11 Cavadias, George S. Reappraisal of Snow-Melt as a Factor in Quebec Streamflow. Proccedings of the Annual Meeting of the Eastern Snow Conference, 1955, pp. 67-77.

12 Cavadias, George S. An Approach to Forecasting the Spring Run-Off in Quebec. Proceedings of the 26th Annual Meeting of the Western Snow Conference, 1958, pp. 34-35.

13 Cavadias, George S. Methods of Analysis and Interpretation. Proceedings of Hydrology Symposium No 4, National Research Council, 1964.

14 Cavadias, G. and Brennen, L. M. Discussion on Distribution-Free Methods. Proceedings of Hydrology Symposium No 5, National Research Council, 1966.

15 Cavadias, George S. and Brisebois, R. Long-Term Forecasting of Snow-Melt Runoff. Proceedings of Hydrology Symposium No 8, National Research Council, 1971, pp. 239-257.

16 Chow, V. T.(Editor-in-Chief). Handbook of Applied Hydrology. McGraw-Hill Book Company, Toronto, 1964.

17 Crawford, Norman H. and Linsley, Ray K. Digital Simulation in Hydrology: Stanford Watershed Model IV. Technical Report No 39, Department of Civil Engineering, Stanford University, Stanford, California, 1966.

18 Darlington, Richard B. Multiple Regression in Psychological Research and Practice. Psychological Bulletin, Vol 69, No 3, 1968.

19 Davar, K. S. and Bray, D. I. Preliminary Results of Snowmelt-Streamflow Studies in the Tobique Basin. Proceedings of the Eastern Snow Conference, 1964, pp. 78-96.

20 DeCoursey, D. G. and Deal, R. B. General Aspects of Multivariate Analysis with Applications to Some Problems in Hydrology. Proceedings of the Symposium on Statistical Hydrology, Miscellaneous Publication No 1275, U.S. Department of Agriculture, Washington, D.C., 1974, pp. 47-68.

21 Diskin, M. H. Definition and Uses of the Linear Regression Model. Water Resources Research, Vol 6, No 6, 1970.

22 Draper, N. R. and Smith, H. Applied Regression Analysis. John Wiley and Sons Inc., New York, 1966.

23 Fiering, Myron B. Multivariate Techniques for Synthetic Hydrology. Proceedings of the American Society for Civil Engineers, Journal of the Hydraulics Division, Vol 90, HY-5, 1964.

24 Finlayson, H. M. Snow-Melt as a Factor in Quebec Streamflow. Proceedings of the Annual Meeting of the Eastern Snow Conference, 1952, pp. 27-45.

25 Fogarasi, S. and Mokievsky-Zubok, O. Principal Components Analysis on Glacier-Climatological Data for Sentinal Glacier, British Colombia. Scientific Series No 95, Inland Waters Directorate, Water RFsourcces Branch, Ottawa.

26 Ford, P. M. A Study of Hydrometeorological Relationships in Alaska. Journal of Geophysical Research, Vol 67, No 6, 1962.

27 Golding, Douglas L. The Correlation of Snowpack with Topography and Snowmelt Runoff on Marmot Creek Basin, Alberta. Atmosphere, Vol 12, No 1, 1974.

28 Gray, D. M. (Editor-in-Chief) Handbook on the Principles of Hydrology Water Information Center, Inc., Port Washington, New York, 1970.

29 Green, Paul E. Analyzing Multivariate Data. The Dryden Press, Hinsdale, Illinois, 1978.

30 Haan, Charles T. Statistical Methods in Hydrology. The IOWA State University Press, Ames, Iowa, 1977.

31 Harman, Harry H. Modern Factor Analysis. The University of Chicago Press, Third Edition, 1976.

32 Hawley, Mark E., McCuen, Richard H., and Moreland, Ronald E. Evaluation of Practices in Water Supply Forecasting. Water Resources Bulletin, Vol 16, No 2, 1980.

33 Hopkins, Bryant L. Use of Snow Surveys in Volumetric Prediction of Runoff. Proceedings of the Eastern Snow Conference, 1954.

34 Horowitz, Joseph L. and Shulman, Mark D. An Investigation into the Relationship between Meteorological Variables and Streamflow. Water Resources Bulletin, Vol 3, No 3, 1967. pp. 21-31.

35 Hotelling, Harold Analysis of a Complex of Statistical Variables into Principal Components. The Journal of Educational Psychology, Vol 24, 1933, pp. 417-520.

36 Hotelling, H. The Relations of the Newer Multivariate Statistical Methods to Factor Analysis. British Journal of Statistical Psychology, Vol 10, 1957.

37 Hunter, H. I. Streamflow Forecasting in Mountainous British Colombia, Canada. Water for Peace Conference, Topic VI-B-2, Washington, D.C., 1966.

38 Johnston, J. Econometric Methods. McGraw-Hill Book Company, Toronto, 1963.

39 Kaiser, Henry F. The Varimax Criterion for Analytic Rotation in Factor Analysis. Psychometrika, Vol 23, No 3, 1958.

40 Kaiser, Henry F. The Application of Electronic Computers to Factor Analysis. Educational and Psychological Measurement, Vol XX, No. 1, 1960.

41 Kaiser, Henry F. Psychometric Approaches to Factor Analysis. Proceedings of the Invitational Conference on Testing Problems, Educational Testing Service, Princeton, N.J., 1964.

42 Kaiser, Henry F. An Index of Factorial Simplicity. Psychometrika, Vol 39, No 1, 1974.

43 Kendall, M. G. A Course in Multivariate Analysis. Charles Griffin and Company Limited, London, 1957.

44 Kendall, M. G. Multivariate Analysis. Hafner Press, New York, 1975.

45 Koelzer, Victor A. and Ford, Perry M. Effect of Various Hydroclimatic Factors on Snowmelt Runoff. Transactions, American Geophysical Union, Vol 37, No 5, 1956, pp. 578-587.

46 Lawley, D. N. and Maxwell, A. E. Factor Analysis as a Statistical Method. Butterworth and Co.(publishers) Ltd., London, 1963.

47 Lilliefors, Hubert W. On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown. The American Statistical Association Journal, 1967.

48 Marsden, Michael A., and Davis, Robert T. Regression on Principal Components as a Tool in Water Supply Forecasting. Proceedings of the Thirty-sixth Annual Meeting of the Western Snow Conference, 1968.

49 Massey, William F. Principal Components Regression in Exploratory Statistical Research. American Statistical Association Journal, 1965.

50 Matalas, Nicholas C. and Maddock, Thomas III Hydrologic Semantics. Water Resources Research, Vol 12, No 2, 1976.

51 Matalas, N. C. and Reiher, Barbara J. Some Comments on the Use of Factor Analysis. Water Resources Research, Vol 3, No 1, 1967.

52 McCuen, Richard H., Rawls, Walter J., and Whaley, Bob L. Comparative Evaluation of Statistical Methods for Water Supply Forecasting. Water Resources Bulletin, Vol 15, No 4, 1979, pp. 935-947.

53 Mustonen, Seppo E. Effects of Climatological and Basin Characteristics on Annual Runoff. Water Resources Research, Vol 3, No 1, 1967.

54 Namias, Jerome The Art and Science of Long-Range Forecasting. EOS Transactions, American Geophysical Union, Vol 61, No 19, 1980.

55 Nelson, Morlan W. Water Supply Forecasting as it Applies to Winter Snow Cover and Spring Runoff. Proceedings of the Eastern Snow Conference, 1966, pp. 18-25.

56 Pearson, Karl On Lines and Planes of Closest Fit to Systems of Points in Space. Phil. Mag., Ser 2, No 6, 1901, pp. 559-572.

57 Quick, M. S. Forecasting Runoff Operational Practices. IAHS-AISH, Publication No 107, 1972.

58 Rango, Albert, Hannaford, Jack F., Hall, Roderick L., Rosenzweig, Michael, and Brown, A. Jean Snow-Covered Area Utilization in Runoff Forecasts. Journal of the Hydraulics Division, HY1, 1979, pp.53-66.

59 Rice, Raymond M. Multivariate Methods useful in Hydrology. Proceedings of the International Hydrology Symposium, Fort Collins, Colorado, 1967.

60 Riggs, H. C. discussion of Paper by A.L. Sharp, A.E. Owen, and B. Harris. 'Application of the Multiple Regression Approach in Evaluating Parameters Affecting Water Yields of River Basins', Journal of Geophysical Research, Vol 65, No 10, 1960.

61 Roscoe, John T. Fundamental Research statistics for the Behavioral Sciences. Holt, Rinehart and Winston, Inc., Toronto, 1969.

62 Sharp, A. L. et al Application of the Multiple Regression Approach in Evaluating Parameters Affecting Water Yields of River Basins. Journal of Geophysical Research, Vol 65, No 4, 1960.

63 Shrader, Michael L., McCuen, Richard H., and Rawls, Walter J. The Effect of Data Independence in Model Calibration and Model Testing. Water Resources Bulletin, Vol 16, No 1, 1978.

64 Silver, L. Ray Alcan Fine Tunes its Hydroelectric Power Plants. Modern Power and Engineering, 1979, pp. 22-23.

65 Snyder, Willard M. Some Possibilities for Multivariate Analysis in Hydrologic Studies. Journal of Geophysical Research, Vol 67, No 2, 1962, pp. 18-25.

66 Soil Conservation Service, U.S.D.A. Water Supply Forecasting as it Applies to Winter Snow Cover and Spring Runoff. Proceedings of the Eastern Snow Conference, 1966.

67 Spearman, C. General Intelligence, Objectively Determined and Measured. American Journal of Psychology, No 15, 1904.

68 S.P.S.S. Statistical Package for the Social Sciences, Second Edition. McGraw-Hill Book Company, Toronto, 1975.

69 Stammers, W. N. The Application of Multivariate Techniques in Hydrology. Proceedings of Hydrology Symposium No 5, National Research Council, 1966, pp. 255-270.

70 Thompstone, R. Frequency Analysis and Forecasting of Spring Runoff Volumes - A Case Study. Presented at the Stochastic Modelling of River Flows, Advanced Specialized Seminar, sponsored by the National Research Council of Canada Associate Committee on Hydrology in Cooperation with the Department of Civil Engineering and Applied Mechanics, McGill University, Montreal, 1978.

71 Thompstone, R. and Bergeron, R. Prevision hydrologique interactive dans un systeme d'information hydrometeorologique. Deuxieme Colloque d'Hydrotechnique du Quebec de la Societe Canadienne de Genie Civil, Universite de Sherbrooke, 1980.

72 Thompstone, R. and Pilon, P. J. Square-Grid Spatial Interpolation of Snow Cover for a Hydrological Information System. Proceedings of the Eastern Snow Conference, 1979.

73 Thompstone, R., Poire, A., and Pilon, P. J. Frequency Analysis and Forecasting Spring Inflow Events for Water Resources Management. Canadian Hydrology Symposium-79, Cold Climate Hydrology, Vancouver, B.C., May 1979.

74 Thompstone, R., Sen, D., and Divi, R. Hydrologic Modelling and Optimization Techniques for Operating Multi-Reservoir Water Resources Systems. C.S.C.E. Fourth National Hydrotechnical Conference on River Basin Management, Vancouver, B.C., May 1979.

75 University of Waterloo, Computing Centre. Library Program MLREGR : Multiple Linear Regression Analysis. undated documentation of program.

76 Unny, T. E., Divi, R., Hinton, B., and Robert, A. A Model for Real-Time Operation of a Large Multi-Reservoir Hydroelectric System. Proceedings of the International Symposium on Real-Time Operation of Hydrosystems, University of Waterloo, Waterloo, Ontario, June 1981, pp. 284-304.

77 U.S. Army Corps of Engineers, North Pacific Division. Program Description and User Manual for SSARR Model Streamflow Synthesis and Reservoir Regulation. Program K5-G0010, Portland, Oregon, 202 p.

78 Valdes, Juan B. and Rodriquez-Iturbe, Ignacio. Linear Model Discrimination Theory Applied to the Choice of Structure and Form of Hydrologic Regression Models. Report No 212, Ralph M. Parsons Laboratory for Water Resources and Hydrodynamics, Department of Civil Engineering, M.I.T..

79 Wallis, James R. Multivariate Statistical Methods in Hydrology - A Comparison Using Data of Known Functional Relationship. Water Resources Research, Vol 1, No 4, 1965.

80 Wallis, James R. Factor Analysis in Hydrology - An Agnostic View. Water Resources Research, Vol 4, No 3, 1968.

81 Yevjevich, Vujica. Probability and Statistics in Hydrology, Water Resources Publications, Fort Collins, Colorado, 1972.

82 Zuzel, John F. and Cox, LLoyd M. Relative Importance of Hydrometeorological Variables in Snowmelt. Water Resources Research, Vol 11, No 1, 1975.

Appendix A

HYDROMETEOROLOGICAL DATA BASE

TABLE A1: Total Monthly Precipitation for the Total Watershed in Inches

YEAR	MONTH											
	Jan.	Feb.	March	April	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
1934	2.57	1.48	–	–	2.18	4.72	3.43	4.25	3.44	3.28	2.65	2.52
1935	1.98	1.57	2.25	2.83	1.68	5.32	5.76	2.33	3.82	3.54	3.45	1.54
1936	2.93	2.24	2.30	2.51	6.25	3.43	4.51	3.09	3.48	5.25	2.60	3.47
1937	3.35	2.39	1.42	1.10	2.52	3.82	3.79	4.53	3.98	4.78	3.16	1.83
1938	2.08	2.02	2.63	3.06	1.65	2.64	4.99	6.53	4.02	1.48	2.20	3.25
1939	1.84	3.38	0.94	1.62	3.73	3.87	5.51	3.66	3.42	3.76	1.49	3.35
1940	2.20	1.32	2.45	2.71	3.12	5.71	2.90	3.77	3.87	2.67	3.35	2.92
1941	1.59	0.85	2.47	3.25	0.76	2.97	3.92	4.16	4.97	4.39	4.34	2.73
1942	2.40	2.90	2.13	0.96	2.96	3.69	4.90	3.31	4.79	3.84	2.33	4.39
1943	1.87	2.41	2.39	2.51	2.10	4.68	3.98	5.36	1.40	4.15	1.80	2.20
1944	1.43	1.82	1.77	0.89	3.15	3.01	4.68	3.99	4.60	2.79	3.02	2.60
1945	2.85	2.46	2.46	3.31	3.32	3.78	3.79	3.50	4.19	3.89	2.92	2.80
1946	2.26	2.10	2.58	3.24	2.72	2.98	4.59	2.20	5.50	3.92	3.30	3.62
1947	2.78	3.44	1.93	2.87	4.56	3.45	4.33	2.69	4.11	1.91	2.43	2.17
1948	1.15	1.81	2.59	1.93	2.82	3.65	4.45	3.42	3.29	2.60	3.29	2.39
1949	2.68	1.57	1.96	1.47	3.16	7.12	3.27	4.11	4.52	2.77	2.22	1.81
1950	3.08	1.60	1.81	2.11	1.17	7.38	4.39	2.71	1.86	3.66	4.55	2.95
1951	3.37	2.09	2.35	2.35	1.42	1.85	4.72	–	4.37	3.65	4.34	2.36
1952	2.87	2.06	1.36	1.80	2.33	4.71	5.55	4.57	3.63	3.25	1.58	4.07
1953	3.42	1.81	2.83	2.40	1.09	2.42	4.12	1.57	5.21	0.84	2.17	2.93
1954	2.15	1.96	1.55	2.37	4.37	3.17	3.87	4.30	4.51	1.96	3.70	3.55
1955	3.16	2.11	4.11	1.19	3.75	1.87	4.68	3.57	4.63	2.44	2.31	1.79

TABLE A1: (continued)

YEAR	MONTH											
	Jan.	Feb.	March	April	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
1956	1.68	2.18	1.50	1.94	2.87	3.50	4.90	4.36	3.87	2.59	2.22	2.94
1957	1.25	1.99	1.32	1.96	1.78	5.14	4.68	2.83	6.18	2.51	3.48	3.15
1958	2.04	2.66	0.98	2.49	3.67	5.72	4.12	4.86	3.62	3.43	2.67	1.89
1959	2.83	1.01	2.07	1.56	2.23	4.90	5.80	3.52	3.03	2.51	3.45	2.37
1960	2.13	3.90	0.99	2.24	2.38	3.61	5.34	2.94	3.89	2.16	3.68	1.79
1961	2.01	2.49	2.18	1.01	1.90	4.50	3.65	6.31	3.83	1.21	2.36	2.54
1962	3.14	2.03	0.64	2.50	4.06	2.51	6.26	3.11	4.20	2.34	1.03	3.14
1963	1.82	1.67	2.06	2.79	1.81	2.95	5.34	5.06	3.53	2.76	4.95	1.23
1964	3.19	1.67	3.03	1.10	4.40	6.09	4.74	4.08	3.38	2.46	2.83	3.33
1965	1.74	2.21	0.47	0.70	4.28	2.56	4.87	5.51	4.12	4.68	3.22	2.73
1966	3.42	1.23	3.22	1.14	2.59	3.46	3.84	5.75	5.76	4.23	4.73	3.69
1967	2.95	2.42	1.21	1.32	2.58	2.48	4.06	3.99	4.40	5.37	3.60	3.05
1968	1.41	2.47	2.85	1.71	0.62	2.69	5.31	3.94	2.38	3.04	2.49	4.65
1969	2.92	1.34	1.69	1.40	2.92	3.54	4.32	4.52	3.93	3.10	3.98	3.52
1970	1.42	2.75	2.02	1.84	2.65	4.35	6.74	3.37	5.35	1.71	2.56	1.66
1971	2.17	3.07	2.64	2.27	2.86	2.70	4.01	6.03	4.00	3.88	2.19	3.55
1972	1.89	2.84	3.72	0.43	1.97	4.72	3.81	5.00	5.05	3.19	1.91	2.90
1973	2.59	2.80	2.03	3.32	3.81	3.72	6.02	6.07	3.78	3.27	2.44	4.69
1974	2.61	1.78	2.77	3.42	4.76	2.29	4.79	3.74	5.00	3.43	1.89	2.80
1975	2.53	1.58	2.90	2.05	2.93	2.98	5.99	2.90	5.09	2.51	3.48	2.64
1976	2.78	3.32	2.71	2.56	5.04	2.78	4.53	4.72	2.70	3.80	2.20	3.85
1977	2.53	2.43	1.74	2.62	1.65	4.25	5.14					

TABLE A2: Average Monthly Temperature for the Total Watershed in Deg. Far.

YEAR	MONTH											
	Jan.	Feb.	March	April	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
1934	-1.67	-6.58	-	-	47.58	57.00	62.93	57.92	57.91	39.27	28.00	2.91
1935	-5.21	1.45	14.97	35.26	45.79	60.38	65.07	64.23	49.51	40.62	27.63	8.03
1936	1.09	1.09	26.99	34.08	44.28	58.31	62.42	58.48	51.88	37.56	18.21	9.37
1937	4.13	12.05	12.32	36.02	51.71	59.66	66.22	65.86	52.51	41.34	28.20	9.45
1938	0.53	3.00	17.06	34.55	48.52	61.59	64.15	62.12	51.05	43.12	27.25	13.19
1939	0.54	0.80	8.03	31.72	45.63	57.55	64.99	63.72	50.76	38.19	23.91	12.73
1940	4.92	4.16	17.72	31.39	46.99	55.65	63.31	61.49	53.09	38.24	25.49	6.50
1941	-1.70	10.01	16.04	35.45	48.72	58.53	65.82	57.82	52.00	39.29	26.84	12.59
1942	-1.90	7.24	22.86	36.38	51.02	60.92	60.22	61.14	51.66	41.89	25.45	1.82
1943	-4.38	0.92	15.07	25.38	44.43	55.48	63.61	59.12	51.96	43.61	26.78	5.17
1944	9.85	0.72	13.34	29.29	48.98	58.23	63.82	63.76	54.82	38.98	29.39	9.99
1945	-5.10	7.38	23.92	37.00	46.39	56.07	62.79	61.65	51.75	38.28	25.53	6.62
1946	0.28	-3.15	23.42	32.25	43.94	55.19	60.22	59.42	55.05	43.07	23.83	5.38
1947	0.94	9.27	21.90	23.31	40.51	57.14	65.84	63.93	49.85	46.46	27.05	6.28
1948	-2.88	-3.54	10.26	30.79	47.03	55.88	63.13	62.41	54.37	41.47	33.07	14.82
1949	3.55	2.21	15.71	35.42	45.16	61.25	64.36	61.07	51.74	43.32	21.82	14.36
1950	2.95	0.48	12.07	32.23	48.36	54.88	61.78	58.18	49.54	40.07	30.02	12.95
1951	2.61	7.76	21.01	39.51	49.51	56.62	48.73	-	52.67	42.25	22.71	7.69
1952	-1.75	10.56	22.37	35.60	48.14	58.56	67.21	61.60	52.15	37.25	28.20	18.08
1953	4.47	9.94	20.34	36.78	47.58	58.96	62.44	60.10	52.18	42.23	33.77	14.33
1954	-6.70	13.78	17.00	30.96	46.41	59.41	60.79	58.84	46.40	41.67	27.52	10.20
1955	0.23	4.49	13.24	34.67	48.38	62.61	64.58	62.70	49.07	41.81	28.33	5.91

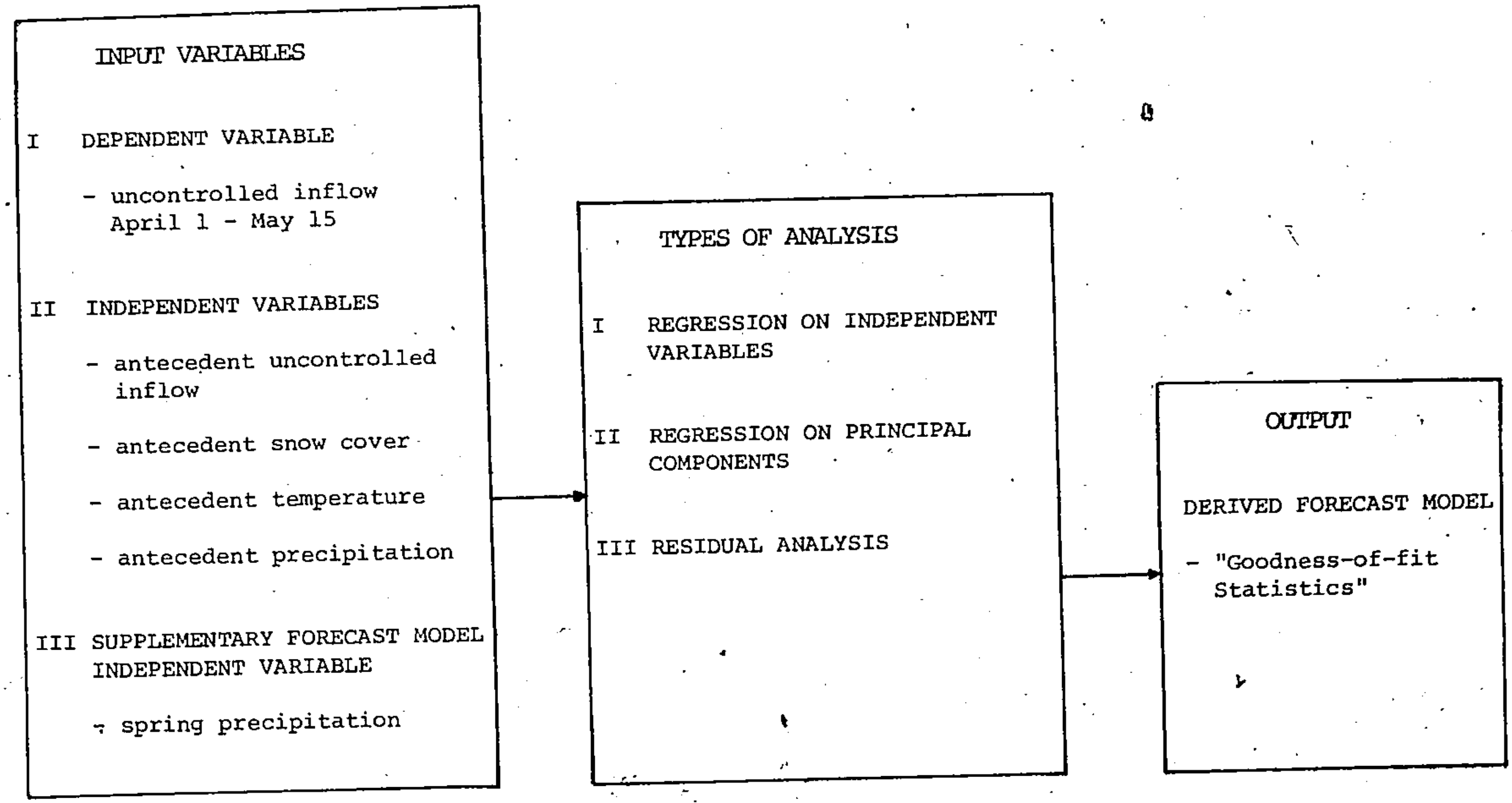
YEAR	MONTH											
	Jan.	Feb.	March	April	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
1956	12.18	6.65	9.43	32.44	39.37	55.79	58.43	57.14	47.62	42.90	26.25	2.68
1957	-9.52	9.15	19.75	31.48	44.87	57.96	61.57	55.89	51.58	42.50	30.34	14.81
1958	8.46	5.90	29.64	37.11	44.04	52.55	62.27	58.83	51.80	39.53	26.78	-1.59
1959	4.17	-5.54	10.53	32.82	49.57	57.01	66.33	61.52	54.05	39.57	24.16	12.08
1960	3.04	14.60	14.25	31.54	53.39	58.77	60.20	61.05	51.76	39.62	30.36	7.67
1961	-6.82	3.87	15.98	37.68	45.00	56.10	62.81	61.29	57.28	43.27	32.20	15.05
1962	-2.88	-5.56	22.80	33.75	47.85	59.17	59.26	61.90	50.98	40.60	25.16	7.67
1963	0.97	-1.18	13.72	33.85	45.76	60.18	65.52	56.65	47.97	45.25	30.86	0.55
1964	6.12	6.55	15.59	33.71	48.13	55.07	63.03	56.73	49.36	37.48	25.20	8.32
1965	-2.30	1.81	16.41	33.13	45.36	58.65	56.59	57.03	49.34	37.13	20.45	9.68
1966	7.51	9.89	21.48	34.46	43.60	58.41	61.93	58.98	49.11	39.96	30.28	13.71
1967	4.05	-3.34	10.16	29.21	40.91	59.85	64.12	61.39	51.64	40.05	25.79	11.41
1968	-5.14	0.15	19.06	36.92	47.34	56.97	61.59	54.90	56.34	44.49	23.98	10.83
1969	11.90	10.98	18.46	29.39	42.62	55.96	61.09	62.02	48.20	38.31	30.83	10.01
1970	-5.98	1.01	18.84	33.01	45.35	57.66	65.75	62.22	49.80	43.12	28.61	1.45
1971	-2.72	5.34	17.21	32.77	48.38	56.55	60.72	58.32	54.21	42.90	23.06	5.33
1972	0.39	-2.15	10.63	30.92	46.94	55.68	62.75	57.29	49.70	35.18	23.96	0.77
1973	4.83	-0.56	24.35	33.79	46.74	60.69	64.08	63.82	50.49	42.30	24.70	11.74
1974	-5.90	0.59	12.49	30.03	40.64	60.49	63.20	60.52	47.87	34.22	27.79	13.92
1975	2.09	4.35	14.98	30.54	51.45	61.67	66.75	61.15	51.45	40.81	28.53	2.57
1976	-3.93	6.65	13.83	33.45	47.32	61.95	62.43	61.02	49.03	36.00	22.30	-1.20
1977	-0.86	8.20	26.57	30.92	49.66	57.08	60.95					

Table A3: Historic monthly uncontrolled inflow (m^3/s) for the total Lac St-Jean watershed.

YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
13	610,194	400,691	616.588	1600;702	5002.035	2457.185	2617,347	1677,394	16-3,180	2212,865	1915;185	958,315
14	531,362	432,851	364,934	445,527	3196,302	2648,525	921,685	568,814	620,151	643,079	666,874	645,149
15	505,556	408,883	328,481	1403,409	4528,676	1899,946	1428,544	524,434	671,703	1647,980	1148,553	691,401
16	510,261	494,700	384,294	2050,886	8196,816	2049,800	833,901	469,977	1638,446	1661,636	1256,631	753,684
17	543,511	440,638	360,179	957,599	3787;220	4478,283	2424,425	1622,312	1315,343	1557,365	1579,449	657,420
18	504,689	443;572	359,539	1115;044	4920,188	3137,470	2551,579	1562,572	2454,512	1733,572	1762,191	822,443
19	601,425	473,305	459,929	1567;051	5219,715	2066,016	812,161	1064,551	1336,581	2190,851	1144,022	705,560
20	497,472	391,285	437,184	1500;727	4882,098	2687,559	1291,912	1071,311	194,019	875,372	1005,550	577,492
21	411,150	356;293	459,564	2039;135	4338,859	1271,639	672,389	994,964	1409,450	2355,732	1154,688	693,867
22	472,078	366,532	357,895	1666;958	3971;832	1967,492	1495,340	1230,527	1457,684	1281,590	811,293	491,078
23	362,006	282,971	226,448	459,308	4670;172	1362,916	680,165	1208,604	2039,849	839,930	644,031	664,454
24	480,208	380,565	330,674	687;168	5141;063	3699,854	2561,444	1809,755	2498,430	1422,537	1193,766	1041,675
25	541,591	494,846	443,761	1449,661	4165;941	2431,991	1406,643	1408;013	181,443	405,218	594,571	627,458
26	408,084	318,368	188,082	266,750	3177,393	3479,545	2139,879	1519,913	554,673	1120,272	2078,684	1283,680
27	822,912	497,627	339,717	1184,683	3752,400	2835,230	2260,274	1777,692	1201,289	2099,084	2531,767	896,930
28	573,536	449,645	353,401	833,286	7000;969	4095,920	1961,023	1150,272	1301,497	1997,379	1311;095	638,511
29	418,458	390,678	252,755	677;918	4242;582	5156,641	2047,163	2278,909	1556,262	1726,264	1342,527	686,129
30	487,607	342,640	280,707	747,843	4937;957	3625,946	2681,792	2585,970	1738,148	1221,118	1128,542	763,976
31	453,124	355,636	317,931	1525;929	2738,245	2441,865	1616,237	992,867	1029,761	1637,987	2121.213	1240,758
32	696,763	505,162	352,231	1177;271	3024;459	1636,660	2348,697	2065,158	2453,880	3478,226	1724,623	767,264
33	560,266	459,551	409,049	1564,158	4781;523	2920,749	1024,906	628,189	508,297	840,569	635,442	423,847
34	289,477	210,560	216,765	782,031	4411;938	3033,875	2068,082	1559,374	1585,773	1773,399	1434,087	1248,249
35	542,689	360,339	257,871	1335;920	3313;863	2663,721	1869,312	1764,173	1120,896	1585,042	1539,522	680,420
36	391,693	277;308	1075,512	1460,327	5183;633	3411,866	1163,753	985,243	571,097	2391,883	1426,535	649,016
37	542,963	393,308	267,371	1066,905	6084;582	2035,454	818,920	1657,755	1569,208	1923,479	1627,589	752,694
38	490,804	488;893	375,251	1960,129	3855;913	2304,762	1677,715	2361,578	1383,240	1302,874	1194,616	749,863
39	458,554	384,764	233,207	501,406	4321,504	2728,380	2652,698	1495,523	1697,771	1595,822	1325,725	761,463
40	498,843	346,607	254,582	562,572	5229;215	3746,466	1751,750	1118,681	1167,337	911,363	898,794	519,669
41	388,770	272,150	172,827	1953;710	2673;800	1110,796	998,873	568,175	2607,544	2490,376	1689,437	1125,205
42	563,612	423,041	303,818	1666,195	5662;109	2248,508	933,560	682,540	1780,366	2276,169	2253,497	669,569
43	489,434	388,251	286,828	380,280	5327;777	3757,717	1168,777	1631,538	1078,947	711,954	1071,624	494,001
44	453,169	306,839	175,568	292,235	3980;237	2111,438	1128,676	1425,804	1783,479	1417,422	1356,308	577,218
45	495,554	297;838	472,169	3018;368	3459,580	2036,491	1786,899	1007,916	1707,349	1964,221	1437,674	654,795
46	470,068	483;823	405,487	1287;642	4696;754	2780,439	1958,748	968,089	566,472	2115,948	1392,649	642,074
47	538,030	589,507	447,871	475,071	5254;973	5796,285	1303,422	1155,897	1249,676	1333,789	888,883	529,169
48	453,717	352,349	322,453	746;446	4340,777	1770,592	1222,397	1618,480	1126,386	851,531	1081,913	597,771
49	540,222	431,537	354,789	1272;961	4530,847	4186,344	1752,572	1343,157	1555,002	1492,417	973,080	650,112
50	626,088	606,599	357,621	1045,477	4003;530	4395,895	2276,900	966,902	882,749	1389,196	1768,421	1232,354
51	645,088	500,206	484,891	3111,797	3696;058	1217,553	774,161	1015,955	944,479	1719,413	1681,392	1198,282
52	553,559	489,680	473,905	1121;746	4379;675	2769,534	1330,003	1597,807	1639,293	1259,484	872,836	1294,196
53	670,795	580,506	763,564	3184,194	3802;566	1547,923	1287,345	541,227	1711,031	1154,162	1037,171	1016,421
54	497,655	478,564	446,227	1295,992	4799,116	1940,387	1078,253	1103,190	1207,548	1445,009	1363,482	848,699
55	549,814	406,557	411,059	1698,571	4814,410	1794,189	899,385	576,852	926,450	1398,604	1542,636	583,864
56	429,876	355,484	259,332	842,914	2346,962	3777,066	2068,882	1956,914	2292,103	1761,432	850,465	547,256
57	354,881	333,235	356,799	1069,925	3097,191	2701,479	2774,898	1214,455	2022,427	1303,684	1683,185	1494,421
58	723,737	499,397	477,011	1848,087	3399,638	3421,777	1490,870	1434,504	1993,544	1556,086	854,619	576,396
59	466,059	411,917	360,309	758,434	4723,783	2802,005	1623,134	1358,138	1088,331	1557,913	1467,973	758,083
60	419,280	431,334	341,361	868,694	5918,059	1423,873	1735,033	2509,559	1283,061	1298,854	1374,057	761,007
61	454,580	316,345	310,030	1109,875	3469,152	1534,991	1187,777	1353,388	1505,258	960,414	867,361	582,059
62	497,290	362,867	294,866	970,720	3838,191	2051,971	1289,444	1390,189	1406,524	910,267	544,166	438,515
63	387,583	293,590	284,361	1033,773	3569,098	2337,883	1881,552	1231,423	749,655	722,733	1074,816	783,752
64	448,419	430,121	361,589	1302,977	5188,040	2809,084	1745,995	1835,888	1270,507	1479,446	821,865	826,685
65	609,737	425,570	315,967	588,341	3834,174	2079,156	1850,484	2215,240	1891,111	2129,649	1034,528	456,963
66	473,996	423,648	484,684	1603,047	3804,210	2665,763	1140,008	1893,792	1505,352	2689,449	2343,357	1449,458
67	636,339	455,303	359,813	614,298	2824,339	2821,355	1448,754	1393,581	878,571	3029,776	2140,665	833,992
68	429,967	388,757	388,953	2547;412	2705,770	1091,824	1815,315	1933,386	1368,957	1486,388	928,149	636,410
69	503,501	440,416	384,558	906,534	3939,477	3219,874	1514,523	1485,384	1347,813	1313,813	1873,478	916,478
70	558,035	488,677	362,188	955,334	4443,633	3462,366	2333,535	1542,932	1872,533	1289,788	938,415	552,646
71	381,462	353,563	329,304	951,464	4267;152	1088,992	942,329	1533,797	1495,630	2477,313	1123,539	578,081
72	430,241	389,971	364,015	542,183	3644,263	2788,861	1235,734	930,728	1146,948	1803,178	722,471	495,371
73	430,150	425,974	417,362	1813,274	5593;418	2542,239	1656,118	2147,827	1440,480	1831,587	1047,271	695,146
74	482,857	422,839	366,208	841;781	5585;188	5388,512	1255,899	1333,475	1167,998	1596,444	1521,384	891,986
75	585,530	389,667	388,770	833;758	3976,308	2484,755	1365,354	1179,921	1214,344	1187,960	1291,658	747,132
76	468,158	452,370	426,588	2485,417	7811;199	2472,281	1471,958	1347,999	1473,825	1642,134	815,669	532,549
77	437,184	355;788	404,664	1896;981	5358;797	2326,839	1295,189	1484,816	1489,450	1159,368	1823,862	657,511

TABLE A4: The total Lac St-Jean watershed snow cover, in equivalent cm of water, as estimated by the arithmetic averaging of reporting snow survey stations from 1955-1979, inclusive.

YEAR	MONTH		
	January	February	March
1955	21.84	26.3	29.3
1956	12.45	16.6	23.3
1957	13.97	17.9	20.7
1958	12.19	15.5	16.4
1959	14.22	18.4	21.9
1960	16.00	21.7	26.0
1961	13.21	19.3	21.7
1962	20.32	24.4	22.4
1963	13.46	17.8	20.0
1964	15.49	22.4	27.3
1965	13.46	19.7	22.8
1966	15.75	21.4	21.4
1967	12.19	17.6	20.4
1968	14.99	21.5	23.5
1969	23.62	26.9	28.1
1970	11.18	18.7	26.8
1971	11.18	19.5	27.3
1972	17.78	25.8	32.4
1973	14.48	20.2	23.6
1974	22.10	27.3	32.8
1975	12.70	16.6	21.3
1976	15.24	25.1	24.9
1977	23.11	26.8	32.0
1978	19.56	20.8	27.5
1979	20.57	22.8	23.5



Organization Chart depicting analyses performed in computer programs of Appendix B

Appendix B

COMPUTER PROGRAMS USED

```
//  JOB ,'PILON',MSGLEVEL=(1,1),CLASS=K
//  EXEC FORTGCLG
//FORT.SYSIN DD *
C-----------------------------------------------------------------------
C
C     PROGRAM TO PERFORM
C     1- REGRESSION ON PRINCIPLE COMPONENTS
C     2-REGRESSION ON INDEPENT VARIABLES
C     3-RESIDUALS ANALYSIS OF THE ABOVE TWO(2) MODELS
C
C     PROGRAM WRITTEN BY PAUL J. PILON
C     UNIVERSITY OF OTTAWA----CIVIL ENG.
C     SUMMER 1981
C
C-----------------------------------------------------------------------
C
C     N = NUMBER OF OBSERVATIONS
C     M = NUMBER OF EXPLANATORY VARIABLES
C     NSP1 = STEP AT WHICH STEP-WISE REGRESSION WILL STOP CONCERNING
C            CONSTRUCTION OF PRINCIPLE COMPONENT MODEL
C     NSP2 = ANALOGOUS TO NSP1 BUT FOR ORIGINAL IND. VARIABLES
C     NSP1&2  MUST BE EQUAL TO OR LESS THAN M.
C     M1 = THE NUMBER OF INDEPENDENT VARIABLES IN THE DATA SET
C     KK = THE DIMENSION OF THE C ARRAY
C     XY = THE DEPENDENT VARIABLE UNDER STUDY
C
C-----------------------------------------------------------------------
C
      REAL *8 XBAR(6),STD(6),RX(6,6),R(6,6),B(6),D(6),
     1T(6),A(6,6),H(6),F(6),C(21),Z(6,6),
     2ROTPC(23,6),V(6,6),XSTAND(23,6),PC(23,6)
      REAL*4 X(23,6),X1(43,34),XY(23),XZ(6),XFLOOD(43),
     1B1(58),SIGB(58),S(23),SS(23)
      INTEGER ICHAR1/'X'/
      INTEGER ID(60),ITITLE(20)
      DATA ITITLE/'    ','RESI','DUAL','S FR','OM T',
     1'HE R','EGRE','SSIO','N   ',11*'    '/
      DATA ISUBTI/'RSDL'/
C     INTEGER NCTR(4)/5,6,10,21/
C     NCTR(14)/1,2,3,4,5,6,7,10,11,12,13,15,21,22/
      COMMON/PAUL/ID,NIN,NVIN,NVAR,NTRAN,NCONS,FIN,FOUT,IRES,IANNEE,
     1     NSP,BSUBO,B1
      COMMON/PAUL2/SIGB
C
```

```
      NSP1=6
      NSP2=6
C
      NSP=NSP1
      N=23
      M1=34
      M=6
      NSTAN=1955
      NFINAN=1977
      NSTAN=NSTAN-1934
      NFINAN=NFINAN-1934
      IF((NFINAN-NSTAN+1).EQ.N) GO TO 1
      WRITE(6,400)
 400  FORMAT('1','  ERROR WITH NSTAN,NFINAN, AND N!!!!')
      STOP
C
C-----------------------------------------------------------------------
C
C     INITIALIZATION OF COEF. FOR THE REGRESSION
C
C-----------------------------------------------------------------------
    1 MAM=0
      XMIN=400.
      XMAX=1200.
      YMIN=XMIN
      YMAX=XMAX
      NVIN=M+1
      NVAR=M+1
      NTRAN=0
      NCONS=0
      IRES=1
      NOBS=N
      IMONTH=1
      NCASE=1
C
C     MAM1=0  IMPLIES THE SCALE OF THE PLOT (X1MIN,X1MAX)
C             ARE GIVEN BY THE  PROGRAM.
C         =1  IMPLIES THE SCALES ARE COMPUTED IN THE PLOT SUBROUTINE.
C
C     X1MIN   THE MIN. SCALE VALUE.
C     X1MAX   THE MAX. SCALE VALUE.
C
      MAM1=0
      X1MIN=-500.0
      X1MAX=500.0
C     FIN,FOUT CONTROL THE REJECTION AND
C     ACCEPTANCE LEVELS FOR THE REGRESSION.
      FIN=0.0
      FOUT =0.0
C
      KK=(M*(M+1))/2
C
C     READ THE DEPENDENT VARIABLE.
      READ(5,490) (XFLOOD(I),I=1,43)
```

```
490   FORMAT(10F8.0/10F8.0/10F8.0/10F8.0/3F8.0)
C     TRANSFER OF SPRING VOLUMES FROM XFLOOD TO XY ARRAY.
      DO 5 I=1,N
  5   XY(I)=XFLOOD(NSTAN-1+I)
C     READ THE INDEPENT VARIABLES IN THE DATA SET.
      DO 10 I=1,43
 10   READ(5,500) (X1(I,J),J=1,M1)
500   FORMAT(10F8.3/10F8.3/10F8.3/4F8.3)
C     NATURAL LOG TRANSFORM OF THE DEPENDENT VARIABLE.
C     DO 15 I=1,N
C 15  XY(I)=ALOG(XY(I))
C
      NC=1
 50   CONTINUE
      DO 20 I=1,N
      I1=NSTAN-1+I
      X(I,1)=X1(I1,12)+X1(I1,13)
      X(I,2)=X1(I1,1)+X1(I1,2)
      X(I,3)=X1(I1,33)
      X(I,4)=X1(I1,30)+X1(I1,31)
      X(I,5)=X1(I1,8)+X1(I1,9)
      X(I,6)=X1(I1,6)
 20   CONTINUE
      IF(NC.EQ.2) GO TO 60
      CALL COOR2(N,M,KK,X,XBAR,STD,RX,R,B,D,T,A,V,H,F,C,Z,
     1           XSTAND,ROTPC,XY,ICHAR1,PC,XZ,
     2            XMIN,XMAX,YMIN,YMAX,MAM,NOBS,
     3           ITITLE,ISUBTI,MAM1,X1MIN,X1MAX)
      NC=NC+1
C-----------------------------------------------------------------
C     THE FOLLOWING STATEMENTS VERIFY THE GENERATED COEF.
C     BY ESTIMATING THE DEPENDENT VARIABLE FROM HISTORIC DATA.
C-----------------------------------------------------------------
      WRITE(6,600)
      DO 40 I=1,N
      EST1=0.0
      DO 30 J=1,M
 30   EST1=D(J)*X(I,J)+EST1
      EST1=EST1+BSUBO
      WRITE(6,601) I,EST1
 40   CONTINUE
600   FORMAT('1',5X,'OBS',3X,'ESTIMATED',/)
601   FORMAT(6X,1I3,1F11.3)
C     ANALYSIS OF RESIDUALS
C
      CALL ANLRSD(NOBS,ITITLE,S,SS)
      IF(NC.EQ.2) GO TO 50
 60   CONTINUE
      NSP=NSP2
      FIN=0.0
      FOUT=0.0
      DO 70 I=1,N
      WRITE(11) (X(I,J),J=1,M),XY(I)
 70   CONTINUE
```

```
      REWIND 11
      CALL MLFEG4(MAM,XMIN,XMAX,YMIN,YMAX,ICHAR1,NOBS,
     1           ITITLE,ISUBTI,MAM1,X1MIN,X1MAX)
      FFWIND 11
      CALL ANLRSD(NOBS,ITITLE,S,SS)
      STOP
      REP
        SUBROUTINE ANLRSD(NOBS,ITITLE,S,SS)
C
C       THIS SUBROUTINE PERFORMS ANALYSIS OF RESIDUALS
C       FOR THE CALIBRATED REGRESSION MODEL
C
C       1-KOLMOGOROV-SMIRNOV TEST
C       2-TENDANCY TO CLUSTER IN ORDER OF MAGNITUDE
C       3-TENDANCY TO CLUSTER IN DIRECTION
C       4-CHECK FOR PREDOMINANCE OF UPWARD OR DOWNWARD TRENDS
C
        COMMON/PAUL2/SIGB
        INTEGER ITITLE(20)
        REAL SIGB(58),S(NOBS),SS(NOBS)
C
C
        SUM1=0.0
        SUM2=0.0
        DO 100 I=1,NOBS
        S(I)=SIGB(I)
        SUM1=SUM1+S(I)
        SUM2=SUM2+S(I)*S(I)
 100    CONTINUE
        U=SUM1/NOBS
        S1=SQRT(SUM2/NOBS-U**2)
        CALL KOLMO4(S,NOBS,Z,PROB,1,U,S1,IER)
C
        IF(IER.EQ.1)GO TO 200
C
        ABSZ=Z/SQRT(FLOAT(NOBS))
C
        WRITE(6,600)ITITLE,U,S1,NOBS,Z,PROB,ABSZ
        GO TO 300
 200    WRITE(6,601)
 300    CONTINUE
        DO 400 I=1,NOBS
        S(I)=SIGB(I)
 400    SS(I)=S(I)
        CALL RAND(NOBS,S,SS,ITITLE)
        RETURN
 600    FORMAT('1',5X,20A4,/,6X,'KOLMOGOROV-SMIRNOV TEST',/,
     1  6X,'MEAN=',F10.3,/,6X,'S.D.=',F10.3,/,6X,'NOBS=',
     2  I4,/,6X,'Z-STAT=',F10.4,/,6X,'PROB=',F10.4,/,
     3  6X,'MAX ABS DEV=',F10.4)
 601    FORMAT('1',5X,'ERROR WITH THE VARIANCE')
        END
        SUBROUTINE COOR2(N,M,KK,X,XBAR,STD,RX,R,B,D,T,A,V,H,
     1                   F,C,Z,XSTAND,ROTPC,XY,ICHAR1,PC,XZ,
```

```
     2                      XMIN,XMAX,YMIN,YMAX,MAM,NOBS,ITITLE,
     3                      ISUBTI,MAM1,X1MIN,X1MAX)
      REAL*8 XBAR(M),STD(M),RX(M,M),R(M,M),B(M),D(M),
     * T(M),A(M,M),TV(51),H(M),F(M),C(KK),Z(M,M),
     * ROTPC(N,M),V(M,M),XSTAND(N,M),PC(N,M)
      REAL*4 X(N,M),XY(N),XZ(M),B1(58),SIGB(58)
      INTEGER ID(60),ITITLE(20)
      COMMON/PAUL/ID,NIN,NVIN,NVAR,NTPAN,NCONS,FIN,FOUT,IRES,IANNER,
     1            NSP,BSUBO,B1
      COMMON/PAUL2/SIGB
C
C
      IO=1
C
C     CAILING OF SSP SUROUTINE CORRE , WHERE
C     X=INPUT DATA MATRIX
C     XBAR=COMPUTED MEAN
C     STD=COMPUTED STANDARD DEVIATION
C     R=CORRELATION MATRIX
C
C     COMPUTATION OF THE NUMBER OF PAGES FOR THE MATRIX.
      NP=M/8
      IF(MOD(M,8).GT.0)NP=NP+1
      CALL CORRE(N,M,IO,X,XBAR,STD,RX,R,B,D,T)
      WRITE(6,600)
      WRITE(6,601) (I,XBAR(I),STD(I),I=1,M)
C     WRITE (6,602)
C     WRITE(6,603) ((RX(I,J),J=1,M),I=1,M)
      J1=1
      J2=8
      DO 3 K=1,NP
      WRITE (6,602)
      IF(J2.GT.M) J2=M
      DO 2 I=1,M
    2 WRITE(6,603) (RX(I,J),J=J1,J2)
      J1=J2+1
      J2=J1+7
    3 CONTINUE
      K=0
      DO 5 J=1,M
      DO 5 I=1,M
      K=K+1
      IF(K.GT.((M*(M+1))/2))GO TO 10
      C(K)=R(I,J)
    5 CONTINUE
   10 CONTINUE
      DO 15 I=1,M
      DO 15 J=1,M
   15 Z(I,J)=0.0
      K=0
      DO 20 J=1,M
      DO 20 I=1,J
      K=K+1
      IF(K.GT.((M*(M+1))/2))GO TO 25
```

```
      Z(I,J)=C(K)
      Z(J,I)=Z(I,J)
 20   CONTINUE
 25   CONTINUE
      NP=M/13
      IF(MOD(M,13).GT.0)NP=NP+1
      J1=1
      J2=13
      DO 33 K=1,NP
      WRITE (6,604)
      IF(J2.GT.M)  J2=M
      DO 30 I=1,M
 30   WRITE(6,605) (Z(I,J),J=J1,J2)
      J1=J2+1
      J2=J1+12
 33   CONTINUE
C     INVERSE OF R MATRIX
      CALL MINV(Z,M,D,H,F)
      J1=1
      J2=13
      DO 38 K=1,NP
      WRITE(6,606)
      IF(J2.GT.M)  J2=M
      DO 35 I=1,M
 35   WRITE(6,605) (Z(I,J),J=J1,J2)
      J1=J2+1
      J2=J1+12
 38   CONTINUE
C     STANDARDIZE DATA IN X(N,M)  TO XSTAND(N,M)
      DO 40 J=1,M
      DO 40 I=1,N
      XSTAND(I,J)=(X(I,J)-XBAR(J))/STD(J)
 40   CONTINUE
C     WRITE STANDARDIZED DATA
      J1=1
      J2=13
      DO 48 K=1,NP
      WRITE(6,607)
      IF(J2.GT.M)  J2=M
      DO 45 I=1,N
 45   WRITE(6,605) (XSTAND(I,J),J=J1,J2)
      J1=J2+1
      J2=J1+12
 48   CONTINUE
C
C     SUBROUTINE EIGEN COMPUTES BOTH THE
C     EIGENVALUES AND THE EIGEN VECTORS.
C
      MV=0
      CALL EIGEN(R,A,M,MV)
      K=0
      DO 50 J=1,M
      DO 50 I=1,M
      K=K+1
```

```
      IF(K.GT.((M*(M+1))/2))GO TO 55
   50 C(K)=R(I,J)
   55 CONTINUE
      DO 60 I=1,M
      DO 60 J=1,M
   60 RX(I,J)=0.0
      K=0
      DO 65 J=1,M
      DO 65 I=1,J
      K=K+1
   65 RX(I,J)=C(K)
C     MAKES ENGINVALUE MATRIX PERFECTLY SYMMETRIC BY
C     REPLACING NON-DIAGONAL VALUES EQUAL TO 0.0
      DO 70 J=1,M
      DO 70 I=1,M
      IF(RX(I,J).NE.RX(J,I))RX(I,J)=0.0
   70 CONTINUE
      J1=1
      J2=13
      DO 78 K=1,NP
      WRITE(6,608)
      IF(J2.GT.M) J2=M
      DO 75 I=1,M
   75 WRITE(6,605)(RX(I,J),J=J1,J2)
      J1=J2+1
      J2=J1+12
   78 CONTINUE
C     COMPUTE SQRT OF EIGENVALUES D LAMDA
      DO 80 J=1,M
      DO 80 I=1,M
      IF(RX(I,J).NE.0.0)RX(I,J)=DSQRT(RX(I,J))

   80 CONTINUE
C     LISTS MATRIX OF D LAMDA
      J1=1
      J2=13
      DO 88 K=1,NP
      WRITE(6,609)
      IF(J2.GT.M) J2=M
      DO 85 I=1,M
   85 WRITE(6,605)(RX(I,J),J=J1,J2)
      J1=J2+1
      J2=J1+12
   88 CONTINUE
C     LISTS MATRIX OF EIGENVECTORS
      J1=1
      J2=13
      DO 93 K=1,NP
      WRITE(6,610)
      IF(J2.GT.M) J2=M
      DO 90 I=1,M
   90 WRITE(6,605)(A(I,J),J=J1,J2)
      J1=J2+1
      J2=J1+12
```

```
   93   CONTINUE
C       SUBROUTINE TRACE IS USED TO FIND THE
C       NO OF EIGENVALUES TO RETAIN,THAT ARE GRPATER
C       THAN CON.
C
        CON=.001
        CALL TRACE(M,R,CON,K,D)
        WRITE(6,611)K
        WRITE(6,612)
        WRITE(6,613) (D(I),I=1,K)
C
C       LOAD MATRIX COMPUTATION
C       BY USE OF SSP SUBROUTINE
C
        CALL LOAD(M,K,R,A)
C       LISTS FACTOR LOADING MATRIX
        NP1=K/13
        IF(MOD(K,13).GT.0)NP1=NP1+1
        J1=1
        J2=13
        DO 98 K1=1,NP1
        WRITE(6,614)
        IF(J2.GT.K) J2=K
        DO 95 I=1,M
   95   WRITE(6,605) (A(I,J),J=J1,J2)
        J1=J2+1
        J2=J1+12
   98   CONTINUE
C       REGRESSION ON PRINCIPAL COMPONENTS
C           -1  1/2
C       Z=XR  LD
C        CALL GMPRD(A,RX,R,M,M,M)
C        CALL GMPRD(Z,R,V,M,M,M)
C        CALL GMPRD(XSTAND,V,PC,N,M,M)
        CALL GMPRD(XSTAND,A,PC,N,M,M)
        NCOUN=0
  100   CONTINUE
        NCOUN=NCOUN+1
C      TO STOP REGRESSION ON VARIMAX MATRIX, SIMPLY
C      1)PLACE 'GO TO 120' AFTER 'NCOUN=NCOUN+1'
C      2)PLACE 'C' IN FRONT OF 'IF(NCOUN.EQ.1)...'AFTER 'REWIND 11'
C        FOLLOWING THE CALL OF 'MLREG4'.
       GO TO 120
        IF(NCOUN.EQ.1)GO TO 120
C
C       USE OF SSP TO COMPUTE
C       VARIMAX ROTATION OF FACTOR LOADING MATRIX
C
        CALL VARMX(M,K,A,NC,TV,H,F,D)
C       LISTS ROTATED FACTOR LOADING MATRIX
        NP1=K/13
        IF(MOD(K,13).GT.0)NP1=NP1+1
        J1=1
        J2=13
```

```
      DO 113 K1=1,NP1
      WRITE(6,615)
      IF(J2.GT.K) J2=K
      DO 110 I=1,M
  110 WRITE(6,605) (A(I,J),J=J1,J2)
      J1=J2+1
      J2=J1+12
  113 CONTINUE
C     START OF MATRIX MULTIPLICATION
C      *    -1 * 1/2
C     Z =XR   L D           (HAAN PAGE 260)
C
C     ROTATED PRINCIPLE COMPONENTS=ROTPC=XSTAND*Z*A*RX
      CALL GMPRD(A,RX,R,M,M,M)
      CALL GMPRD(Z,R,V,M,M,M)
      CALL GMPRD(XSTAND,V,ROTPC,N,M,M)
C     LISTS MATRIX OF ROTATED PRINCIPLE COMPONENTS
      J1=1
      J2=13
      DO 118 K=1,NP
      WRITE(6,616)
      IF(J2.GT.M) J2=M
      DO 115 I=1,N
  115 WRITE(6,605) (ROTPC(I,J),J=J1,J2)
      J1=J2+1
      J2=J1+12
  118 CONTINUE
  120 CONTINUE
      IF(NCOUN.EQ.1)GO TO 135
      DO 130 I=1,N
      DO 125 J=1,M
  125 XZ(J)=SNGL(ROTPC(I,J))
      WRITE(11) (XZ(J),J=1,M),XY(I)
  130 CONTINUE
      GO TO 150
  135 DO 145 I=1,N
      DO 140 J=1,M
  140 XZ(J)=SNGL(PC(I,J))
      WRITE(11) (XZ(J),J=1,M),XY(I)
  145 CONTINUE
  150 CONTINUE
      REWIND 11
      CALL MLREG4(MAM,XMIN,XMAX,YMIN,YMAX,ICHAR1,NOBS,
     1 ITITLE,ISUBTI,MAM1,X1MIN,X1MAX)
      REWIND 11
C      IF(NCOUN.EQ.1)GO TO 100
C     THE FOLLOWING STATEMENTS CONVERTS THE COEF. FOR
C     THE REGRESSION TO THE ORIGINAL VARIABLE FORMAT.
      DO 155 J=1,M
  155 D(J)=0.0
      BETA=0.0
      DO 160 J=1,NIN
      DO 160 K=1,M
      D(K)=B1(J)*A(K,ID(J))/STD(K)+D(K)
```

```
      BETA=B1(J)*A(K,ID(J))*XBAR(K)/STD(K)+BETA
160   CONTINUE
      BSUBO=BSUBO-BETA
      WRITE(6,617) BSUBO,(K,D(K),K=1,M)
      RETURN
600   FORMAT('1',3X,'VAR',2X,'MEAN',6X,'STAN. DEV.',//)
601   FORMAT(4X,I2,2X,F10.5,6X,F10.5)
602   FORMAT('1',5X,'CROSS PRODUCTS OF DEVIATIONS FROM MEANS',/,
     *6X,40('_'),//)
603   FORMAT(2X,8F15.5)
604   FORMAT('1',3X,'MATRIX OF CORRELATION COEFF.',/,
     *4X,28('_'),//)
605   FORMAT(2X,13F10.5)
606   FORMAT('1',4X,'INVERSION OF CORRELATION MATRIX',///)
607   FORMAT('1',4X,'LIST OF STANDARDIZED DATA',///)
608   FORMAT('1',3X,' EIGENVALUES',//)
609   FORMAT('1',4X,'CHECK OF SQRT OF D LAMDA',///)
610   FORMAT('1',3X,' EIGENVECTORS',//)
611   FORMAT('1',5X,'THERE ARE',I3,' FACTORS',///)
612   FORMAT(' ',5X,'CUMULATIVE % OF EIGENVALUES GREATER OR ',
     *'EQUAL TO CON',///)
613   FORMAT(F15.5)
614   FORMAT('1',10X,'FACTOR LOADING MATRIX',///)
615   FORMAT('1',10X,'VARIMAX ROTATION OF FACTOR LOADING MATRIX',
     *///)
616   FORMAT('1',4X,'ROTATED PRINCIPLE COMPONENTS Z*',///)
617   FORMAT('1',5X,' Y INTERCEPT =',F15.5,/,
     *30(5X,'COFF.(',1I3,')=',1F12.5,/))
      END
C        SUBROUTINE CORRE
C
C        PURPOSE
C           COMPUTE MEANS, STANDARD DEVIATIONS, SUMS OF CROSS-PRODUCTS
C           OF DEVIATIONS, AND CORRELATION COEFFICIENTS.
C
C
C        USAGE
C           CALL CORRE (N,M,IO,X,XBAR,STD,RX,R,B,D,T)
C
C        DESCRIPTION OF PARAMETERS
C           N     - NUMBER OF OBSERVATIONS.
C           M     - NUMBER OF VARIABLES.
C           IO    - OPTION CODE FOR INPUT DATA
C                   0 IF DATA ARE TO BE READ IN FROM INPUT DEVICE IN THE
C                     SPECIAL SUBROUTINE NAMED DATA.  (SEE SUBROUTINES
C                     USED BY THIS SUBROUTINE BELOW.)
C                   1 IF ALL DATA ARE ALREADY IN CORE.
C           X     - IF IO=0, THE VALUE OF X IS 0.0.
C                   IF IO=1, X IS THE INPUT MATRIX (N BY M) CONTAINING
C                            DATA.
C           XBAR  - OUTPUT VECTOR OF LENGTH M CONTAINING MEANS.
C           STD   - OUTPUT VECTOR OF LENGTH M CONTAINING STANDARD
C                   DEVIATIONS.
C           RX    - OUTPUT MATRIX (M X M) CONTAINING SUMS OF CROSS-
C                   PRODUCTS OF DEVIATIONS FROM MEANS.
```

POOR COPY
COPIE DE QUALITEE INFERIEURE


```
C        R      - OUTPUT MATRIX (ONLY UPPER TRIANGULAR PORTION OF THE
C                 SYMMETRIC MATRIX OF M BY M) CONTAINING CORRELATION
C                 COEFFICIENTS.  (STORAGE MODE OF 1)
C        B      - OUTPUT VECTOR OF LENGTH M CONTAINING THE DIAGONAL
C                 OF THE MATRIX OF SUMS OF CROSS-PRODUCTS OF
C                 DEVIATIONS FROM MEANS.
C        D      - WORKING VECTOR OF LENGTH M.
C        T      - WORKING VECTOR OF LENGTH M.
C
C     REMARKS
C        NONE
C
C     SUBROUTINES AND FUNCTION SUBPROGRAMS REQUIRED
C        DATA(M,D) - THIS SUBROUTINE MUST BE PROVIDED BY THE USER.
C                    (1) IF IO=0, THIS SUBROUTINE IS EXPECTED TO
C                        FURNISH AN OBSERVATION IN VECTOR D FROM AN
C                        EXTERNAL INPUT DEVICE.
C                    (2) IF IO=1, THIS SUBROUTINE IS NOT USED BY
C                        CORRE BUT MUST EXIST IN JOB DECK. IF USER
C                        HAS NOT SUPPLIED A SUBROUTINE NAMED DATA,
C                        THE FOLLOWING IS SUGGESTED.
C                             SUBROUTINE DATA
C                             RETURN
C                             END
C
C     METHOD
C        PRODUCT-MOMENT CORRELATION COEFFICIENTS ARE COMPUTED.
C
C     ..................................................................
C
      SUBROUTINE CORRE (N,M,IO,X,XBAR,STD,RX,R,B,D,T)
      DIMENSION X(1),XBAR(1),STD(1),RX(1),R(1),B(1),D(1),T(1)
C
C        ...............................................................
C
C        IF A DOUBLE PRECISION VERSION OF THIS ROUTINE IS DESIRED, THE
C        C IN COLUMN 1 SHOULD BE REMOVED FROM THE DOUBLE PRECISION
C        STATEMENT WHICH FOLLOWS.
C
C     DOUBLE PRECISION XBAR,STD,RX,R,B,T
C
C        THE C MUST ALSO BE REMOVED FROM DOUBLE PRECISION STATEMENTS
C        APPEARING IN OTHER ROUTINES USED IN CONJUNCTION WITH THIS
C        ROUTINE.
C
C        THE DOUBLE PRECISION VERSION OF THIS SUBROUTINE MUST ALSO
C        CONTAIN DOUBLE PRECISION FORTRAN FUNCTIONS.  SQRT AND ABS IN
C        STATEMENT 220 MUST BE CHANGED TO DSQRT AND DABS.
C
C        ...............................................................
C
C     INITIALIZATION
C
      DO 100 J=1,M
```

```
      B(J)=0.0
  100 T(J)=0.0
      K=(M*M+M)/2
      DO 102 I=1,K
  102 R(I)=0.0
      FN=N
      L=0
C
      IF(IO) 105, 127, 105
C
C     DATA ARE ALREADY IN CORE
C
  105 DO 108 J=1,M
      DO 107 I=1,N
      L=L+1
  107 T(J)=T(J)+X(L)
      XBAR(J)=T(J)
  108 T(J)=T(J)/FN
C
      DO 115 I=1,N
      JK=0
      L=I-N
      DO 110 J=1,M
      L=L+N
      D(J)=X(L)-T(J)
  110 B(J)=B(J)+D(J)
      DO 115 J=1,M
      DO 115 K=1,J
      JK=JK+1
  115 R(JK)=R(JK)+D(J)*D(K)
      GO TO 205
C
C     READ OBSERVATIONS AND CALCULATE TEMPORARY
C     MEANS FROM THESE DATA IN T(J)
C
  127 IF(N-M) 130, 130, 135
  130 KK=N
      GO TO 137
  135 KK=M
  137 DO 140 I=1,KK
      CALL DATA (M,D)
      DO 140 J=1,M
      T(J)=T(J)+D(J)
      L=L+1
  140 RX(L)=D(J)
      FKK=KK
      DO 150 J=1,M
      XBAR(J)=T(J)
  150 T(J)=T(J)/FKK
C
C     CALCULATE SUMS OF CROSS-PRODUCTS OF DEVIATIONS
C     FROM TEMPORARY MEANS FOR M OBSERVATIONS
C
      L=0
```

```
      DO 180 I=1,KK
      JK=0
      DO 170 J=1,M
      L=L+1
  170 D(J)=RX(L)-T(J)
      DO 180 J=1,M
      B(J)=B(J)+D(J)
      DO 180 K=1,J
      JK=JK+1
  180 R(JK)=R(JK)+D(J)*D(K)
C
      IF(N-KK) 205, 205, 185
C
C     READ THE REST OF OBSERVATIONS ONE AT A TIME, SUM
C     THE OBSERVATION, AND CALCULATE SUMS OF CROSS-
C     PRODUCTS OF DEVIATIONS FROM TEMPORARY MEANS
C
  185 KK=N-KK
      DO 200 I=1,KK
      JK=0
      CALL DATA (M,D)
      DO 190 J=1,M
      XBAR(J)=XBAR(J)+D(J)
      D(J)=D(J)-T(J)
  190 B(J)=B(J)+D(J)
      DO 200 J=1,M
      DO 200 K=1,J
      JK=JK+1
  200 R(JK)=R(JK)+D(J)*D(K)
C
C     CALCULATE MEANS
C
  205 JK=0
      DO 210 J=1,M
      XBAR(J)=XBAR(J)/FN
C
C     ADJUST SUMS OF CROSS-PRODUCTS OF DEVIATIONS
C     FROM TEMPORARY MEANS
C
      DO 210 K=1,J
      JK=JK+1
  210 R(JK)=R(JK)-B(J)*B(K)/FN
C
C     CALCULATE CORRELATION COEFFICIENTS
C
      JK=0
      DO 220 J=1,M
      JK=JK+J
  220 STD(J)=DSQRT(DABS(R(JK)))
      DO 230 J=1,M
      DO 230 K=J,M
      JK=J+(K*K-K)/2
      L=M*(J-1)+K
      RX(L)=R(JK)
```

```
      L=M*(K-1)+J
      RX(L)=R(JK)
      IF(STD(J)*STD(K)) 225, 222, 225
  222 R(JK)=0.0
      GO TO 230
  225 R(JK)=R(JK)/(STD(J)*STD(K))
  230 CONTINUE
C
C     CALCULATE STANDARD DEVIATIONS
C
      FN=SQRT(FN-1.0)
      DO 240 J=1,M
  240 STD(J)=STD(J)/FN
C
C     COPY THE DIAGONAL OF THE MATRIX OF SUMS OF CROSS-PRODUCTS OF
C     DEVIATIONS FROM MEANS.
C
      L=-M
      DO 250 I=1,M
      L=L+M+1
  250 B(I)=RX(L)
      RETURN
      END
        SUBROUTINE DATA
        RETURN
        END
         SUBROUTINE EIGEN
C
C        PURPOSE
C           COMPUTE EIGENVALUES AND EIGENVECTORS OF A REAL SYMMETRIC
C           MATRIX
C
C        USAGE
C           CALL EIGEN(A,R,N,MV)
C
C        DESCRIPTION OF PARAMETERS
C           A - ORIGINAL MATRIX (SYMMETRIC), DESTROYED IN COMPUTATION.
C               RESULTANT EIGENVALUES ARE DEVELOPED IN DIAGONAL OF
C               MATRIX A IN DESCENDING ORDER.
C           R - RESULTANT MATRIX OF EIGENVECTORS (STORED COLUMNWISE,
C               IN SAME SEQUENCE AS EIGENVALUES)
C           N - ORDER OF MATRICES A AND R
C           MV- INPUT CODE
C                   0   COMPUTE EIGENVALUES AND EIGENVECTORS
C                   1   COMPUTE EIGENVALUES ONLY (R NEED NOT BE
C                       DIMENSIONED BUT MUST STILL APPEAR IN CALLING
C                       SEQUENCE)
C
C        REMARKS
C           ORIGINAL MATRIX A MUST BE REAL SYMMETRIC (STORAGE MODE=1)
C           MATRIX A CANNOT BE IN THE SAME LOCATION AS MATRIX R
C
C        SUBROUTINES AND FUNCTION SUBPROGRAMS REQUIRED
C           NONE
```

```
C
C        METHOD
C           DIAGONALIZATION METHOD ORIGINATED BY JACOBI AND ADAPTED
C           BY VON NEUMANN FOR LARGE COMPUTERS AS FOUND IN 'MATHEMATICAL
C           METHODS FOR DIGITAL COMPUTERS', EDITED BY A. RALSTON AND
C           H.S. WILF, JOHN WILEY AND SONS, NEW YORK, 1962, CHAPTER 7
C
C        ..................................................................
C
      SUBROUTINE EIGEN(A,R,N,MV)
      DIMENSION A(1),R(1)
C
C        ..................................................................
C
C        IF A DOUBLE PRECISION VERSION OF THIS ROUTINE IS DESIRED, THE
C        C IN COLUMN 1 SHOULD BE REMOVED FROM THE DOUBLE PRECISION
C        STATEMENT WHICH FOLLOWS.
C
C     DOUBLE PRECISION A,R,ANORM,ANRMX,THR,X,Y,SINX,SINX2,COSX,
C    1                 COSX2,SINCS,RANGE
C
C        THE C MUST ALSO BE REMOVED FROM DOUBLE PRECISION STATEMENTS
C        APPEARING IN OTHER ROUTINES USED IN CONJUNCTION WITH THIS
C        ROUTINE.
C
C        THE DOUBLE PRECISION VERSION OF THIS SUBROUTINE MUST ALSO
C        CONTAIN DOUBLE PRECISION FORTRAN FUNCTIONS. SQRT IN STATEMENTS
C        40, 68, 75, AND 78 MUST BE CHANGED TO DSQRT. ABS IN STATEMENT
C        62 MUST BE CHANGED TO DABS. THE CONSTANT IN STATEMENT 5 SHOULD
C        BE CHANGED TO 1.0D-12.
C
C        ..................................................................
C
C        GENERATE IDENTITY MATRIX
C
    5 RANGE=1.0D-12
      IF(MV-1) 10,25,10
   10 IQ=-N
      DO 20 J=1,N
      IQ=IQ+N
      DO 20 I=1,N
      IJ=IQ+I
      R(IJ)=0.0
      IF(I-J) 20,15,20
   15 R(IJ)=1.0
   20 CONTINUE
C
C        COMPUTE INITIAL AND FINAL NORMS (ANORM AND ANORMX)
C
   25 ANORM=0.0
      DO 35 I=1,N
      DO 35 J=I,N
      IF(I-J) 30,35,30
   30 IA=I+(J*J-J)/2
```

```
      ANORM=ANORM+A(IA)*A(IA)
   35 CONTINUE
      IF(ANORM) 165,165,40
   40 ANORM=1.414*DSQRT(ANORM)
      ANRMX=ANORM*RANGE/FLOAT(N)
C
C        INITIALIZE INDICATORS AND COMPUTE THRESHOLD , THR
C
      IND=0
      THR=ANORM
   45 THR=THR/FLOAT(N)
   50 L=1
   55 M=L+1
C
C        COMPUTE SIN AND COS
C
   60 MQ=(M*M-M)/2
      LQ=(L*L-L)/2
      LM=L+MQ
   62 IF(DABS(A(LM))-THR) 130,65,65
   65 IND=1
      LL=L+LQ
      MM=M+MQ
      X=0.5*(A(LL)-A(MM))
   69 Y=-A(LM)/DSQRT(A(LM)*A(LM)+X*X)
      IF(X) 70,75,75
   70 Y=-Y
   75 SINX=Y/DSQRT(2.0*(1.0+(DSQRT(1.0-Y*Y))))
      SINX2=SINX*SINX
   78 COSX=DSQRT(1.0-SINX2)
      COSX2=COSX*COSX
      SINCS=SINX*COSX
C
C        ROTATE L AND M COLUMNS
C
      ILQ=N*(L-1)
      IMQ=N*(M-1)
      DO 125 I=1,N
      IQ=(I*I-I)/2
      IF(I-L) 80,115,80
   80 IF(I-M) 85,115,90
   85 IM=I+MQ
      GO TO 95
   90 IM=M+IQ
   95 IF(I-L) 100,105,105
  100 IL=I+LQ
      GO TO 110
  105 IL=L+IQ
  110 X=A(IL)*COSX-A(IM)*SINX
      A(IM)=A(IL)*SINX+A(IM)*COSX
      A(IL)=X
  115 IF(MV-1) 120,125,120
  120 ILR=ILQ+I
      IMR=IMQ+I
```

```
      X=R(ILR)*COSX-R(IMR)*SINX
      R(IMR)=R(ILR)*SINX+R(IMR)*COSX
      R(ILR)=X
  125 CONTINUE
      X=2.0*A(LM)*SINCS
      Y=A(LL)*COSX2+A(MM)*SINX2-X
      X=A(LL)*SINX2+A(MM)*COSX2+X
      A(LM)=(A(LL)-A(MM))*SINCS+A(LM)*(COSX2-SINX2)
      A(LL)=Y
      A(MM)=X
C
C        TESTS FOR COMPLETION
C
C        TEST FOR M = LAST COLUMN
C
  130 IF(M-N) 135,140,135
  135 M=M+1
      GO TO 60
C
C        TEST FOR L = SECOND FROM LAST COLUMN
C
  140 IF(L-(N-1)) 145,150,145
  145 L=L+1
      GO TO 55
  150 IF(IND-1) 160,155,160
  155 IND=0
      GO TO 50
C
C        COMPARE THRESHOLD WITH FINAL NORM
C
  160 IF(THR-ANRMX) 165,165,45
C
C        SORT EIGENVALUES AND EIGENVECTORS
C
  165 IQ=-N
      DO 185 I=1,N
      IQ=IQ+N
      LL=I+(I*I-I)/2
      JQ=N*(I-2)
      DO 185 J=I,N
      JQ=JQ+N
      MM=J+(J*J-J)/2
      IF(A(LL)-A(MM)) 170,185,185
  170 X=A(LL)
      A(LL)=A(MM)
      A(MM)=X
      IF(MV-1) 175,185,175
  175 DO 180 K=1,N
      ILR=IQ+K
      IMR=JQ+K
      X=R(ILR)
      R(ILR)=R(IMR)
  180 R(IMR)=X
  185 CONTINUE
```

```
      RETURN
      END
C        SUBROUTINE GMPRD
C
C        PURPOSE
C           MULTIPLY TWO GENERAL MATRICES TO FORM A RESULTANT GENERAL
C           MATRIX
C
C        USAGE
C           CALL GMPRD(A,B,R,N,M,L)
C
C        DESCRIPTION OF PARAMETERS
C           A - NAME OF FIRST INPUT MATRIX
C           B - NAME OF SECOND INPUT MATRIX
C           R - NAME OF OUTPUT MATRIX
C           N - NUMBER OF ROWS IN A
C           M - NUMBER OF COLUMNS IN A AND ROWS IN B
C           L - NUMBER OF COLUMNS IN B
C
C        REMARKS
C           ALL MATRICES MUST BE STORED AS GENERAL MATRICES
C           MATRIX R CANNOT BE IN THE SAME LOCATION AS MATRIX A
C           MATRIX R CANNOT BE IN THE SAME LOCATION AS MATRIX B
C           NUMBER OF COLUMNS OF MATRIX A MUST BE EQUAL TO NUMBER OF ROW
C           OF MATRIX B
C
C        SUBROUTINES AND FUNCTION SUBPROGRAMS REQUIRED
C           NONE
C
C        METHOD
C           THE M BY L MATRIX B IS PREMULTIPLIED BY THE N BY M MATRIX A
C           AND THE RESULT IS STORED IN THE N BY L MATRIX R.
C
C     ..................................................................
C
      SUBROUTINE GMPRD(A,B,R,N,M,L)
      DIMENSION A(1),B(1),R(1)
      DOUBLE PRECISION A,B,R
C
      IR=0
      IK=-M
      DO 10 K=1,L
      IK=IK+M
      DO 10 J=1,N
      IR=IR+1
      JI=J-N
      IB=IK
      R(IR)=0
      DO 10 I=1,M
      JI=JI+N
      IB=IB+1
   10 R(IR)=R(IR)+A(JI)*B(IB)
      RETURN
      END
```

```
      SUBROUTINE KOLMO4
C
C
C     PURPOSE
C        TESTS THE DIFFERENCE BETWEEN EMPIRICAL AND THEORETICAL
C        DISTRIBUTIONS USING THE KOLMOGOROV-SMIRNOV TEST
C
C     USAGE
C        CALL KOLMO4(X,N,Z,PROB,IFCOD,U,S,IER)
C
C     DESCRIPTION OF PARAMETERS
C        X    - INPUT VECTOR OF N INDEPENDENT OBSERVATIONS. ON
C               RETURN FROM KOLMO4, X HAS BEEN SORTED INTO A
C               MONOTONIC NON-DECREASING SEQUENCE.
C        N    - NUMBER OF OBSERVATIONS IN X
C        Z    - OUTPUT VARIABLE CONTAINING THE GREATEST VALUE WITH
C               RESPECT TO X OF SQRT(N)*ABS(FN(X)-F(X)) WHERE
C               F(X) IS A THEORETICAL DISTRIBUTION FUNCTION AND
C               FN(X) AN EMPIRICAL DISTRIBUTION FUNCTION.
C        PROB - OUTPUT VARIABLE CONTAINING THE PROBABILITY OF
C               THE STATISTIC BEING GREATER THAN OR EQUAL TO Z IF
C               THE HYPOTHESIS THAT X IS FROM THE DENSITY UNDER
C               CONSIDERATION IS TRUE.  E.G., PROB = 0.05 IMPLIES
C               THAT ONE CAN REJECT THE NULL HYP. THAT THE SET
C               X IS FROM THE DENSITY UNDER CONSIDERATION WITH 5%
C               PROBABILITY OF BEING INCORRECT. PROB=1.-
C               SMIRN(Z)
C        IFCOD- A CODE DENOTING THE PARTICULAR THEORETICAL
C               PROB. DISTRIBUTION FUNCTION BEING CONSIDERED.
C               =1---F(X) IS THE NORMAL PDF.
C               =2---F(X) IS THE EXPONENTIAL PDF.
C               =3---F(X) IS THE CAUCHY PDF.
C               =4---F(X) IS THE UNIFORM PDF.
C               =5---F(X) IS USER SPECIFIED.
C        U    - WHEN IFCOD IS 1 OR 2, U IS THE MEAN OF THE DENSITY
C               GIVEN ABOVE.
C               WHEN IFCOD IS 3, U IS THE MEDIAN OF THE CAUCHY
C               DENSITY.
C               WHEN IFCOD IS 4, U IS THE LEFT ENDPOINT OF THE
C               UNIFORM DENSITY.
C               WHEN IFCOD IS 5, U IS USER SPECIFIED.
C        S    - WHEN IFCOD IS 1 OR 2 , S IS THE STAN. DEV. OF
C               DENSITY GIVEN ABOVE, AND SHOULD BE POSITIVE.
C               WHEN IFCOD IS 3, U - S SPECIFIES THE FIRST QUARTILE
C               OF THE CAUCHY DENSITY. S SHOULD BE NON-ZERO.
C               IF IFCOD IS 4, S IS THE RIGHT ENDPOINT OF THE UNIFORM
C               DENSITY. S SHOULD BE GREATER THAN U.
C               IF IFCOD IS 5, S IS USER SPECIFIED.
C        IER  - ERROR INDICATOR WHICH IS NON-ZERO IF S VIOLATES ABOVE
C               CONVENTIONS. ON RETURN NO TEST HAS BEEN MADE, AND X
C               AND Y HAVE BEEN SORTED INTO MONOTONIC NON-DECREASING
C               SEQUENCES. IER IS SET TO ZERO ON ENTRY TO KOLMO4.
C               IER IS CURRENTLY SET TO ONE IF THE USER-SUPPLIED PDF
C               IS REQUESTED FOR TESTING. THIS SHOULD BE CHANGED
```

```
C            (SEE REMARKS) WHEN SOME PDF IS SUPPLIED BY THE USER.
C
C        REMARKS
C        N SHOULD BE GREATER THAN OR EQUAL TO 100. (SEE THE
C        MATHEMATICAL DESCRIPTION GIVEN FOR THE PROGRAM SMIRN,
C        CONCERNING ASYMPTOTIC FORMULAE)  ALSO, PROBABILITY LEVELS
C        DETERMINED BY THIS PROGRAM WILL NOT BE CORRECT IF THE
C        SAME SAMPLES ARE USED TO ESTIMATE PARAMETERS FOR THE
C        CONTINUOUS DISTRIBUTIONS WHICH ARE USED IN THIS TEST.
C        SEE THE MATHEMATICAL DESCRIPTION FOR THIS PROGRAM)
C        F(X) SHOULD BE A CONTINUOUS FUNCTION.
C        ANY USER SUPPLIED CUMULATIVE PROBABILITY DISTRIBUTION
C        SHOULD BE CODED BEGINNING WITH STATEMENT 26 BELOW,
C        AND SHOULD RETURN TO STATEMENT 27.
C
C        DOUBLE PRECISION USAGE---IT IS DOUBTFUL THAT THE USER WILL
C        WISH TO PERFORM THIS TEST USING DOUBLE PRECISION ACCURACY.
C        IF ONE WISHES TO COMMUNICATE WITH KOLMO4 IN A DOUBLE
C        PRECISION PROGRAM, HE SHOULD CALL THE FORTRAN SUPPLIED
C        PROGRAM SNGL(X) PRIOR TO CALLING KOLMO4, AND CALL THE
C        FORTRAN SUPPLIED PROGRAM DBLE(X) AFTER EXITING FROM KOLMO4.
C        (NOTE THAT SUBROUTINE SMIRN DOES HAVE DOUBLE PRECISION
C        CAPABILITY AS SUPPLIED BY THIS PACKAGE.)
C
C
C        SUBROUTINES AND FUNCTION SUBPROGRAMS REQUIRED
C        SMIRN, NDTR, AND ANY USER SUPPLIED SUBROUTINES REQUIRED.
C
C        METHOD
C        FOR REFERENCE, SEE (1) W. FELLER--ON THE KOLMOGOROV-SMIRNOV
C        LIMIT THEOREMS FOR EMPIRICAL DISTRIBUTIONS--
C        ANNALS OF MATH. STAT., 19, 1948.  177-189,
C        (2) N. SMIRNOV--TABLE FOR ESTIMATING THE GOODNESS OF FIT
C        OF EMPIRICAL DISTRIBUTIONS--ANNALS OF MATH. STAT., 19,
C        1948.  279-281.
C        (3) R. VON MISES--MATHEMATICAL THEORY OF PROBABILITY AND
C        STATISTICS--ACADEMIC PRESS, NEW YORK,1964.  490-493,
C        (4) B.V. GNEDENKO--THE THEORY OF PROBABILITY--CHELSEA
C        PUBLISHING COMPANY, NEW YORK, 1962.  384-401.
C
C..................................................................
C
      SUBROUTINE KOLMO4(X,N,Z,PROB,IFCOD,U,S,IER)
      DIMENSION X(1)
C
C        NON DECREASING ORDERING OF X(I)'S  (DUBY METHOD)
C
      IER=0
      DO 5 I=2,N
      IF(X(I)-X(I-1))1,5,5
    1 TEMP=X(I)
      IM=I-1
      DO 3 J=1,IM
      L=I-J
```

```
      IF(TEMP-X(L))2,4,4                                              KLMO1110
    2 X(L+1)=X(L)                                                     KLMO1120
    3 CONTINUE                                                        KLMO1130
      X(1)=TEMP                                                       KLMO1140
      GO TO 5                                                         KLMO1150
    4 X(L+1)=TEMP                                                     KLMO1160
    5 CONTINUE                                                        KLMO1170
C                                                                     KLMO1180
C     COMPUTES MAXIMUM DEVIATION DN IN ABSOLUTE VALUE BETWEEN         KLMO1190
C     EMPIRICAL AND THEORETICAL DISTRIBUTIONS                         KLMO1200
C                                                                     KLMO1210
      NM1=N-1                                                         KLMO1220
      XN=N                                                            KLMO1230
      DN=0.0                                                          KLMO1240
      FS=0.0                                                          KLMO1250
      IL=1                                                            KLMO1260
    6 DO 7 I=IL,NM1                                                   KLMO1270
      J=I                                                             KLMO1280
      IF(X(J)-X(J+1))9,7,9                                            KLMO1290
    7 CONTINUE                                                        KLMO1300
    8 J=N                                                             KLMO1310
    9 IL=J+1                                                          KLMO1320
C     USE OF GUMBEL'S PROBABILITY PLOTTING FUNCTION                   KLMO1330
      FS=FLOAT(J)/(XN+1.0)                                            KLMO1340
      IF(IFCOD-2)10,13,17                                             KLMO1350
   10 IF(S)11,11,12                                                   KLMO1360
   11 IER=1                                                           KLMO1370
      GO TO 29                                                        KLMO1380
   12 Z =(X(J)-U)/S                                                   KLMO1390
      CALL NDTR(Z,Y,D)                                                KLMO1400
      GO TO 27                                                        KLMO1410
   13 IF(S)11,11,14                                                   KLMO1420
   14 Z=(X(J)-U)/S+1.0                                                KLMO1430
      IF(Z)15,15,16                                                   KLMO1440
   15 Y=0.0                                                           KLMO1450
      GO TO 27                                                        KLMO1460
   16 Y=1.-EXP(-Z)                                                    KLMO1470
      GO TO 27                                                        KLMO1480
   17 IF(IFCOD-4)18,20,26                                             KLMO1490
   18 IF(S)19,11,19                                                   KLMO1500
   19 Y=ATAN((X(J)-U)/S)*0.3183099+0.5                                KLMO1510
      GO TO 27                                                        KLMO1520
   20 IF(S-U)11,11,21                                                 KLMO1530
   21 IF(X(J)-U)22,22,23                                              KLMO1540
   22 Y=0.0                                                           KLMO1550
      GO TO 27                                                        KLMO1560
   23 IF(X(J)-S)25,25,24                                              KLMO1570
   24 Y=1.0                                                           KLMO1580
      GO TO 27                                                        KLMO1590
   25 Y=(X(J)-U)/(S-U)                                                KLMO1600
      GO TO 27                                                        KLMO1610
   26 IER=1                                                           KLMO1620
      GO TO 29                                                        KLMO1630
   27 ES=ABS(Y-FS)                                                    KLMO1640
```

```
      DN=AMAX1(DN,ES)
      IF(IL-N)6,8,28
C
C     COMPUTES Z=DN*SQRT(N)   AND  PROBABILITY
C
   28 Z=DN*SQRT(XN)
      CALL SMIRN(Z,PROB)
      PROB=1.0-PROB
   29 RETURN
      END
C
C     ..................................................................
C
C        SUBROUTINE LOAD
C
C        PURPOSE
C           COMPUTE A FACTOR MATRIX (LOADING) FROM EIGENVALUES AND
C           ASSOCIATED EIGENVECTORS.  THIS SUBROUTINE NORMALLY OCCURS
C           IN A SEQUENCE OF CALLS TO SUBROUTINES CORRE, EIGEN, TRACE,
C           LOAD, AND VARMX IN THE PERFORMANCE OF A FACTOR ANALYSIS.
C
C        USAGE
C           CALL LOAD (M,K,R,V)
C
C        DESCRIPTION OF PARAMETERS
C           M     - NUMBER OF VARIABLES.
C           K     - NUMBER OF FACTORS. K MUST BE GREATER THAN OR EQUAL
C                   TO 1 AND LESS THAN OR EQUAL TO M.
C           R     - A MATRIX (SYMMETRIC AND STORED IN COMPRESSED FORM
C                   WITH ONLY UPPER TRIANGLE BY COLUMN IN CORE) CON-
C                   TAINING EIGENVALUES IN DIAGONAL.  EIGENVALUES ARE
C                   ARRANGED IN DESCENDING ORDER, AND FIRST K
C                   EIGENVALUES ARE USED BY THIS SUBROUTINE.  THE ORDER
C                   OF MATRIX R IS M BY M.  ONLY M*(M+1)/2 ELEMENTS ARE
C                   IN STORAGE.  (STORAGE MODE OF 1)
C           V     - WHEN THIS SUBROUTINE IS CALLED, MATRIX V (M X M)
C                   CONTAINS EIGENVECTORS COLUMNWISE.  UPON RETURNING TO
C                   THE CALLING PROGRAM, MATRIX V CONTAINS A FACTOR
C                   MATRIX (M X K).
C
C        REMARKS
C           NONE
C
C        SUBROUTINES AND FUNCTION SUBPROGRAMS REQUIRED
C           NONE
C
C        METHOD
C           NORMALIZED EIGENVECTORS ARE CONVERTED TO THE FACTOR PATTERN
C           BY MULTIPLYING THE ELEMENTS OF EACH VECTOR BY THE SQUARE
C           ROOT OF THE CORRESPONDING EIGENVALUE.
C
C     ..................................................................
C
      SUBROUTINE LOAD (M,K,R,V)
```

```
      DIMENSION P(1),V(1)
C
C        ...............................................................
C
C        IF A DOUBLE PRECISION VERSION OF THIS ROUTINE IS DESIRED, THE
C        C IN COLUMN 1 SHOULD BE REMOVED FROM THE DOUBLE PRECISION
C        STATEMENT WHICH FOLLOWS.
C
      DOUBLE PRECISION P,V,SQ
C
C        THE C MUST ALSO BE REMOVED FROM DOUBLE PRECISION STATEMENTS
C        APPEARING IN OTHER ROUTINES USED IN CONJUNCTION WITH THIS
C        ROUTINE.
C
C        THE DOUBLE PRECISION VERSION OF THIS SUBROUTINE MUST ALSO
C        CONTAIN DOUBLE PRECISION FORTRAN FUNCTIONS.  SQRT IN STATEMENT
C        150 MUST BE CHANGED TO DSQRT.
C
C        ...............................................................
C
      L=0
      JJ=0
      DO 160 J=1,K
      JJ=JJ+J
  150 SQ=DSQRT(P(JJ))
      DO 160 I=1,M
      L=L+1
  160 V(L)=SQ*V(L)
      RETURN
      END
C        SUBROUTINE MINV
C
C        PURPOSE
C           INVERT A MATRIX
C
C        USAGE
C           CALL MINV(A,N,D,L,M)
C
C        DESCRIPTION OF PARAMETERS
C           A - INPUT MATRIX, DESTROYED IN COMPUTATION AND REPLACED BY
C               RESULTANT INVERSE.
C           N - ORDER OF MATRIX A
C           D - RESULTANT DETERMINANT
C           L - WORK VECTOR OF LENGTH N
C           M - WORK VECTOR OF LENGTH N
C
C        REMARKS
C           MATRIX A MUST BE A GENERAL MATRIX
C
C        SUBROUTINES AND FUNCTION SUBPROGRAMS REQUIRED
C           NONE
C
C        METHOD
C           THE STANDARD GAUSS-JORDAN METHOD IS USED. THE DETERMINANT
```

```
C        IS ALSO CALCULATED. A DETERMINANT OF ZERO INDICATES THAT
C        THE MATRIX IS SINGULAR.
C
C     ..................................................................
C
      SUBROUTINE MINV(A,N,D,L,M)
      DIMENSION A(1),L(1),M(1)
C
C     ..................................................................
C
C        IF A DOUBLE PRECISION VERSION OF THIS ROUTINE IS DESIRED, THE
C        C IN COLUMN 1 SHOULD BE REMOVED FROM THE DOUBLE PRECISION
C        STATEMENT WHICH FOLLOWS.
C
      DOUBLE PRECISION A,D,BIGA,HOLD
C
C        THE C MUST ALSO BE REMOVED FROM DOUBLE PRECISION STATEMENTS
C        APPEARING IN OTHER ROUTINES USED IN CONJUNCTION WITH THIS
C        ROUTINE.
C
C        THE DOUBLE PRECISION VERSION OF THIS SUBROUTINE MUST ALSO
C        CONTAIN DOUBLE PRECISION FORTRAN FUNCTIONS. ABS IN STATEMENT
C        10 MUST BE CHANGED TO DABS.
C
C     ..................................................................
C
C        SEARCH FOR LARGEST ELEMENT
C
      D=1.0
      NK=-N
      DO 80 K=1,N
      NK=NK+N
      L(K)=K
      M(K)=K
      KK=NK+K
      BIGA=A(KK)
      DO 20 J=K,N
      IZ=N*(J-1)
      DO 20 I=K,N
      IJ=IZ+I
   10 IF(DABS(BIGA)-DABS(A(IJ))) 15,20,20
   15 BIGA=A(IJ)
      L(K)=I
      M(K)=J
   20 CONTINUE
C
C        INTERCHANGE ROWS
C
      J=L(K)
      IF(J-K) 35,35,25
   25 KI=K-N
      DO 30 I=1,N
      KI=KI+N
      HOLD=-A(KI)
```

```
      JI=KI-K+J
      A(KI)=A(JI)
   30 A(JI) =HOLD
C
C        INTERCHANGE COLUMNS
C
   35 I=M(K)
      IF(I-K)  45,45,38
   38 JP=N*(I-1)
      DO 40 J=1,N
      JK=NK+J
      JI=JP+J
      HOLD=-A(JK)
      A(JK)=A(JI)
   40 A(JI) =HOLD
C
C        DIVIDE COLUMN BY MINUS PIVOT (VALUE OF PIVOT ELEMENT IS
C        CONTAINED IN BIGA)
C
   45 IF(BIGA) 48,46,48
   46 D=0.0
      RETURN
   48 DO 55 I=1,N
      IF(I-K) 50,55,50
   50 IK=NK+I
      A(IK)=A(IK)/(-BIGA)
   55 CONTINUE
C
C        REDUCE MATRIX
C
      DO 65 I=1,N
      IK=NK+I
      HOLD=A(IK)
      IJ=I-N
      DO 65 J=1,N
      IJ=IJ+N
      IF(I-K) 60,65,60
   60 IF(J-K) 62,65,62
   62 KJ=IJ-I+K
      A(IJ)=HOLD*A(KJ)+A(IJ)
   65 CONTINUE
C
C        DIVIDE ROW BY PIVOT
C
      KJ=K-N
      DO 75 J=1,N
      KJ=KJ+N
      IF(J-K) 70,75,70
   70 A(KJ)=A(KJ)/BIGA
   75 CONTINUE
C
C        PRODUCT OF PIVOTS
C
      D=D*BIGA
```

```
C
C       REPLACE PIVOT BY RECIPROCAL
C
        A(KK)=1.0/BIGA
   80   CONTINUE
C
C       FINAL ROW AND COLUMN INTERCHANGE
C
        K=N
  100   K=(K-1)
        IF(K) 150,150,105
  105   I=L(K)
        IF(I-K) 120,120,108
  108   JQ=N*(K-1)
        JR=N*(I-1)
        DO 110 J=1,N
        JK=JQ+J
        HOLD=A(JK)
        JI=JR+J
        A(JK)=-A(JI)
  110   A(JI)=HOLD
  120   J=M(K)
        IF(J-K) 100,100,125
  125   KI=K-N
        DO 130 I=1,N
        KI=KI+N
        HOLD=A(KI)
        JI=KI-K+J
        A(KI)=-A(JI)
  130   A(JI)=HOLD
        GO TO 100
  150   RETURN
        END
      SUBROUTINE MLREG4(MAM,XMIN,XMAX,YMIN,YMAX,ICHAR1,NOBS,
     * ITITLE,ISUBTI,MAM1,X1MIN,X1MAX)
C       NOTE THIS IS MLREG2 RENAMED MLREG4 FOR LOADING PURPOSES ONLY
C     360 STEPWISE MULTIPLE REGRESSION PROGRAM, 3/14/66
C     PHASES 1 AND 2 CAN BE OVERLAID TO CONSERVE CORE. THE STEPS TO
C     READY PHASES 1 AND 2 FOR OVERLAY ARE:
C        1. SET UP A COMMON AREA CONSISTING OF RIJ,XBAR,SIGMA,FIN,
C        FOUT,OBS,NVAR,NOBS,NINDV.
C        2. SET SIGMA AND DATA EQUIVALENT IN PHASE 2.
C        3. REMOVE STATEMENT 101-2 FROM PHASE 1 AND INSERT IT
C        BEHIND DIMENSION COMMENTS CARD IN PHASE 2.
C     TO MODIFY THE PROGRAM TO OPERATE ON A DISK SYSTEM MAKE THE
C     FOLLOWING CHANGES TO THE SOURCE DECK:
C        1. REMOVE STATEMENT 630 PLUS 1.
C        2. REMOVE THE C1 FROM COMMENT C1.
C        3. REMOVE THE C2 FROM COMMENTS CARD C2  AND INSERT IT
C        FOLLOWING STATEMENT 101.
C        4. REPLACE STATEMENT 870 WITH COMMENTS CARD C3 AFTER
C        REMOVING THE C3.
C        5. REPLACE STATEMENT 610 WITH COMMENTS CARD C4 AFTER
C        REMOVING THE C4.
```

```
C        6. REPLACE STATEMENT 67 PLUS 2 WITH COMMENTS CARD C5 AFTER
C          REMOVING THE C5.
C1      DEFINE FILE 1(1000,60,U,IFA)
C2      IFA=1
C3870 WRITE(1'IFA) (DATA(J),J=1,NVAR)
C4610 IFA=1
C5      READ(1'IFA) (DATA(I),I=1,NVAR)
C       PHASE 1. TRANSFORM ORIGINAL DATA, COMPUTE AND PRINT MEANS,
C       STANDARD DEVIATIONS, AND SIMPLE CORRELATION COEFFICIENTS.
C       DIMENSIONS
         COMMON/PAUL/ID,NIN,NVIN,NVAR,NTRAN,NCONS,FIN,FOUT,IRES,IANNEE,
       *        NSP,BSUBO,B
          COMMON/PAUL2/SIGB
        DIMENSIONDATA(58),CONST(12),ITRAN(60),JTRAN(60),KTRAN(60)
        DIMENSIONLTRAN(60),X1(58),Y1(58),ITITLE(20)
        REAL*8 FIJ(58,58),XBAR(58),SIGMA(58)
        INTEGER AID(20)
        DIMENSIONSIGB(58),B(58),ID(60),FMT(20)
       *
C       EQUIVALENCES
        EQUIVALENCE(SIGMA(1),DATA(1))
C       STATEMENT LABEL 101 IS NOT REFERENCED. IT MARKS THE FIRST
C       EXECUTABLE STATEMENT OF THE SOURCE PROGRAM.
C       ICOM IS FIXED DECIMAL REPRESENTATION OF ALPHABETIC COMMA.
  101 ICOM=1799372864
        REWIND3
C       READ I.D.
C        READ IN VARIABLE FORMAT
        IF(FIN-FOUT)1020,690,690
690     IF(NTRAN)1000,730,700
C       READ TRANSFORMATION CARDS
  700 READ(5,71)(ITRAN(I),JTRAN(I),KTRAN(I),LTRAN(I),I=1,NTRAN)
71      FORMAT(36I2)
       *
        IF(NCONS)1000,730,720
C       READ CONSTANT CARD
  720 READ(5,72)(CONST(I),I=1,NCONS)
72      FORMAT(12F6.3)
C       INITIALIZE.
  730 OBS=NOBS
        NINDV=NVAR-1
        DO90I=1,NVIN
        XBAR(I)=0.0
        DO90J=1,NVIN
90      FIJ(I,J)=0.0
       *
C        READ DATA, FORM SUMS VECTOR, SUMS OF SQUARES MATRIX.
        DO110I=1,NOBS
        READ(11)(DATA(J),J=1,NVIN)
3       FORMAT(12F6.0)
        IF(NTRAN)1000,860,750
C       TRANSFORMATION OF RAW DATA
  750 DO850M=1,NTRAN
        II=ITRAN(M)
```

```
      JJ=JTRAN(M)
      KK=KTRAN(M)
      LL=LTRAN(M)
      GOTO(760,770,780,790,800,810,820,830,840),TI
     *
C     X(J)=X(K)
  760 DATA(JJ)=DATA(KK)
      GOTO850
C     X(J)=-X(K)
  770 DATA(JJ)=-DATA(KK)
      GOTO850
C     X(J)=LOG X(K)
  780 DATA(JJ)=ALOG(DATA(KK))
      GOTO850
C     X(J)=1/X(K)
  790 DATA(JJ)=1.0/DATA(KK)
      GOTO850
C     X(J)=X(K)+X(L)
  800 DATA(JJ)=DATA(KK)+DATA(LL)
      GOTO850
C     X(J)=X(K)*X(L)
  810 DATA(JJ)=DATA(KK)*DATA(LL)
      GOTO850
C     X(J)=X(K)/X(L)
  820 DATA(JJ)=DATA(KK)/DATA(LL)
      GOTO850
C     X(J)=X(K)+C(L)
  830 DATA(JJ)=DATA(KK)+CONST(LL)
      GOTO850
C     X(J)=X(K)*C(L)
  840 DATA(JJ)=DATA(KK)*CONST(LL)
  850 CONTINUE
  860 IF(IRES)870,880,870
C     WRITE DATA FILE
  870 WRITE(3)(DATA(J),J=1,NVAR)
  880 DO100J=1,NVAR
      XBAR(J)=XBAR(J)+DATA(J)
      DO100K=1,NVAR
  100 RIJ(J,K)=RIJ(J,K)+DATA(J)*DATA(K)
  110 CONTINUE
C     COMPUTE STANDARD DEVIATIONS*SQR ROUTE (OBS-1)
      DO120I=1,NVAR
  120 SIGMA(I)=(RIJ(I,I)-XBAR(I)*XBAR(I)/OBS)**.5
C     COMPUTE CORRELATION MATRIX
      DO130I=1,NVAR
      DO130J=1,NVAR
  130 RIJ(I,J)=(RIJ(I,J)-XBAR(I)*XBAR(J)/OBS)/(SIGMA(I)*SIGMA(J))
     *
C     COMPUTE MEANS AND STANDARD DEVIATIONS
      DO140I=1,NVAR
      XBAR(I)=XBAR(I)/OBS
  140 SIGMA(I)=SIGMA(I)/(OBS-1.0)**.5
     *
C     SKIP TO NEW PAGE, WRITE I.D., AVERAGES, STANDARD DEVIATIONS,
```

```
C     AND SIMPLE CORRELATION MATRIX.
      WRITE(6,65)IANNEE,FIN,FOUT
65    FORMAT('1',I4,2X,'FIN =',F4.1,2X,'FOUT=',F4.1)
      WRITE(6,51)
51    FORMAT('0AVERAGES')
      WRITE(6,52) (I,XBAR(I),ICOM,I=1,NINDV),NVAR,XBAR(NVAR)
52    FORMAT(4(' VAR(',I2,')=',E14.7,A1))
      WRITE(6,53)
53    FORMAT('0STANDARD DEVIATIONS')
      WRITE(6,52) (I,SIGMA(I),ICOM,I=1,NINDV),NVAR,SIGMA(NVAR)
      WRITE(6,55)
55    FORMAT('0SIMPLE CORRELATION COEFFICIENTS')
      DO150I=1,NINDV
150   WRITE(6,56) (I,J,RIJ(I,J),ICOM,J=I,NINDV),I,NVAR,RIJ(I,NVAR)
56    FORMAT(4(' VARS(',I2,',',I2,')=',F7.3,A1))
C     PHASE 2. PERFORM STEPWISE CALCULATIONS AND PRINT RESULTS.
C     DIMENSIONS
C     INITIALIZE
      DO190I=1,NVAR
      SIGB(I)=0.0
190   B(I)=0.0
      NENT=0
      DF=OBS-1.0
      NSTEP=-1
C     TRANSFORM SIGMA VECTOR FROM STANDARD DEVIATIONS TO SQUARE
C     ROOTS OF SUMS OF SQUARES.
      DO310I=1,NVAR
310   SIGMA(I)=SIGMA(I)*(OBS-1.0)**.5
C     BEGIN STEP NUMBER NSTEP.
200   NSTEP=NSTEP+1
      STDEE=((RIJ(NVAR,NVAR)/DF)**.5)*SIGMA(NVAR)
      DF=DF-1.0
      IF(DF) 1010,1010,205
205   VMIN=0.0
      VMAX=0.0
      NIN=0
C     FIND MINIMUM VARIANCE CONTRIBUTION OF VARIABLES IN REGRESSION
C     EQUATION. FIND MAXIMUM VARIANCE CONTRIBUTION OF VARIABLES
C      NOT IN REGRESSION EQUATION.
      DO300I=1,NINDV
      IF(RIJ(I,I)-.001) 300,300,210
210   VI=RIJ(I,NVAR)*RIJ(NVAR,I)/RIJ(I,I)
      IF(VI) 240,300,220
220   IF(VI-VMAX) 300,300,230
230   VMAX=VI
      NMAX=I
      GOTO300
240   NIN=NIN+1
      ID(NIN)=I
C     COMPUTE REGRESSION COEFFICIENT AND ITS STANDARD DEVIATION.
      B(NIN)=RIJ(I,NVAR)*SIGMA(NVAR)/SIGMA(I)
      SIGB(NIN)=(STDEE*RIJ(I,I)**.5)/SIGMA(I)
      IF(VMIN) 250,260,1000
250   IF(VI-VMIN) 300,300,260
```

```
260   VMIN=VI
      NMIN=I
300   CONTINUE
      IF(NIN)1000,460,400
C     COMPUTE CONSTANT TERM.
  400 BSUBO=XBAR(NVAR)
      DO410I=1,NIN
      J=ID(I)
410   BSUBO=BSUBO-B(I)*XBAR(J)
      IF(NENT)1000,480,420
C     OUTPUT FOR VARIABLE ADDED
  420 WRITE(6,57)NSTEP,K
57    FORMAT('0STEP NUMBER ',I2,10X,'ENTER VARIABLE ',I2)
  425 WRITE(6,58)STDEE
58    FORMAT(' STANDARD ERROR OF ESTIMATE=',F11.3)
      R=(1.-RIJ(NVAR,NVAR))**.5
      WRITE(6,59)R
59    FORMAT(' MULTIPLE CORRELATION COEFFICIENT =',F6.3)
     *
      IDFN=OBS-DF-2.0
      IDFD=DF+1.0
      F=(SIGMA(NVAR)**2-(STDEE**2)*(DF+1.0))/((OBS-DF-2.0)*STDEE**2)
      WRITE(6,66)IDFN,IDFD,F
66    FORMAT(' GOODNESS OF FIT,F(',I4,',',I4,')=',F8.4)
     *
      WRITE(6,60)BSUBO
60    FORMAT(' CONSTANT TERM=',F12.4)
      WRITE(6,61)
61    FORMAT('0VAR     COEFF      STD DEV     T VALUE      BETA COEFF')
     *
      WRITE(6,62)
62    FORMAT('                               COEFF')
      DO430I=1,NIN
      J=ID(I)
      T=B(I)/SIGB(I)
430   WRITE(6,63)ID(I),B(I),SIGB(I),T,RIJ(J,NVAR)
63    FORMAT(' ',I3,F10.4,F12.4,F10.4,F12.4)
C     COMPUTE F LEVEL FOR MINIMUM VARIANCE CONTRIBUTION VARIABLE
C     IN REGRESSION EQUATION.
C       THIS INSERT STOPS THE REGRESSION AT THE DESIRED STEP-NSP.
        IF(NSTEP.EQ.NSP)GO TO 600
      FLEVL=VMIN*DF/RIJ(NVAR,NVAR)
      IF(FOUT+FLEVL)460,460,450
C     INITIALIZE FOR REMOVAL OF VARIABLE K FROM EQUATION.
  450 K=NMIN
      NENT=0
      DF=DF+2.0
      GOTO500
C     COMPUTE F LEVEL FOR MAXIMUM VARIANCE CONTRIBUTION VARIABLE
C     NOT IN EQUATION.
  460 FLEVL=VMAX*DF/(RIJ(NVAR,NVAR)-VMAX)
      IF(FLEVL-FIN)600,600,470
C     INITIALIZE FOR ENTRY OF VARIABLE K INTO EQUATION.
  470 K=NMAX
```

```
      NENT=K
      GOTO500
C     OUTPUT FOR VARIABLE DELETED
  480 WRITE(6,64)NSTEP,K
64    FORMAT('OSTEP NUMBER ',I2,10X,'DELETE VARIABLE ',I2)
      GOTO425
C     UPDATE MATRIX
  500 DO540I=1,NVAR
      IF(I-K)510,540,510
510   DO530J=1,NVAR
      IF(J-K)520,530,520
520   RIJ(I,J)=RIJ(I,J)-RIJ(I,K)*RIJ(K,J)/RIJ(K,K)
530   CONTINUE
  540 CONTINUE
      DO560J=1,NVAR
      IF(J-K)550,560,550
550   RIJ(K,J)=RIJ(K,J)/RIJ(K,K)
560   CONTINUE
      DO580I=1,NVAR
      IF(I-K)570,580,570
570   RIJ(I,K)=-RIJ(I,K)/RIJ(K,K)
580   CONTINUE
      RIJ(K,K)=1.0/RIJ(K,K)
      GOTO200
600   IF(IRES)610,640,610
C     PRINT RESIDUALS
  610 REWIND3
      WRITE(6,67)
67    FORMAT('0 OBS        ACTUAL     ESTIMATE    RESIDUAL')
     *
      DO630K=1,NOBS
      READ(3) (DATA(I),I=1,NVAR)
      EST=BSUBO
      DO620I=1,NIN
      J=ID(I)
620   EST=EST+B(I)*DATA(J)
      RESID=DATA(NVAR)-EST
        SIGB(K)=RESID
        ITRAN(K)=K
      WRITE(6,68) K,DATA(NVAR),EST,RESID
68    FORMAT(' ',I4,3F12.2)
      X1(K)=DATA(NVAR)
      Y1(K)=EST
  630 CONTINUE
      CALL PLOTXY(AID,1,MAM,XMIN,XMAX,YMIN,YMAX,
     *X1,Y1,NOBS,ICHAR1,X2,Y2,N2,ICHAR2,X3,Y3,N3,ICHAR3)
      CALL PLOT1(ITITLE,ISUBTI,ITRAN,SIGB,NOBS,MAM1,X1MIN,X1MAX)
      REWIND4
C     NORMAL END OF JOB
  640 RETURN
1000  WRITE(6,1231)
1231  FORMAT(' ERRORNEWT,VMIN,NCONS,OR NTRANS IS NEGATIVE.',/,'CHECK FOR
     * CONTROL CARD ERROR')
      RETURN
```

```
 1010 WRITE(6,1232)
 1232 FORMAT(' ERROR. DEGREES OF FREEDOM=0.EITHER ADD MORE DATA',/,'OBSE
     *RVATIONS OR DELETE ONE OR MORE INDEPENDANT VARIABLES.SAMPLE',/,'SI
     *ZE MUST EXCEED NUMBER OF INDEPENDANT VARIABLES BY AT LE AST 2.')
     *
      RETURN
 1020 WRITE(6,1233)
 1233 FORMAT('.ERROR.F LEVEL FOR INCOMING VARIABLE IS LESS ',/,'THAN F L
     *EVEL FOR OUTGOING VARIABLE')
      END
C
C..................................................................
C
C        SUBROUTINE NDTR
C
C        PURPOSE
C           COMPUTES Y = P(X) = PROBABILITY THAT THE RANDOM VARIABLE  U,
C           DISTRIBUTED NORMALLY(0,1), IS LESS THAN OR EQUAL TO X.
C           F(X), THE ORDINATE OF THE NORMAL DENSITY AT X, IS ALSO
C           COMPUTED.
C
C        USAGE
C           CALL NDTR(X,P,D)
C
C        DESCRIPTION OF PARAMETERS
C           X--INPUT SCALAR FOR WHICH P(X) IS COMPUTED.
C           P--OUTPUT PROBABILITY.
C           D--OUTPUT DENSITY.
C
C        REMARKS
C           MAXIMUM ERROR IS 0.0000007.
C
C        SUBROUTINES AND SUBPROGRAMS REQUIRED
C           NONE
C
C        METHOD
C           BASED ON APPROXIMATIONS IN C. HASTINGS, APPROXIMATIONS FOR
C           DIGITAL COMPUTERS, PRINCETON UNIV. PRESS, PRINCETON, N.J.,
C           1955.  SEE EQUATION 26.2.17, HANDBOOK OF MATHEMATICAL
C           FUNCTIONS, ABRAMOWITZ AND STEGUN, DOVER PUBLICATIONS, INC.,
C           NEW YORK.
C
C..................................................................
C
      SUBROUTINE NDTR(X,P,D)
C
      AX=ABS(X)
      T=1.0/(1.0+.2316419*AX)
      D=0.3989423*EXP(-X*X/2.0)
      P = 1.0 - D*T*((((1.330274*T - 1.821256)*T + 1.781478)*T -
     1 0.3565638)*T + 0.3193815)
      IF(X)1,2,2
    1 P=1.0-P
    2 RETURN
```

```
      END
      SUBROUTINE PLOTXY(ITITLE,NX,MAM,XMIN,XMAX,YMIN,YMAX,
     * X1,Y1,N1,ICHAR1,X2,Y2,N2,ICHAR2,X3,Y3,N3,ICHAR3)
      INTEGER IPLOT(101,51)
      INTEGER IBLANK/' '/,IDOT/'.'/,IPLUS/'+'/
      REAL XAXIS(11),YAXIS(6)
      INTEGER ITITLE(20),ICHAR1,ICHAR2,ICHAR3
      REAL X1(1),Y1(1),X2(1),Y2(1),X3(1),Y3(1)
      DO 5 IX=1,101
      DO 5 IY=1,51
      IPLOT(IX,IY)=IBLANK
      IF(IX.EQ.101.OR.IY.EQ.51) GO TO 4
      IF(IX.NE.1.AND.IY.NE.1) GO TO 5
    4 IPLOT(IX,IY)=IDOT
      IF(MOD(IX-1,10).EQ.0.AND.MOD(IY-1,10).EQ.0) IPLOT(IX,IY)=IPLUS
    5 CONTINUE
      IF(MAM.EQ.0)      GO TO 21
      XMIN=X1(1)
      XMAX=X1(1)
      YMIN=Y1(1)
      YMAX=Y1(1)
   21 DO 22 I=1,N1
      IF(X1(I).LT.XMIN) XMIN=X1(I)
      IF(X1(I).GT.XMAX) XMAX=X1(I)
      IF(Y1(I).LT.YMIN) YMIN=Y1(I)
   22 IF(Y1(I).GT.YMAX) YMAX=Y1(I)
      IF(NX.EQ.1) GO TO 25
      DO 23 I=1,N2
      IF(X2(I).LT.XMIN) XMIN=X2(I)
      IF(X2(I).GT.XMAX) XMAX=X2(I)
      IF(Y2(I).LT.YMIN) YMIN=Y2(I)
   23 IF(Y2(I).GT.YMAX) YMAX=Y2(I)
      IF(NX.EQ.2) GO TO 25
      DO 24 I=1,N3
      IF(X3(I).LT.XMIN) XMIN=X3(I)
      IF(X3(I).GT.XMAX) XMAX=X3(I)
      IF(Y3(I).LT.YMIN) YMIN=Y3(I)
   24 IF(Y3(I).GT.YMAX) YMAX=Y3(I)
   25 XSCALE=(XMAX-XMIN)/100
      YSCALE=(YMAX-YMIN)/50
      XAINT=XSCALE*10
      YAINT=YSCALE*10
      XAXIS(1)=XMIN
      YAXIS(1)=YMIN
      DO 26 M=2,11
   26 XAXIS(M)=XAXIS(M-1)+XAINT
      DO 27 M=2,6
   27 YAXIS(M)=YAXIS(M-1)+YAINT
      DO 35 I=1,N1
      IX=((X1(I)-XMIN)/XSCALE)+0.5
      IY=((Y1(I)-YMIN)/YSCALE)+0.5
   35 IPLOT(IX+1,IY+1)=ICHAR1
      IF(NX.EQ.1) GO TO 50
      DO 40 I=1,N2
```

POOR COPY
COPIE DE QUALITEE INFERIEURE


```
      IX=((X2(I)-XMIN)/XSCALE)+0.5
      IY=((Y2(I)-YMIN)/YSCALE)+0.5
40    IPLOT(IX+1,IY+1)=ICHAR2
      IF(NX.EQ.2) GO TO 50
      DO 45 I=1,N3
      IX=((X3(I)-XMIN)/XSCALE)+0.5
      IY=((Y3(I)-YMIN)/YSCALE)+0.5
45    IPLOT(IX+1,IY+1)=ICHAR3
50    WRITE(6,600) ITITLE
      DO 60 IYY=1,51
      IY=52-IYY
      IF(MOD(IY-1,10).EQ.0) GO TO 55
      WRITE(6,603) (IPLOT(IX,IY),IX=1,101)
      GO TO 60
55    MY=(IY-1)/10+1
      WRITE(6,602) YAXIS(MY),(IPLOT(IX,IY),IX=1,101)
60    CONTINUE
      WRITE(6,601) XAXIS
      RETURN
600   FORMAT('1',19X,20A4,///)
601   FORMAT(3X,11F10.3)
602   FORMAT(F8.1,1X,101A1)
603   FORMAT(9X,101A1)
      END
      SUBROUTINE PLOT1(ITITLE,ISUBTI,ID,X1,N1,MAM1,X1MIN,X1MAX)
C
C-----GRAPH OF ONE VARIABLE WITH RESPECT TO ANOTHER (TIME)
C
C     DESCRIPTION OF PARAMETERS
C     -ITITLE(20)-TITLE OF GRAPH
C     -ISUBTI-VARIABLE NAME
C     -ID(I),I=1,2,...,N1-PERIODS (DATES)
C     -X1(I),I=1,2,...,N1-VALUES OF THE VARIABLE
C     -N1-NUMBER OF OBSERVATIONS
C     -MAM1-IF SET EQUAL TO ZERO, THE SCALES X1MIN AND X1MAX ARE
C            USER SPECIFIED BUT CHECKED BY THE SUBROUTINE. IF GIVEN
C            SCALE VALUES ARE INCORRECT, THE PROGRAM REPLACES THEM.
C            IF SET NOT EQUAL TO ZERO, THE SUBROUTINE COMPUTES
C            APPROPRIATE VALUES FOR THE SCALE OF THE GRAPH.
C     -X1MIN-SMALLEST VALUE ON THE GRAPH
C     -X1MAX-LARGEST VALUE ON THE GRAPH
C
      INTEGER ITITLE(20),ID(1),LINE(101),IB/' '/,IDOT/'.'/,IX/'X'/
      REAL X1(1),AXIS1(11)
      IF(MAM1.EQ.0) GO TO 31
      X1MIN=X1(1)
      X1MAX=X1(1)
   31 DO 32 I=1,N1
      IF(X1(I).LT.X1MIN) X1MIN=X1(I)
   32 IF(X1(I).GT.X1MAX) X1MAX=X1(I)
   34 SCALE1=(X1MAX-X1MIN)/100
      AINT=SCALE1*10
      AXIS1(1)=X1MIN
      DO 36 M=2,11
```

```
   36 AXIS1(M)=AXIS1(M-1)+AINT
      WRITE(6,601)ITITLE,ISUBTI,AXIS1
      WRITE(6,603)ISUBTI
      DO 48 L=1,N1
      DO 40 K=1,101
   40 LINE(K)=IB
      DO 42 K=1,101,10
   42 LINE(K)=IDOT
      K=((X1(L)-X1MIN)/SCALE1)+0.5
      LINE(K+1)=IX
   48 WRITE(6,604)ID(L),LINE,X1(L)
      WRITE(6,605)
      RETURN
  601 FORMAT('1',20A4,//,' ',A4,' ',2F7.2,9F10.2)
  602 FORMAT(' ',A4,' ',2F7.2,9F10.2)
  603 FORMAT(' DATE  ',10('.+++++++++'),'.',A4,'   X ')
  604 FORMAT(' ',I6,101A1,F8.2)
  605 FORMAT(7X,10('.+++++++++'),'.')
      END
C       SUBROUTINE RAND
        SUBROUTINE RAND(N,S,SS,ITITLE)
C
C       PURPOSE
C       RANDOMNESS TESTS
C
C       REFERENCE-STATISTICAL METHODS IN HYDROLOGY-
C                 PROCEEDINGS OF HYDROLOGY SYMPOSIUM NO.5
C                 P. 249-251
C
C

        INTEGER R1,R2,R3,R4,FIRST,SECOND,ITITLE(20)
        DIMENSION S(N),SS(N)
C
C

        T=N
C       TEST NO.1 BEGINS
C       SERIES IS TESTED FOR ANY CLUSTERING TENDANCIES.
C       THIS IS DONE BY CLASSIFYING THE OBSERVATIONS INTO 4
C       QUARTILES.
C       THE NUMBER OF RUNS IN EACH CATEGORY IS COUNTED ( A RUN
C       IS DEFINED AS A GROUP OF CONSECUTIVE OBSERVATIONS
C       IN ANY CATEGORY.
C

        ICOUNT=1
        CALL SORTX(N,SS)
C       SS HAS BEEN SORTED IN DESCENDING ORDER OF MAGNITUDE
        K=N/2
        L=N/4
        M=(3*N)/4
        F050=SS(K+1)
        FF025=SS(L+1)
        FF075=SS(M+1)
        AVGFF=((FF025-F050)+(F050-FF075))/2.0
        F025=F050+AVGFF
```

```
      F075=F050-AVGFF
      F1=0
      F2=0
      F3=0
      F4=0
      DO 80 I=1,N
      IF(S(I)-F025) 10,60,60
   10 IF(S(I)-F050) 20,50,50
   20 IF(S(I)-F075) 30,40,40
   30 FIRST=4
      F4=F4+1
      GO TO 70
   40 FIRST=3
      F3=F3+1
      GO TO 70
   50 FIRST=2
      F2=F2+1
      GO TO 70
   60 FIRST=1
      F1=F1+1
   70 CONTINUE
      IF(I.EQ.1)GO TO 80
      IF(FIRST.NE.SECOND) ICOUNT=ICOUNT+1
   80 SECOND=FIRST
      R1=ICOUNT
      SMSQ=F1**2+F2**2+F3**2+F4**2
      SMCU=F1**3+F2**3+F3**3+F4**3
      RM1=(T*(T+1.)-SMSQ)/T
      STDV1=SQRT((SMSQ*(SMSQ+T*(T+1.))
     1 -2.*T*SMCU-T**3)/((T**2)*(T-1.)))
      STNV1=((R1+0.5)-RM1)/STDV1
C
C     TEST NO. 2 BEGINS
C     SERIES IS TESTED FOR ANY CLUSTERING TENDANCY IN DIRECTION
C     OF MOVEMENT.
C     THE RUNS UP AND DOWN ARE COUNTED IN THIS TEST.
C
      ICOUNT=1
      SN=0.0
      SP=0.0
      DO 120 I=2,N
      K=I-1
      IF(S(I)-S(K))100,110,90
   90 SN=SN+1.
      FIRST=-1
      GO TO 110
  100 SP=SP+1.
      FIRST=1
  110 CONTINUE
      IF(I.EQ.2)GO TO 120
      IF(FIRST.NE.SECOND) ICOUNT=ICOUNT+1
  120 SECOND=FIRST
      R2=ICOUNT
      RM2=(2.*T-1.)/3.
```

```
      STDV2=SQRT((16.*T-29.)/90.)
      STNV2=((R2+0.5)-RM2)/STDV2
C
C     TESTNO. 3 BEGINS
C
C     THIS TEST SHOWS IF THERE IS A PREDOMINANCE OF
C     UPWARD OR DOWNWARD CHANGES.
C
      IF(SN-SP) 130,140,140
  130 SMIN=SN
      GO TO 150
  140 SMIN=SP
  150 CONTINUE
      RM3=(T-1.)/2.
      STDV3=SQRT((T+1.)/12.)
      STNV3=(RM3-(SMIN+0.5))/STDV3
      WRITE(6,270)
      WRITE(6,260) ITITLE
      WRITE(6,160)
      WRITE(6,240)
      WRITE(6,220) F025,F050,F075,FF025,FF075,AVGFF
      WRITE(6,250) F1,F2,F3,F4
      WRITE(6,230) SN,SP,SMIN
      WRITE(6,170)
      I=1
      WRITE(6,180) I,STNV1,R1,RM1,STDV1
      I=I+1
      WRITE(6,180) I,STNV2,R2,RM2,STDV2
      I=I+1
      WRITE(6,180) I,STNV3,SMIN,RM3,STDV3
      WRITE(6,190)
      WRITE(6,200)
      WRITE(6,210)
      RETURN
C
  160 FORMAT(9X,'THREE SERIAL CORRELATION TESTS USING RUNS , ',
     1 'OR CLUSTERING , OF LIKE OBSERVATIONS',/)
  170 FORMAT(20X,12HTEST RESULTS/1H ,20X,12H,,,,,,,,,,,,/1H ,
     1 8HTEST NO.,12X,24HSTANDARD NORMAL VARIABLE,9X,7HR,VALUE,10X,
     2 6HR,MEAN,10X,20HR,STANDARD DEVIATION)
  180 FORMAT(1H ,3X,I1,24X,F6.2,16X,F10.2,6X,F10.2,12X,F10.2)
  190 FORMAT(1H ,5X,5HNOTE./1H ,5X,5H,,,,,/1H ,5X,
     1 'THE STANDARD NORMAL VARIABLE(K) INDICATES THE NUMBER',
     2 ' OF STD. DEVIATIONS THAT THE TEST VALUE IS BELOW THE',
     3 ' EXPECTED MEAN VALUE',/,1H ,'IF THE OBSERVATIONS ARE',
     4 ' INDEPENDENT, VALUES OF K LESS THAN -2.33 OCCUR WITH A PROBA',
     1 'BILITY OF LESS THAN 0.01 FOR A SEQUENCE OF')
  200 FORMAT(1H ,'INDEPENDENT OBSERVATIONS (USE K GREATER THAN',
     1 ' 2.58 FOR TEST 3 - 2-TAIL PROBABILITY).',/,1H ,5X,
     2 'SPECIFICALLY, SUCH VALUES OF K INDICATE-',/,1H ,
     3 'TEST 1- A TENDANCY FOR OBSERVATIONS TO CLUSTER INTO THE ',
     4 'RESPECTIVE CATEGORIES.',/,1H ,'TEST 2- A TENDANCY FOR ',
     5 'DIRECTIONS OF MOVEMENT TO PERSIST (I.E.- CLUSTERING NOT')
  210 FORMAT(1H ,8X,'EXCLUSIVELY DUE TO A FEW LARGE CHANGES.)',/,1H ,
```

```
     1  'TEST 3- AN UPWARD OR DOWNWARD TREND DUE TO A RELATIVE',
     2  ' SHORTAGE OF - OR + SIGNS RESPECTIVELY (TREND DUE TO A ',
     3  'LARGER MOVEMENTS',/,1H ,8X,'IN ONE DIRECTION IS NOT ',
     4  'DETECTED BY THIS TEST).')
 220    FORMAT(1H ,31X,'F025',11X,'F050',11X,'F075',10X,
     1  'FF025',10X,'FF075',10X,'AVGFF',/,
     2  1H ,21X,6F15.2)
 230    FORMAT(1H ,'CHANGES.',5X,F5.1,'=NEGATIVE',5X,F5.1,
     1  '=POSITIVE',5X,F5.1,'=MINIMUM')
 240    FORMAT(1H ,32X,'CATEGORY BOUNDARY',/,1H ,31X,
     2  '-----------------------------------')
 250    FORMAT(1H ,'CATEGORY NO.',13X,1H1,14X,1H2,14X,
     1  1H3,14X,1H4,/,1H ,'OCCURANCE',3X,4I15)
 260    FORMAT(9X,20A4,//)
 270    FORMAT(1H1)
        END
C
C   ...................................................................
C
C        SUBROUTINE SMIRN
C
C        PURPOSE
C           COMPUTES VALUES OF THE LIMITING DISTRIBUTION FUNCTION FOR
C           THE KOLMOGOROV-SMIRNOV STATISTIC.
C
C        USAGE
C           CALL SMIRN(X,Y)
C
C        DESCRIPTION OF PARAMETERS
C           X     - THE ARGUMENT OF THE SMIRN FUNCTION
C           Y     - THE RESULTANT SMIRN FUNCTION VALUE
C
C        REMARKS
C           Y IS SET TO ZERO IF X IS NOT GREATER THAN 0.27, AND IS SET
C           TO ONE IF X IS NOT LESS THAN 3.1.  ACCURACY TESTS WERE MADE
C           REFERRING TO THE TABLE GIVEN IN THE REFERENCE BELOW.
C           TWO ARGUMENTS, X= 0.62, AND X = 1.87 GAVE RESULTS WHICH
C           DIFFER FROM THE SMIRNOV TABLES BY 2.9 AND 1.9 IN THE 5TH
C           DECIMAL PLACE.  ALL OTHER RESULTS SHOWED SMALLER ERRORS,
C           AND ERROR SPECIFICATIONS ARE GIVEN IN THE ACCURACY TABLES
C           IN THIS MANUAL.  IN DOUBLE PRECISION MODE, THESE SAME
C           ARGUMENTS RESULTED IN DIFFERENCES FROM TABLED VALUES BY 3
C           AND 2 IN THE 5TH  DECIMAL PLACE.  IT IS NOTED IN
C           LINDGREN (REFERENCE BELOW) THAT FOR HIGH SIGNIFICANCE LEVELS
C           (SAY, .01 AND .05) ASYMPTOTIC FORMULAS GIVE VALUES WHICH ARE
C           TOO HIGH ( BY 1.5 PER CENT,WHEN N = 80).  THAT IS, AT HIGH
C           SIGNIFICANCE LEVELS, THE HYPOTHESIS OF NO DIFFERENCE WILL BE
C           REJECTED TOO SELDOM USING ASYMPTOTIC FORMULAS.
C
C        SUBROUTINES AND FUNCTION SUBPROGRAMS REQUIRED
C           NONE
C
C        METHOD
C           THE METHOD IS DESCRIBED BY W. FELLER-ON THE KOLMOGOROV-
```

```
C          SMIRNOV LIMIT THEOREMS FOR EMPIRICAL DISTRIBUTIONS- ANNALS
C          OF MATH. STAT., 19, 1948, 177-189, BY N. SMIRNOV--TABLE
C          FOR ESTIMATING THE GOODNESS OF FIT OF EMPIRICAL
C          DISTRIBUTIONS- ANNALS OF MATH. STAT., 19, 1948, 279-281,
C          AND GIVEN IN LINDGREN, STATISTICAL THEORY, THE MACMILLAN
C          COMPANY, N. Y., 1962.
C
C     ..................................................................
C
      SUBROUTINE SMIRN(X,Y)
C     DOUBLE PRECISION X,Q1,Q2,Q4,Q8,Y
C
C        IF A DOUBLE PRECISION VERSION OF THIS ROUTINE IS DESIRED, THE C
C        IN COLUMN ONE OF THE DOUBLE PRECISION CARD ABOVE SHOULD BE
C        REMOVED, AND THE C IN COLUMN ONE OF THE STATEMENTS NUMBERED
C        C   3, C   5, AND C   8 SHOULD BE REMOVED, AND THESE CARDS
C        SHOULD REPLACE THE STATEMENTS NUMBERED 3, 5, AND 8,
C        RESPECTIVELY.  ALL ROUTINES CALLING THIS ROUTINE MUST ALSO
C        PROVIDE DOUBLE PRECISION ARGUMENTS TO THIS ROUTINE.
C
C     ..................................................................
C
      IF(X-.27) 1,1,2
    1 Y=0.0
      GO TO 9
    2 IF(X-1.0) 3,6,6
    3 Q1=EXP(-1.233701/X**2)
C   3 Q1=DEXP(-1.233700550136170/X**2)
      Q2=Q1*Q1
      Q4=Q2*Q2
      Q8=Q4*Q4
      IF(Q8-1.0E-25) 4,5,5
    4 Q8=0.0
    5 Y=(2.506628/X)*Q1*(1.0+Q8*(1.0+Q8*Q8))
C   5 Y=(2.506628274631001/X)*Q1*(1.0D0+Q8*(1.0D0+Q8*Q8))
      GO TO 9
    6 IF(X-3.1) 8,7,7
    7 Y=1.0
      GO TO 9
    8 Q1=EXP(-2.0*X*X)
C   8 Q1=DEXP(-2.0D0*X*X)
      Q2=Q1*Q1
      Q4=Q2*Q2
      Q8=Q4*Q4
      Y=1.0-2.0*(Q1-Q4+Q8*(Q1-Q8))
    9 RETURN
      END
C        SUBROUTINE SORTX
C
         SUBROUTINE SORTX(N,X)
C
C        SUBROUTINE TO SORT AN ARRAY OF REAL VARIATES IN
C        DESCENDING ORDER
C
```

```
      DIMENSION X(1)
      K=(N-1)
      DO 10 L=1,K
      M=(N-L)
      DO 10 J=1,M
      IF(X(J).GE.X(J+1))GO TO 10
      XTEMP=X(J)
      X(J)=X(J+1)
      X(J+1)=XTEMP
   10 CONTINUE
      RETURN
      END
C       SUBROUTINE TRACE
C
C       PURPOSE
C          COMPUTE CUMULATIVE PERCENTAGE OF EIGENVALUES GREATER THAN
C          OR EQUAL TO A CONSTANT SPECIFIED BY THE USER.  THIS SUB-
C          ROUTINE NORMALLY OCCURS IN A SEQUENCE OF CALLS TO SUB-
C          ROUTINES CORRE, EIGEN, TRACE, LOAD, AND VARMX IN THE PER-
C          FORMANCE OF A FACTOR ANALYSIS.
C
C       USAGE
C          CALL TRACE (M,R,CON,K,D)
C
C       DESCRIPTION OF PARAMETERS
C          M      - NUMBER OF VARIABLES.
C          R      - INPUT MATRIX (SYMMETRIC AND STORED IN COMPRESSED
C                   FORM WITH ONLY UPPER TRIANGLE BY COLUMN IN CORE)
C                   CONTAINING EIGENVALUES IN DIAGONAL.  EIGENVALUES ARE
C                   ARRANGED IN DESCENDING ORDER.  THE ORDER OF MATRIX R
C                   IS M BY M.  ONLY M*(M+1)/2 ELEMENTS ARE IN STORAGE.
C                   (STORAGE MODE OF 1)
C          CON    - A CONSTANT USED TO DECIDE HOW MANY EIGENVALUES TO
C                   RETAIN.  CUMULATIVE PERCENTAGE OF EIGENVALUES
C                   WHICH ARE GREATER THAN OR EQUAL TO THIS VALUE IS
C                   CALCULATED.
C          K      - OUTPUT VARIABLE CONTAINING THE NUMBER OF EIGENVALUES
C                   GREATER THAN OR EQUAL TO CON.  (K IS THE NUMBER OF
C                   FACTORS.)
C          D      - OUTPUT VECTOR OF LENGTH M CONTAINING CUMULATIVE
C                   PERCENTAGE OF EIGENVALUES WHICH ARE GREATER THAN
C                   OR EQUAL TO CON.
C
C       REMARKS
C          NONE
C
C       SUBROUTINES AND FUNCTION SUBPROGRAMS REQUIRED
C          NONE
C
C       METHOD
C          EACH EIGENVALUE GREATER THAN OR EQUAL TO CON IS DIVIDED BY M
C          AND THE RESULT IS ADDED TO THE PREVIOUS TOTAL TO OBTAIN
C          THE CUMULATIVE PERCENTAGE FOR EACH EIGENVALUE.
C
C       ..............................................................
```

```
C
      SUBROUTINE TRACE (M,R,CON,K,D)
      DIMENSION R(1),D(1)
C
C        ...............................................................
C
C        IF A DOUBLE PRECISION VERSION OF THIS ROUTINE IS DESIRED, THE
C        C IN COLUMN 1 SHOULD BE REMOVED FROM THE DOUBLE PRECISION
C        STATEMENT WHICH FOLLOWS.
C
      DOUBLE PRECISION R,D
C
C        THE C MUST ALSO BE REMOVED FROM DOUBLE PRECISION STATEMENTS
C        APPEARING IN OTHER ROUTINES USED IN CONJUNCTION WITH THIS
C        ROUTINE.
C
C        ...............................................................
C
      FM=M
      L=0
      DO 100 I=1,M
      L=L+I
  100 D(I)=R(L)
      K=0
C
C     TEST WHETHER I-TH EIGENVALUE IS GREATER
C     THAN OR EQUAL TO THE CONSTANT
C
      DO 110 I=1,M
      IF(D(I)-CON) 120, 105, 105
  105 K=K+1
  110 D(I)=D(I)/FM
C
C     COMPUTE CUMULATIVE PERCENTAGE OF EIGENVALUES
C
  120 DO 130 I=2,K
  130 D(I)=D(I)+D(I-1)
      RETURN
      END
C        SUBROUTINE VARMX
C
C        PURPOSE
C           PERFORM ORTHOGONAL ROTATIONS OF A FACTOR MATRIX.  THIS
C           SUBROUTINE NORMALLY OCCURS IN A SEQUENCE OF CALLS TO SUB-
C           ROUTINES CORRE, EIGEN, TRACE, LOAD, VARMX IN THE PERFORMANCE
C           OF A FACTOR ANALYSIS.
C
C        USAGE
C           CALL VARMX (M,K,A,NC,TV,H,F,D)
C
C        DESCRIPTION OF PARAMETERS
C           M      - NUMBER OF VARIABLES AND NUMBER OF ROWS OF MATRIX A.
C           K      - NUMBER OF FACTORS.
C           A      - INPUT IS THE ORIGINAL FACTOR MATRIX, AND OUTPUT IS
```

```
C                 THE ROTATED FACTOR MATRIX.  THE ORDER OF MATRIX A
C                 IS M X K.
C        NC     - OUTPUT VARIABLE CONTAINING THE NUMBER OF ITERATION
C                 CYCLES PERFORMED.
C        TV     - OUTPUT VECTOR CONTAINING THE VARIANCE OF THE FACTOR
C                 MATRIX FOR EACH ITERATION CYCLE.  THE VARIANCE PRIOR
C                 TO THE FIRST ITERATION CYCLE IS ALSO CALCULATED.
C                 THIS MEANS THAT NC+1 VARIANCES ARE STORED IN VECTOR
C                 TV.  MAXIMUM NUMBER OF ITERATION CYCLES ALLOWED IN
C                 THIS SUBROUTINE IS 50.  THEREFORE, THE LENGTH OF
C                 VECTOR TV IS 51.
C        H      - OUTPUT VECTOR OF LENGTH M CONTAINING THE ORIGINAL
C                 COMMUNALITIES.
C        F      - OUTPUT VECTOR OF LENGTH M CONTAINING THE FINAL
C                 COMMUNALITIES.
C        D      - OUTPUT VECTOR OF LENGTH M CONTAINING THE DIFFERENCES
C                 BETWEEN THE ORIGINAL AND FINAL COMMUNALITIES.
C
C     REMARKS
C        IF VARIANCE COMPUTED AFTER EACH ITERATION CYCLE DOES NOT
C        INCREASE FOR FOUR SUCCESSIVE TIMES, THE SUBROUTINE STOPS
C        ROTATION.
C
C     SUBROUTINES AND FUNCTION SUBPROGRAMS REQUIRED
C        NONE
C
C     METHOD
C        KAISER'S VARIMAX ROTATION AS DESCRIBED IN 'COMPUTER PROGRAM
C        FOR VARIMAX ROTATION IN FACTOR ANALYSIS' BY THE SAME AUTHOR,
C        EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT, VOL XIX, NO. 3,
C        1959.
C
C     ..................................................................
C
      SUBROUTINE VARMX (M,K,A,NC,TV,H,F,D)
      DIMENSION A(1),TV(1),H(1),F(1),D(1)
C
C        ...............................................................
C
C        IF A DOUBLE PRECISION VERSION OF THIS ROUTINE IS DESIRED, THE
C        C IN COLUMN 1 SHOULD BE REMOVED FROM THE DOUBLE PRECISION
C
      DOUBLE PRECISION A,TV,H,F,D,TVLT,CONS,AA,BB,CC,DD,U,T,B,COS4T,
     1                 SIN4T,TAN4T,SINP,COSP,CTN4T,COS2T,SIN2T,COST,SINT
C
C        THE C MUST ALSO BE REMOVED FROM DOUBLE PRECISION STATEMENTS
C        APPEARING IN OTHER ROUTINES USED IN CONJUNCTION WITH THIS
C        ROUTINE.
C
C        THE DOUBLE PRECISION VERSION OF THIS SUBROUTINE MUST ALSO
C        CONTAIN DOUBLE PRECISION FORTRAN FUNCTIONS.  SQRT IN STATEMENTS
C        115, 290, 330, 350, AND 355 MUST BE CHANGED TO DSQRT.  ABS IN
C        STATEMENTS 280, 320, AND 375 MUST BE CHANGED TO DABS.
C
```

```
C     ...............................................................
C
C     INITIALIZATION
C
      EPS=0.00116
      TVLT=0.0
      LL=K-1
      NV=1
      NC=0
      FN=M
      FFN=FN*FN
      CONS=0.7071066
C
C     CALCULATE ORIGINAL COMMUNALITIES
C
      DO 110 I=1,M
      H(I)=0.0
      DO 110 J=1,K
      L=M*(J-1)+I
  110 H(I)=H(I)+A(L)*A(L)
C
C     CALCULATE NORMALIZED FACTOR MATRIX
C
      DO 120 I=1,M
  115 H(I)=DSQRT(H(I))
      DO 120 J=1,K
      L=M*(J-1)+I
  120 A(L)=A(L)/H(I)
      GO TO 132
C
C     CALCULATE VARIANCE FOR FACTOR MATRIX
C
  130 NV=NV+1
      TVLT=TV(NV-1)
  132 TV(NV)=0.0
      DO 150 J=1,K
      AA=0.0
      BB=0.0
      LB=M*(J-1)
      DO 140 I=1,M
      L=LB+I
      CC=A(L)*A(L)
      AA=AA+CC
  140 BB=BB+CC*CC
  150 TV(NV)=TV(NV)+(FN*BB-AA*AA)/FFN
      IF(NV-51) 160, 430, 430
C
C     PERFORM CONVERGENCE TEST
C
  160 IF((TV(NV)-TVLT)-(1.E-7)) 170, 170, 190
  170 NC=NC+1
      IF(NC-3) 190, 190, 430
C
C     ROTATION OF TWO FACTORS CONTINUES UP TO
```

```
C      THE STATEMENT 120.
C
  190 DO 420 J=1,LL
      L1=M*(J-1)
      II=J+1
C
C        CALCULATE NUM AND DEN
C
      DO 420 K1=II,K
      L2=M*(K1-1)
      AA=0.0
      BB=0.0
      CC=0.0
      DD=0.0
      DO 230 I=1,M
      L3=L1+I
      L4=L2+I
      U=(A(L3)+A(L4))*(A(L3)-A(L4))
      T=A(L3)*A(L4)
      T=T+T
      CC=CC+(U+T)*(U-T)
      DD=DD+2.0*U*T
      AA=AA+U
  230 BB=BB+T
      T=DD-2.0*AA*BB/PN
      B=CC-(AA*AA-BB*BB)/PN
C
C        COMPARISON OF NUM AND DEN
C
      IF(T-B) 280, 240, 320
  240 IF((T+B)-EPS) 420, 250, 250
C
C        NUM + DEN IS GREATER THAN OR EQUAL TO THE
C        TOLERANCE FACTOR
C
  250 COS4T=CONS
      SIN4T=CONS
      GO TO 350
C
C        NUM IS LESS THAN DEN
C
  280 TAN4T=DABS(T)/DABS(B)
      IF(TAN4T-EPS) 300, 290, 290
  290 COS4T=1.0/DSQRT(1.0+TAN4T*TAN4T)
      SIN4T=TAN4T*COS4T
      GO TO 350
  300 IF(B) 310, 420, 420
  310 SINP=CONS
      COSP=CONS
      GO TO 400
C
C        NUM IS GREATER THAN DEN
C
  320 CTN4T=DABS(T/B)
```

```
      IF(CTN4T-EPS) 340, 330, 330
  330 SIN4T=1.0/DSQRT(1.0+CTN4T*CTN4T)
      COS4T=CTN4T*SIN4T
      GO TO 350
  340 COS4T=0.0
      SIN4T=1.0
C
C         DETERMINE COS THETA AND SIN THETA
C
  350 COS2T=DSQRT((1.0+COS4T)/2.0)
      SIN2T=SIN4T/(2.0*COS2T)
  355 COST=DSQRT((1.0+COS2T)/2.0)
      SINT=SIN2T/(2.0*COST)
C
C         DETERMINE COS PHI AND SIN PHI
C
      IF(B) 370, 370, 360
  360 COSP=COST
      SINP=SINT
      GO TO 380
  370 COSP=CONS*COST+CONS*SINT
  375 SINP=DABS(CONS*COST-CONS*SINT)
  380 IF(T) 390, 390, 400
  390 SINP=-SINP
C
C         PERFORM ROTATION
C
  400 DO 410 I=1,M
      L3=L1+I
      L4=L2+I
      AA=A(L3)*COSP+A(L4)*SINP
      A(L4)=-A(L3)*SINP+A(L4)*COSP
  410 A(L3)=AA
  420 CONTINUE
      GO TO 130
C
C      DENORMALIZE VARIMAX LOADINGS
C
  430 DO 440 I=1,M
      DO 440 J=1,K
      L=M*(J-1)+I
  440 A(L)=A(L)*H(I)
C
C      CHECK ON COMMUNALITIES
C
      NC=NV-1
      DO 450 I=1,M
  450 H(I)=H(I)*H(I)
      DO 470 I=1,M
      F(I)=0.0
      DO 460 J=1,K
      L=M*(J-1)+I
  460 F(I)=F(I)+A(L)*A(L)
  470 D(I)=H(I)-F(I)
```

POOR COPY
COPIE DE QUALITEE INFERIEURE


```
      RETURN                                                                          VAPM2350
      END                                                                             VAPM2360
//GO.SYSIN DD *
  559.98  829.30  800.80  672.81  596.17  753.90  481.29  803.35  796.09  519.07
  767.42  736.92  892.95  580.65  746.41  692.86  709.98  680.55  733.76  655.60
  730.61  542.00  512.28  671.38  667.91  742.27  502.84  592.38  549.09  739.43
  540.37  672.81  507.74  549.05  649.24  702.09  551.79  549.61  846.18  952.60
  585.33 1013.69  825.93
  62.930  57.920  57.910  39.270  28.000   2.910  -5.210   1.450  19.970  35.260
  45.790   3.430   4.250   3.440   3.280   2.650   2.520   1.980   1.570   2.250
   2.830   1.680  73.030  55.070  56.000  62.630  50.640  44.080  19.160  12.730
   9.110 -99.990 -99.990 -99.990
  65.070  64.230  49.510  40.620  27.630   8.030   1.090   1.090  26.990  34.080
  44.280   5.760   2.330   3.820   3.540   3.450   1.540   2.930   2.240   2.300
   2.510   6.250  66.010  62.300  39.580  55.970  54.370  21.200  13.830   9.790
  37.980 -99.990 -99.990 -99.990
  62.420  58.480  51.880  37.560  18.210   9.370   4.130  12.050  12.320  36.020
  51.710   4.510   3.090   3.480   5.250   2.600   3.470   3.350   2.390   1.420
   1.100   2.520  41.100  31.970  34.290  84.440  50.380  22.920  19.170  13.890
   9.440 -99.990 -99.990 -99.990
  66.220  65.860  52.510  41.340  28.200   9.450   0.530   3.000  17.060  34.550
  48.520   3.790   4.530   3.980   4.780   3.160   1.830   2.080   2.020   2.630
   3.060   1.650  28.920  37.350  55.410  67.930  57.490  26.590  17.330  14.160
  13.250 -99.990 -99.990 -99.990
  64.150  62.120  51.050  43.120  27.250  13.190   0.540   0.800   8.030  31.720
  45.630   4.990   6.530   4.020   1.480   2.200   3.250   1.840   3.390   0.940
   1.620   3.730  66.310  93.400  63.970  46.010  42.190  26.480  16.190  12.170
   8.240 -99.990 -99.990 -99.990
  64.990  63.720  50.760  38.190  23.910  12.730   4.920   4.160  17.720  31.390
  46.990   5.510   3.660   3.420   3.760   1.490   3.350   2.200   1.320   2.450
   2.710   3.120  93.680  52.810  38.770  56.360  46.820  26.890  17.620  12.310
   8.990 -99.990 -99.990 -99.990
  63.310  61.490  53.090  38.240  25.490   6.590  -1.700  10.010  16.040  35.450
  48.720   2.900   3.770   3.870   2.670   3.350   2.920   1.590   0.850   2.470
   3.250   0.760  61.860  39.220  41.220  32.180  31.740  18.350  13.730   9.610
   6.100 -99.990 -99.990 -99.990
  65.820  57.820  52.000  39.290  26.840  12.590  -1.900   7.240  22.860  36.380
  51.020   3.920   4.160   4.970   4.390   4.340   2.730   2.400   2.900   2.130
   0.960   2.960  35.270  20.070  85.020  87.940  59.650  39.740  20.610  14.940
  10.730 -99.990 -99.990 -99.990
  60.220  61.140  51.660  41.890  25.450   1.820  -4.380   0.920  15.070  25.380
  44.430   4.900   3.310   4.790   3.840   2.330   4.390   1.870   2.410   2.390
   2.510   2.100  32.970  24.100  61.460  80.380  79.580  23.640  17.280  13.710
  10.130 -99.990 -99.990 -99.990
  63.610  59.120  51.960  43.610  26.780   5.170   9.850   0.720  13.340  29.290
  48.980   3.980   5.360   1.400   4.150   1.800   2.200   1.430   1.820   1.770
   0.890   3.150  41.270  57.620  38.100  25.140  37.840  17.440  16.000  10.840
   6.200 -99.990 -99.990 -99.990
  63.820  63.760  54.820  38.980  29.390   9.990  -5.100   7.380  23.920  37.000
  46.390   4.680   3.990   4.600   2.790   3.020   2.600   2.850   2.460   2.460
   3.310   3.320  39.860  50.320  60.160  50.050  47.900  20.380  17.500  10.520
  16.670 -99.990 -99.990 -99.990
  62.790  61.650  51.750  38.280  25.530   6.620   0.280  -3.150  23.420  32.250
  43.940   3.790   3.500   4.190   3.890   2.920   2.800   2.260   2.100   2.580
```

POOR COPY
COPIE DE QUALITEE INFERIEURE

3.240	2.720	60.280	35.590	60.290	69.360	50.770	23.300	16.600	17.090
14.320	-99.990	-99.990	-99.990						
60.220	59.420	55.050	43.070	23.830	5.380	0.940	9.270	21.900	23.310
40.510	4.590	2.200	5.500	3.920	3.300	3.620	2.780	3.440	1.930
2.870	4.560	69.170	34.190	34.130	74.720	49.180	22.670	19.000	20.820
15.820	-99.990	-99.990	-99.990						
65.840	63.930	49.850	46.460	27.050	6.280	-2.880	-3.540	10.260	30.790
47.030	4.330	2.690	4.110	1.910	2.430	2.170	1.150	1.810	2.590
1.930	2.820	46.030	40.820	58.260	47.100	31.390	18.690	16.020	12.440
11.390	-99.990	-99.990	-99.990						
63.130	62.410	54.370	41.470	33.070	14.820	3.550	2.210	15.710	35.420
45.160	4.450	3.420	3.290	2.600	3.290	2.390	2.680	1.570	1.960
1.470	3.160	43.170	57.860	46.840	30.070	38.210	21.110	19.080	15.240
12.530	-99.990	-99.990	-99.990						
64.360	61.070	51.740	43.320	21.820	14.360	2.950	0.480	12.070	32.230
48.360	3.270	4.110	4.520	2.770	2.220	1.810	3.080	1.600	1.810
2.110	1.170	61.890	47.430	54.910	52.700	34.360	22.960	22.110	21.420
12.630	-99.990	-99.990	-99.990						
61.780	58.180	49.540	40.070	30.020	12.950	2.610	7.760	21.010	39.510
49.510	4.390	2.710	1.860	3.660	4.550	2.950	3.370	2.090	2.350
2.350	1.420	80.410	34.140	15.640	49.060	62.450	43.520	22.780	17.660
17.300	-99.990	-99.990	-99.990						
48.730	60.080	52.670	42.250	22.710	7.690	-1.750	10.560	22.370	35.600
48.140	4.720	2.600	4.370	3.650	4.240	2.360	2.870	2.060	1.360
1.800	2.330	27.340	35.980	33.350	60.720	59.370	42.320	19.550	17.290
16.740	-99.990	-99.990	-99.990						
67.210	61.500	52.150	37.250	28.200	18.080	4.470	9.940	20.340	36.780
47.580	5.550	4.570	3.630	3.250	1.580	4.070	3.420	1.810	2.830
2.400	1.090	46.970	56.250	57.890	44.480	30.820	45.700	23.970	20.500
26.960	-99.990	-99.990	-99.990						
62.440	60.100	52.180	42.230	33.770	14.330	-6.700	13.780	17.000	30.960
46.410	4.120	1.570	5.210	0.840	2.170	2.930	2.150	1.960	1.550
2.370	4.370	45.460	19.110	60.420	40.760	36.630	35.970	17.570	16.900
15.760	-99.990	-99.990	-99.990						
60.790	58.840	46.400	41.670	27.520	10.200	0.230	4.490	13.240	34.670
48.380	3.870	4.300	4.510	1.960	3.700	3.550	3.160	2.110	4.110
1.190	3.750	38.000	39.000	42.600	51.000	48.100	30.000	19.400	14.000
14.500	21.840	26.300	29.300						
64.580	62.700	49.070	41.810	28.330	5.910	12.180	6.650	9.430	32.440
39.370	4.680	3.570	4.630	2.440	2.310	1.790	1.580	2.180	1.500
1.940	2.870	31.800	20.400	32.700	49.400	54.500	20.600	15.200	12.100
9.200	12.450	16.600	23.300						
58.430	57.140	47.620	42.900	26.250	2.680	-9.520	9.150	19.750	31.480
44.870	4.900	4.360	3.870	2.590	2.220	2.940	1.250	1.990	1.320
1.960	1.780	73.000	69.100	80.900	62.200	30.000	19.300	12.500	11.300
12.600	13.970	17.900	20.700						
61.570	55.890	51.580	42.500	30.340	14.810	8.460	5.900	29.640	37.110
44.040	4.680	2.830	6.180	2.510	3.480	3.150	2.040	2.660	0.980
2.490	3.670	98.000	43.600	71.400	46.000	59.400	38.600	25.600	17.100
16.800	12.190	15.500	16.400						
62.270	58.830	51.800	39.530	26.790	-1.590	4.170	-5.540	10.530	32.820
49.570	4.120	4.860	3.620	3.430	2.670	1.890	2.830	1.010	2.070
1.560	2.230	66.800	50.700	70.400	55.000	30.200	20.400	16.500	14.100
13.000	14.220	18.400	21.900						

66.330	61.520	54.050	39.570	24.160	12.090	3.040	14.600	14.250	31.540
53.390	5.800	3.520	3.030	2.510	3.450	2.370	2.130	3.900	0.990
2.240	2.380	57.300	48.000	38.400	55.000	51.800	26.800	14.800	14.700
12.100	16.000	21.700	26.000						
60.200	61.050	51.760	39.620	30.360	7.670	-6.820	3.870	15.980	[illegible]
45.030	5.340	2.940	3.890	2.160	3.680	1.790	2.010	2.490	2.180
1.010	1.900	61.300	88.600	45.300	45.900	48.700	26.900	16.100	10.800
10.900	13.210	19.300	21.700						
62.810	61.290	57.280	43.270	32.200	15.050	-2.980	-5.560	22.800	33.750
47.850	3.650	6.310	3.830	1.210	2.360	2.540	3.140	2.030	0.640
2.500	4.060	41.900	47.800	53.200	33.900	30.600	20.600	17.600	12.400
10.400	20.320	24.400	22.400						
59.260	61.900	50.980	40.600	25.160	7.670	0.970	-1.180	13.720	33.850
45.760	6.260	3.110	4.200	2.340	1.030	3.140	1.920	1.670	2.060
2.790	1.810	45.500	49.100	49.700	32.100	19.200	15.200	13.700	9.900
10.000	13.460	17.800	20.000						
65.520	56.650	47.970	45.250	30.860	0.550	6.120	6.550	15.590	33.710
48.130	5.340	5.060	3.530	2.760	4.950	1.230	3.190	1.670	3.030
1.100	4.400	66.400	43.500	27.200	25.500	38.000	27.700	15.800	14.700
12.800	15.490	22.400	27.300						
63.030	56.730	49.360	37.480	25.200	8.320	-2.300	1.810	16.410	33.130
45.360	4.740	4.080	3.380	2.460	2.830	3.330	1.740	2.210	0.470
0.700	4.280	61.700	64.800	44.900	52.200	29.300	29.200	21.500	14.500
11.200	13.460	19.700	22.800						
56.590	57.030	49.340	37.130	20.450	9.680	7.510	9.890	21.460	34.460
43.600	4.870	5.510	4.120	4.680	3.220	2.730	3.420	1.230	3.220
1.140	2.590	65.300	78.200	66.900	75.200	36.500	23.200	15.700	14.800
17.100	15.750	21.400	21.400						
61.930	58.980	49.110	39.960	30.280	13.710	4.050	-3.340	10.160	29.210
40.910	3.840	5.750	5.760	4.230	4.730	3.690	2.950	2.420	1.210
1.320	2.580	41.000	66.900	53.200	91.800	52.800	51.200	22.500	15.500
12.700	12.190	17.600	20.400						
64.120	61.390	51.640	40.050	25.790	11.410	-5.140	0.150	19.060	36.920
47.340	4.060	3.990	4.400	5.370	3.600	3.050	1.410	2.470	2.850
1.710	0.620	51.200	49.200	30.700	107.000	75.600	29.500	15.200	13.300
13.700	14.990	21.500	23.500						
61.590	54.900	56.340	44.490	23.980	10.830	11.900	10.980	18.460	29.390
42.620	5.310	3.940	2.380	3.040	2.490	4.650	2.920	1.340	1.090
1.400	2.920	35.900	68.300	48.300	52.500	32.800	22.500	17.800	14.900
12.200	23.620	26.900	28.100						
61.090	62.020	48.200	38.310	30.830	10.010	-5.980	1.010	18.840	33.910
45.350	4.320	4.520	3.930	3.100	3.980	3.520	1.420	2.750	2.020
1.840	2.650	53.500	52.500	47.600	46.400	66.200	32.400	19.700	16.500
12.800	11.180	18.700	26.800						
65.750	62.220	49.800	43.120	28.610	1.450	-2.720	5.340	17.210	32.770
48.380	6.740	3.370	5.350	1.710	2.580	1.660	2.170	3.070	2.640
2.270	2.860	82.400	54.500	66.100	42.700	32.900	19.500	13.500	12.600
11.600	11.180	19.500	27.300						
60.720	58.320	54.210	42.900	23.060	5.330	0.390	-2.150	10.630	30.920
46.940	4.010	6.030	4.000	3.880	2.190	3.550	1.890	2.840	3.720
0.430	1.970	33.300	54.200	52.800	87.500	39.700	20.100	15.200	13.300
12.900	17.780	25.800	32.400						
62.750	57.290	49.700	35.180	23.960	0.770	4.830	-0.560	24.350	33.790
46.740	3.810	5.000	5.050	3.190	1.910	2.900	2.590	2.800	2.030

POOR COPY
COPIE DE QUALITEE INFERIEURE


```
  3.320    3.810  43.600  32.900  40.500  63.700  25.500  17.500  15.200  14.600
 14.700   14.480  20.200  23.600
 64.080   63.820  50.490  42.300  24.700  11.740  -5.900   0.590  12.490  30.030
 40.640    6.020   6.070   3.780   3.270   2.440   4.690   2.610   1.780   2.770
  3.420    4.750  58.500  75.900  50.900  64.700  37.000  24.500  17.100  14.300
 12.900   22.100  27.300  32.800
 63.200   60.520  47.870  34.220  27.790  13.920   2.090   4.350  14.980  30.540
 51.450    4.790   3.740   5.000   3.430   1.890   2.800   2.530   1.580   2.900
  2.050    2.930  44.300  47.100  41.200  56.400  53.700  31.500  20.700  13.200
 13.700   12.700  16.600  21.300
 66.750   61.150  51.450  40.810  28.530   2.570  -3.930   6.650  13.830  33.450
 47.320    5.990   2.900   5.090   2.510   3.480   2.640   2.780   3.320   2.710
  2.560    5.040  48.200  41.700  42.900  42.000  45.600  27.800  16.500  15.400
 15.100   15.240  25.100  24.900
 62.430   61.020  49.030  36.000  22.300  -1.200  -0.860   8.200  26.570  30.920
 49.660    4.530   4.720   2.700   3.800   2.200   3.850   2.530   2.430   1.740
  2.620    1.650  37.900  47.600  52.000  58.000  28.800  18.800  15.400  12.200
 14.300   23.110  26.800  32.000
/*
//GO.FT03F001 DD DSN=&&TEMP1,UNIT=SYSDA,DISP=(NEW,DELETE),
//     SPACE=(TRK,(2,1))
//GO.FT11F001 DD DSN=&&FILE1,UNIT=SYSDA,DISP=(NEW,DELETE),
//    SPACE=(49,(23,1))
```

Appendix C

PLOT OF RESIDUALS OF MODELS 1 THROUGH 15

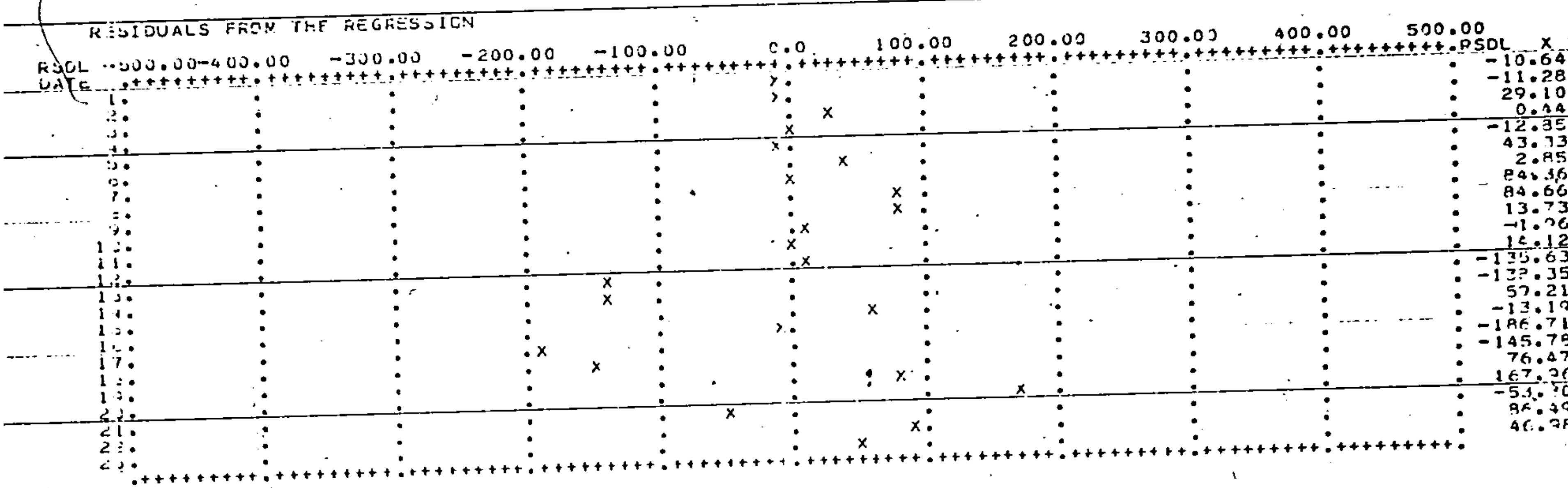


Figure C1: Residuals (BCF) of Model 1.

POOR COPY
COPIE DE QUALITEE INFERIEURE

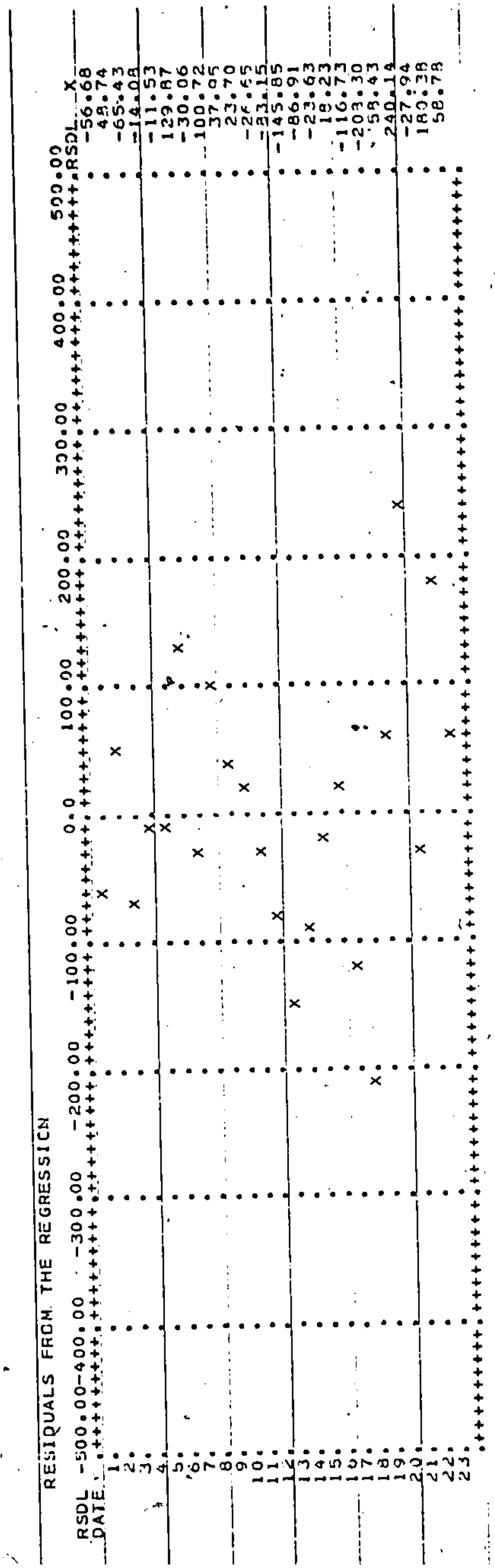


Figure C2: Residuals (BCF) of Model 2.

POOR COPY
COPIE DE QUALITEE INFERIEURE

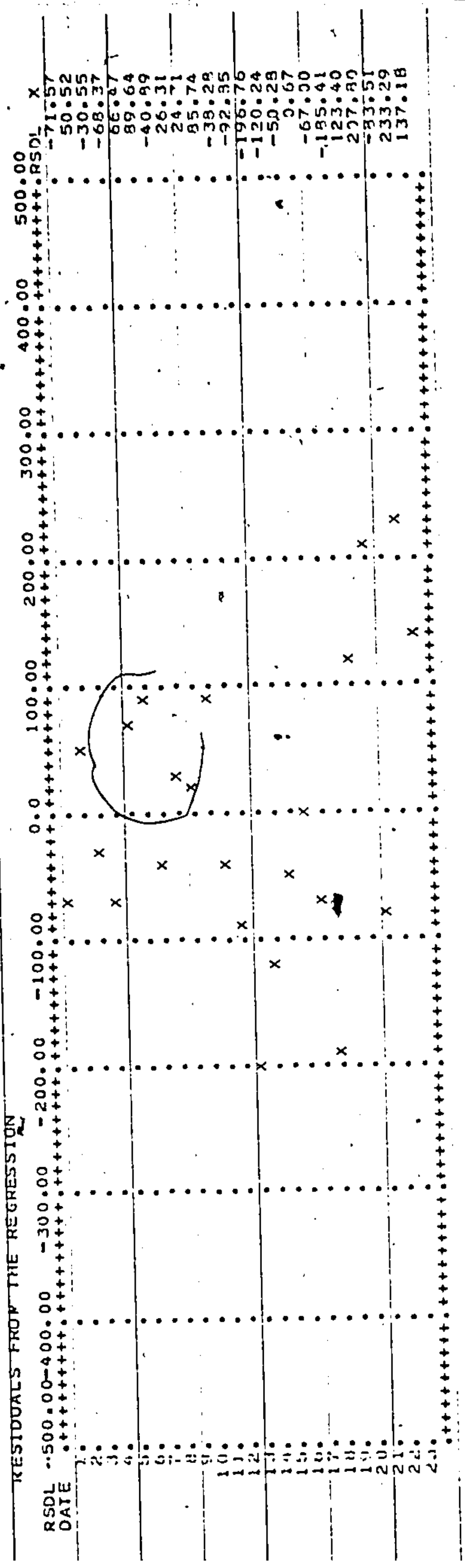


Figure C3: Residuals (BCF) of Model 3.

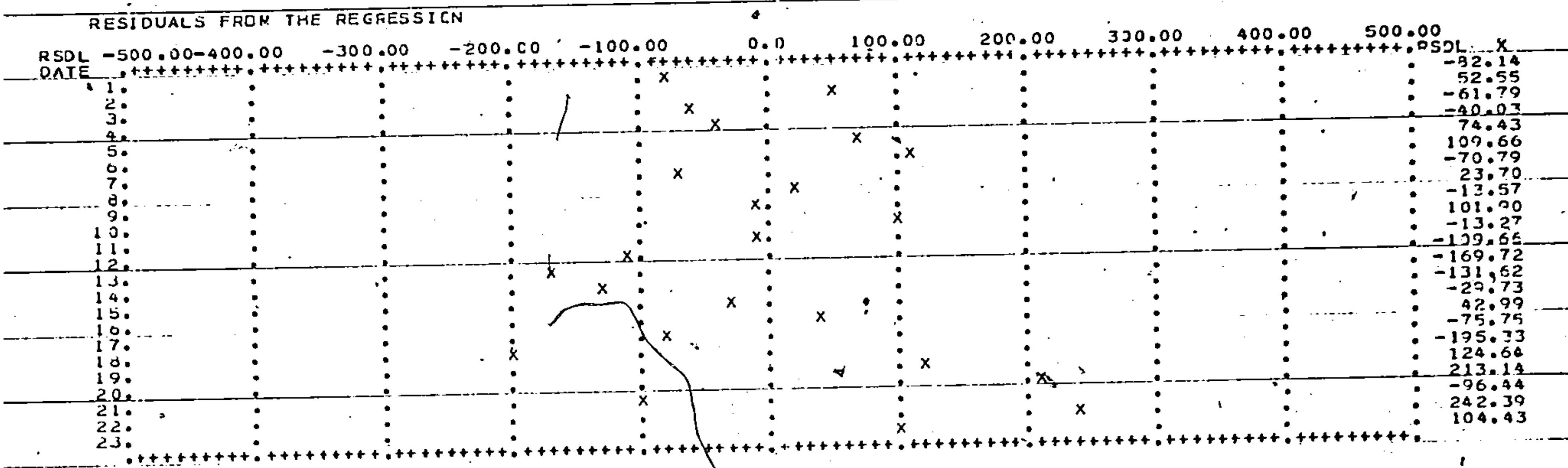


Figure C4: Residuals (BCF) of Model 4.

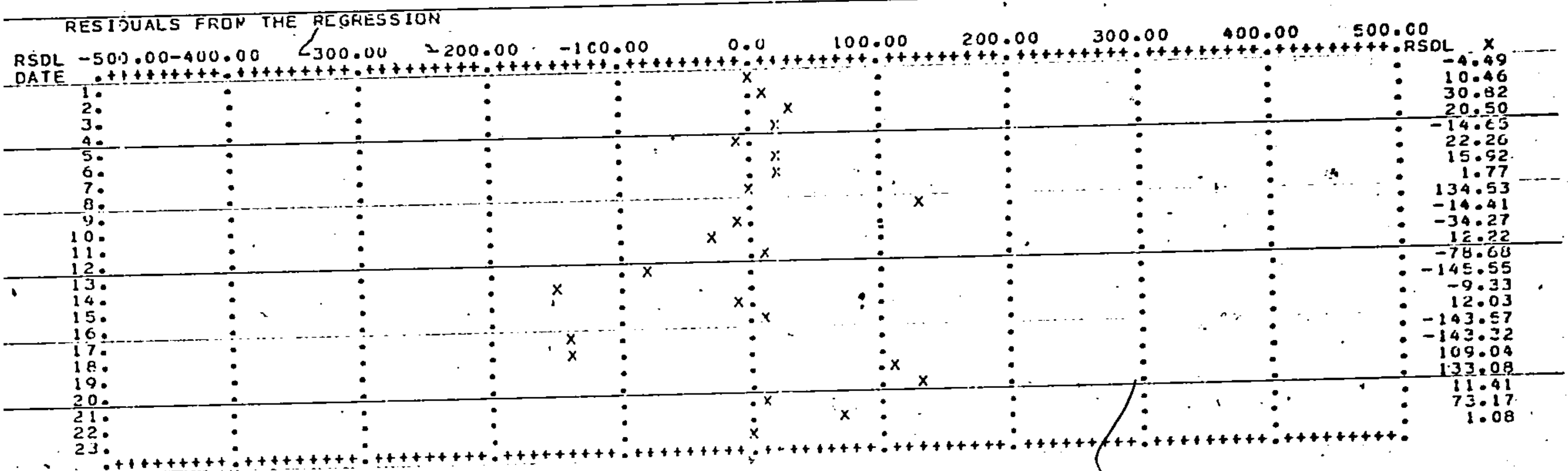


Figure C5: Residuals (BCF) of Model 5.

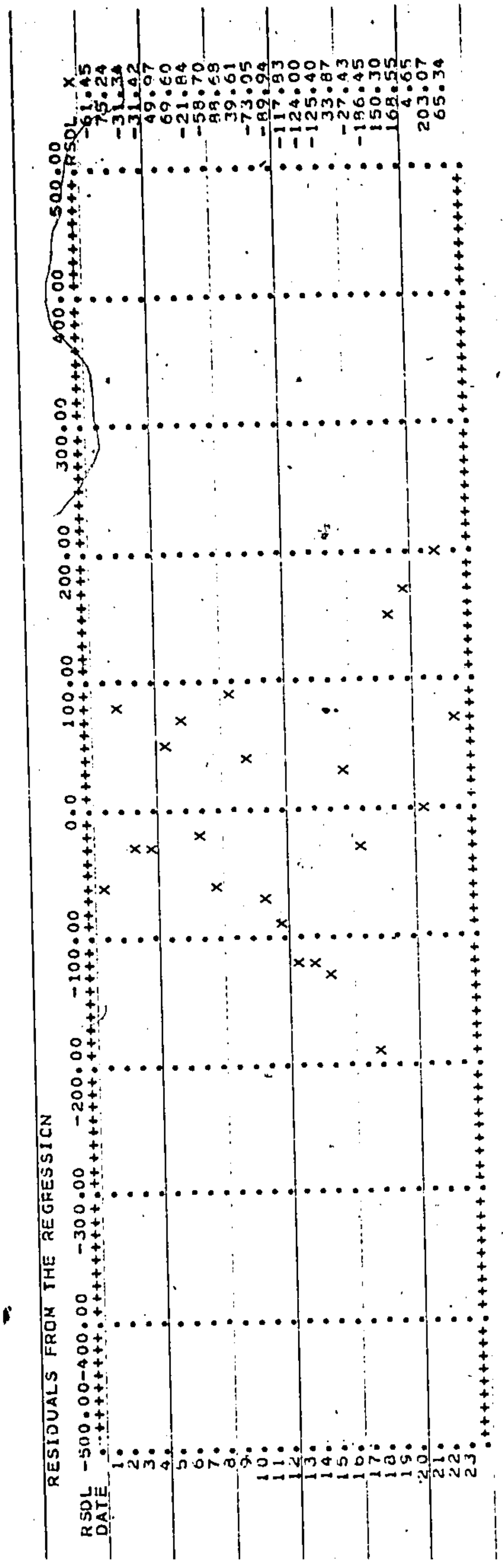


Figure C6: Residuals (BCF) of Model 6.

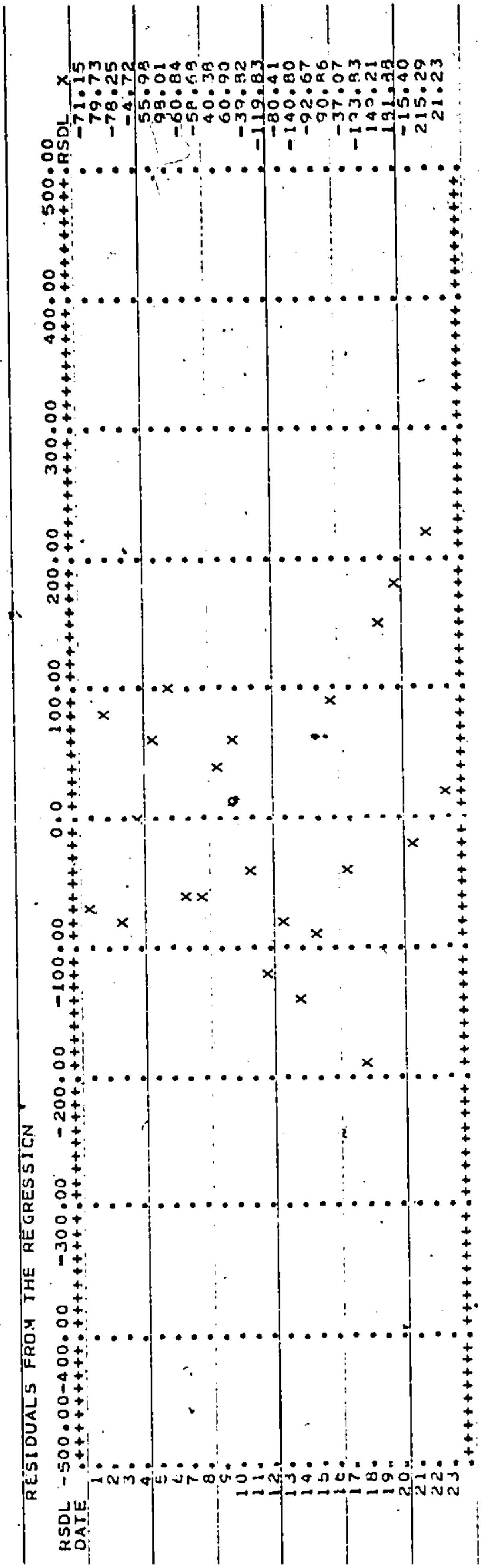


Figure C7: Residuals (BCF) of Model 7.

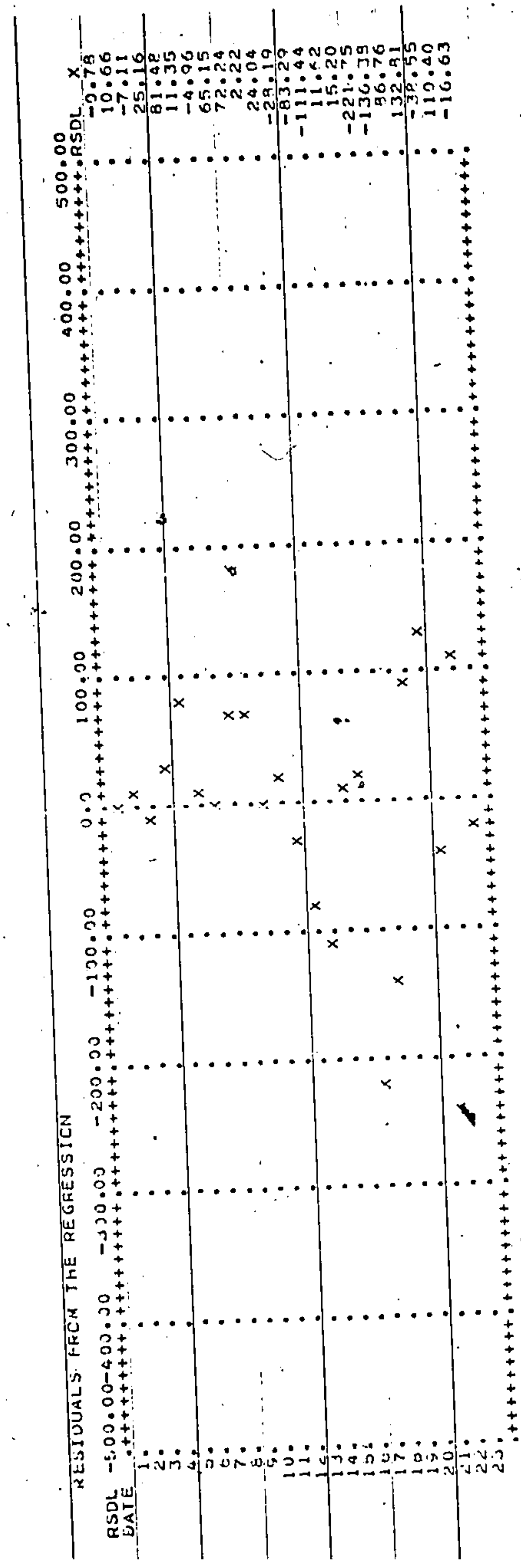


Figure C8: Residuals (BCF) of Model 8.

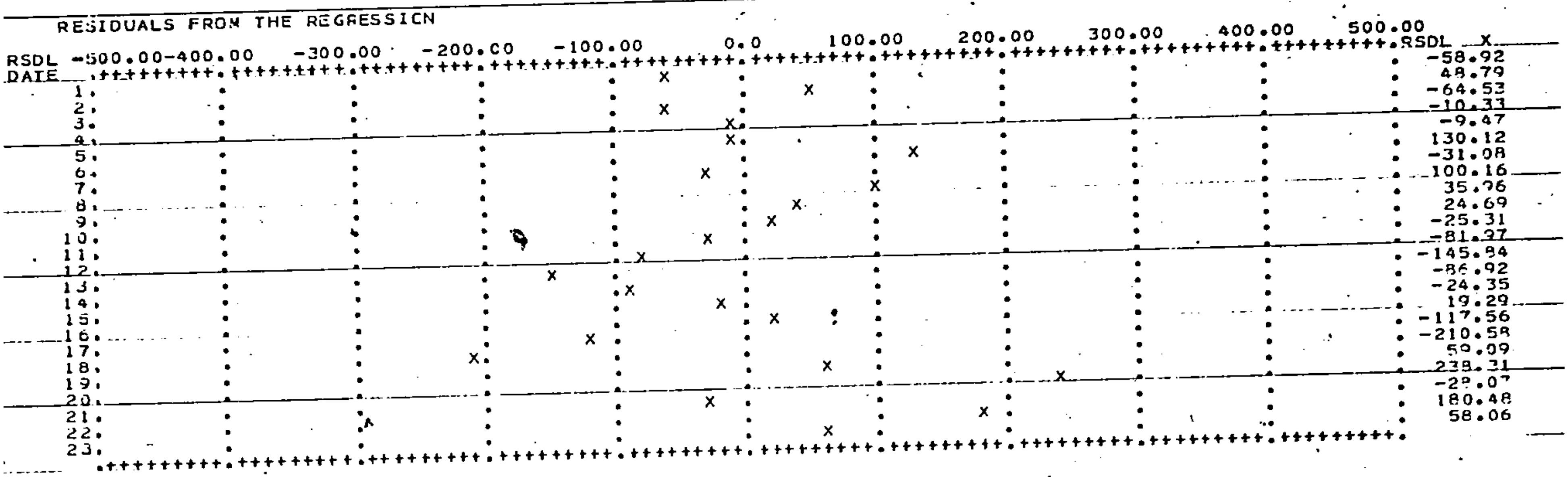


Figure C9: Residuals (BCF) of Model 9.

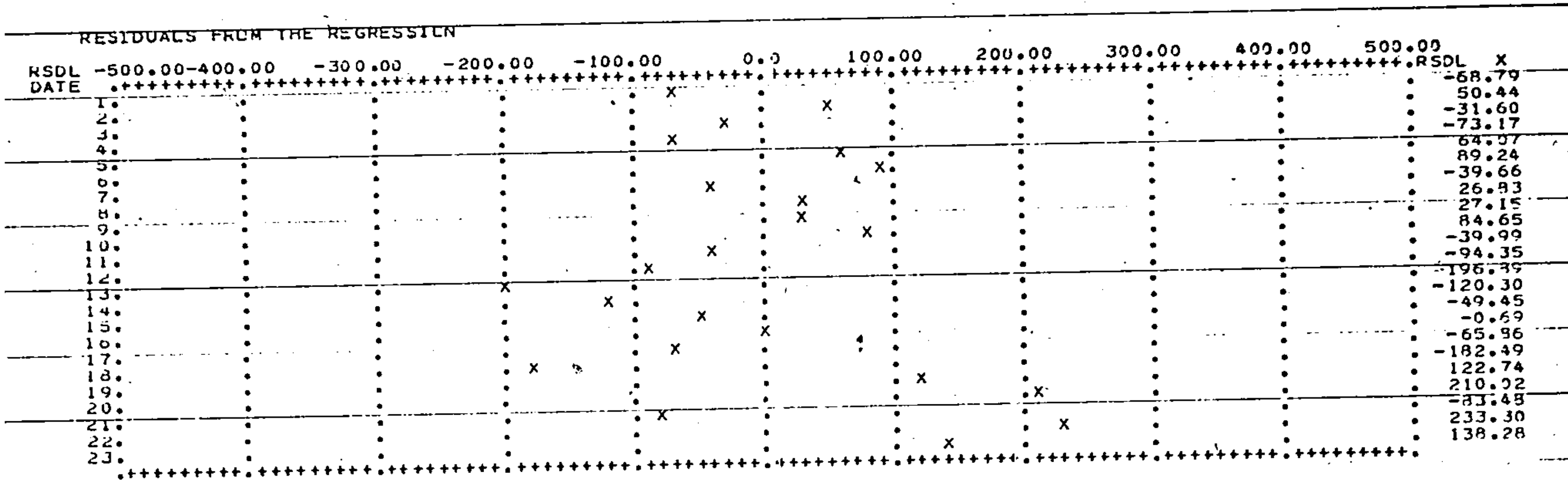


Figure C10: Residuals (BCF) of Model 10.

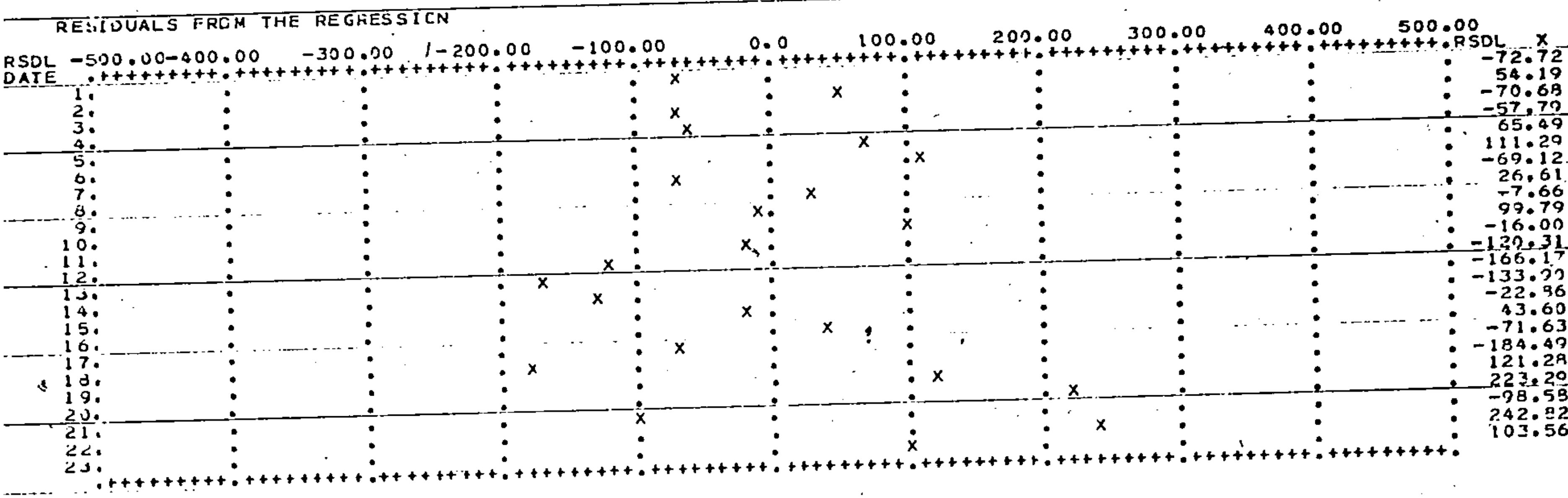


Figure C11: Residuals (BCF) of Model 11.

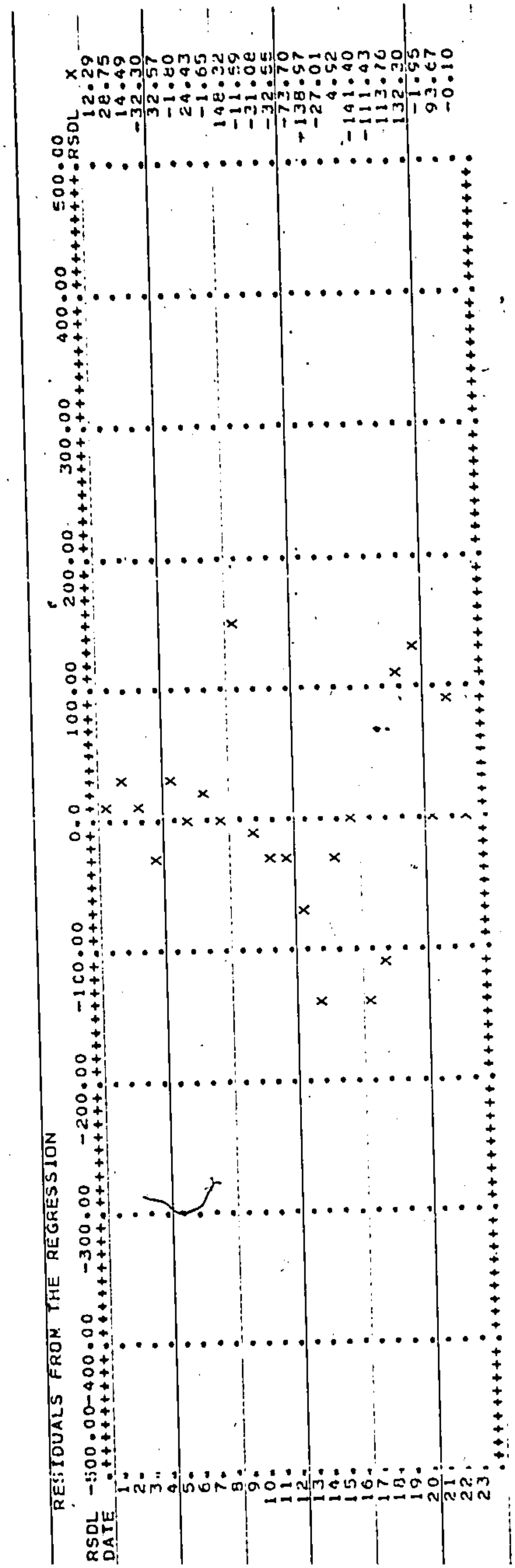


Figure C12: Residuals (BCF) of Model 12.

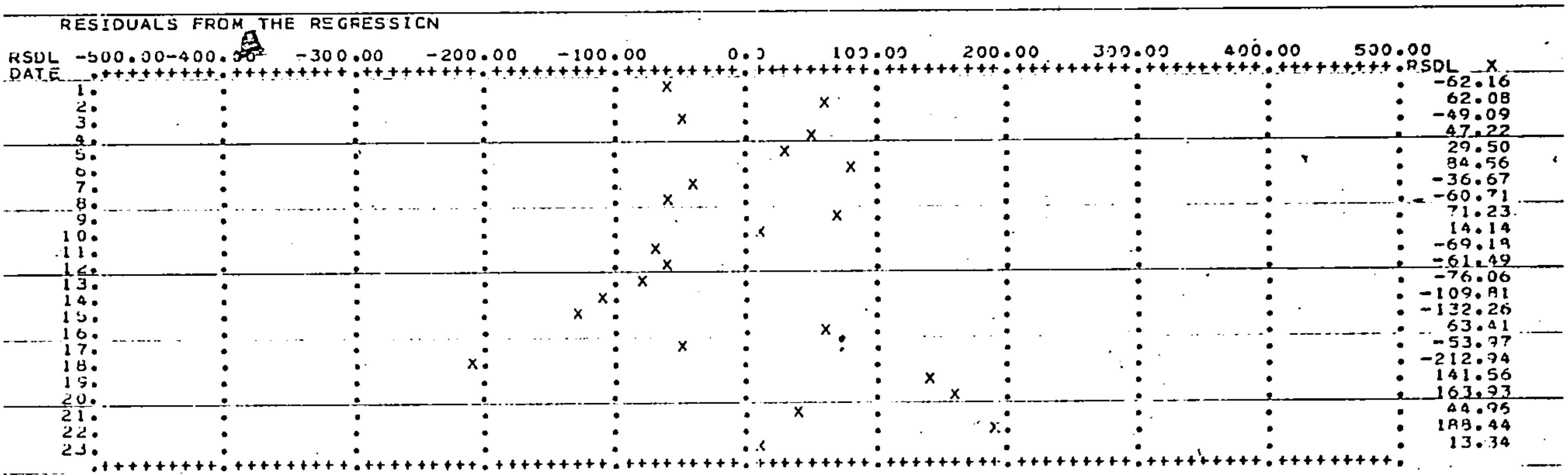


Figure C13: Residuals (BCF) of Model 13.

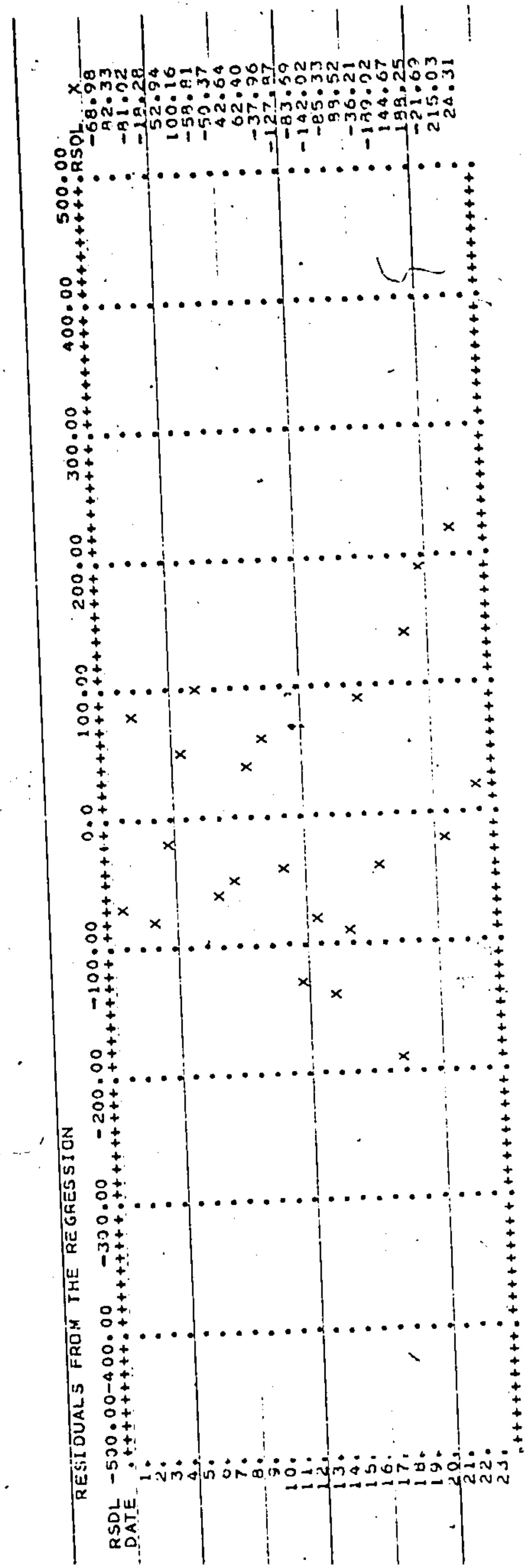


Figure C14: Residuals (BCF) of Model 14.

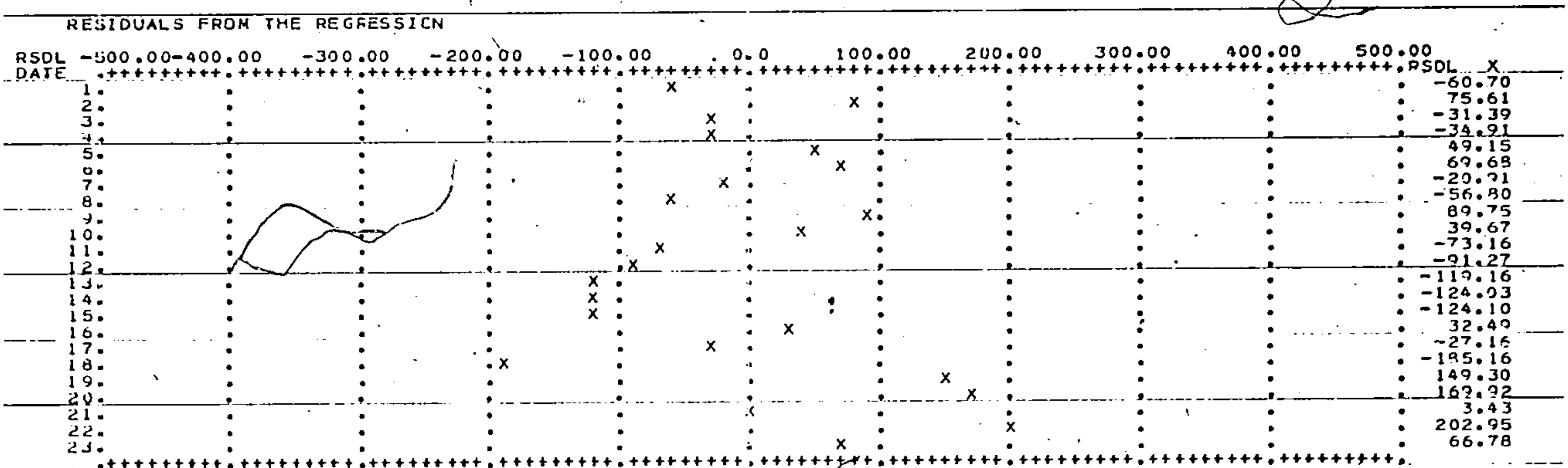


Figure C15: Residuals (BCF) of Model 15.

Appendix D

SCATTERGRAMS OF MODELS 1 THROUGH 15

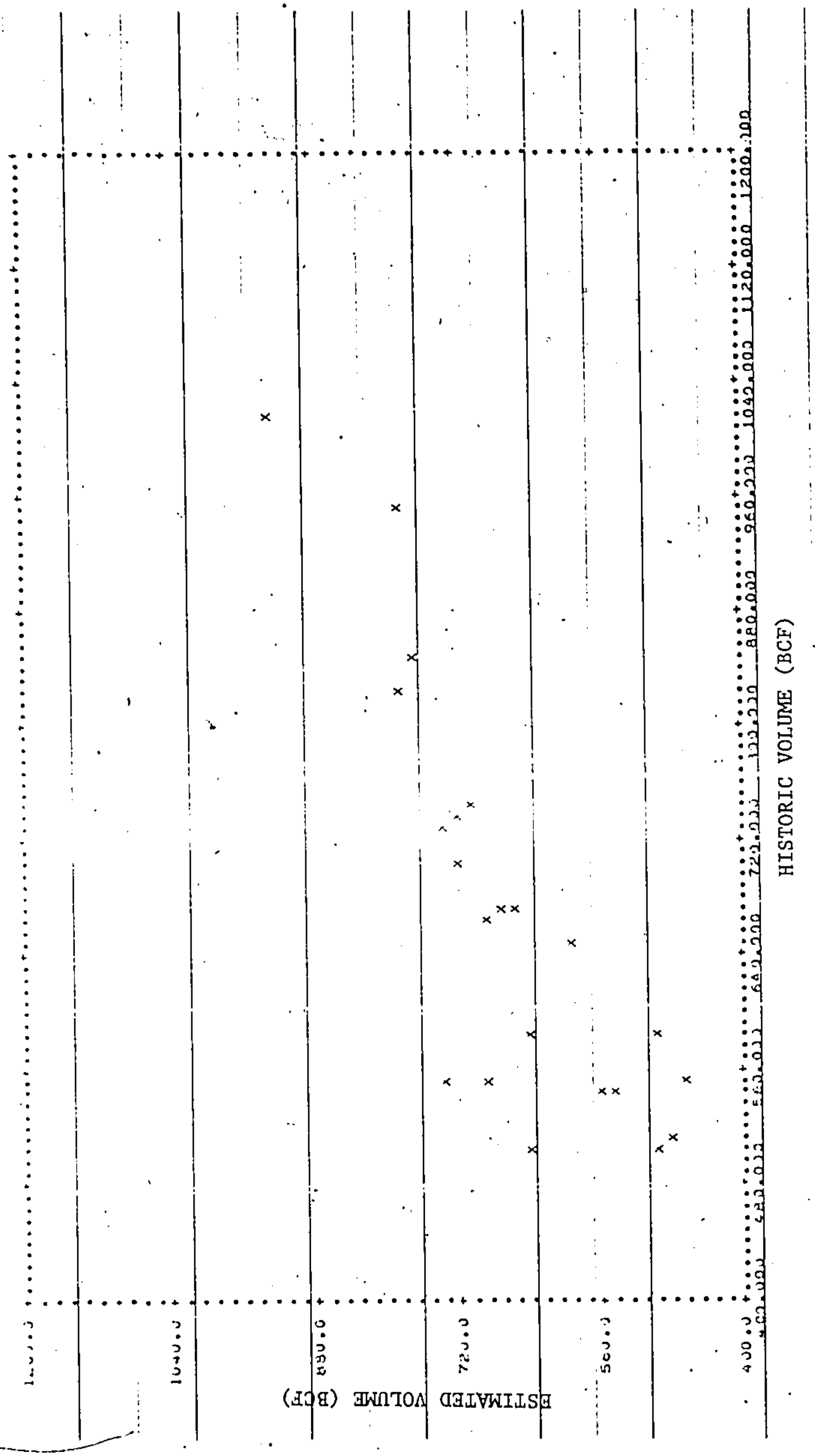


Figure D1: Scattergram for Spring Flood Volume (BCF)- Model 1.

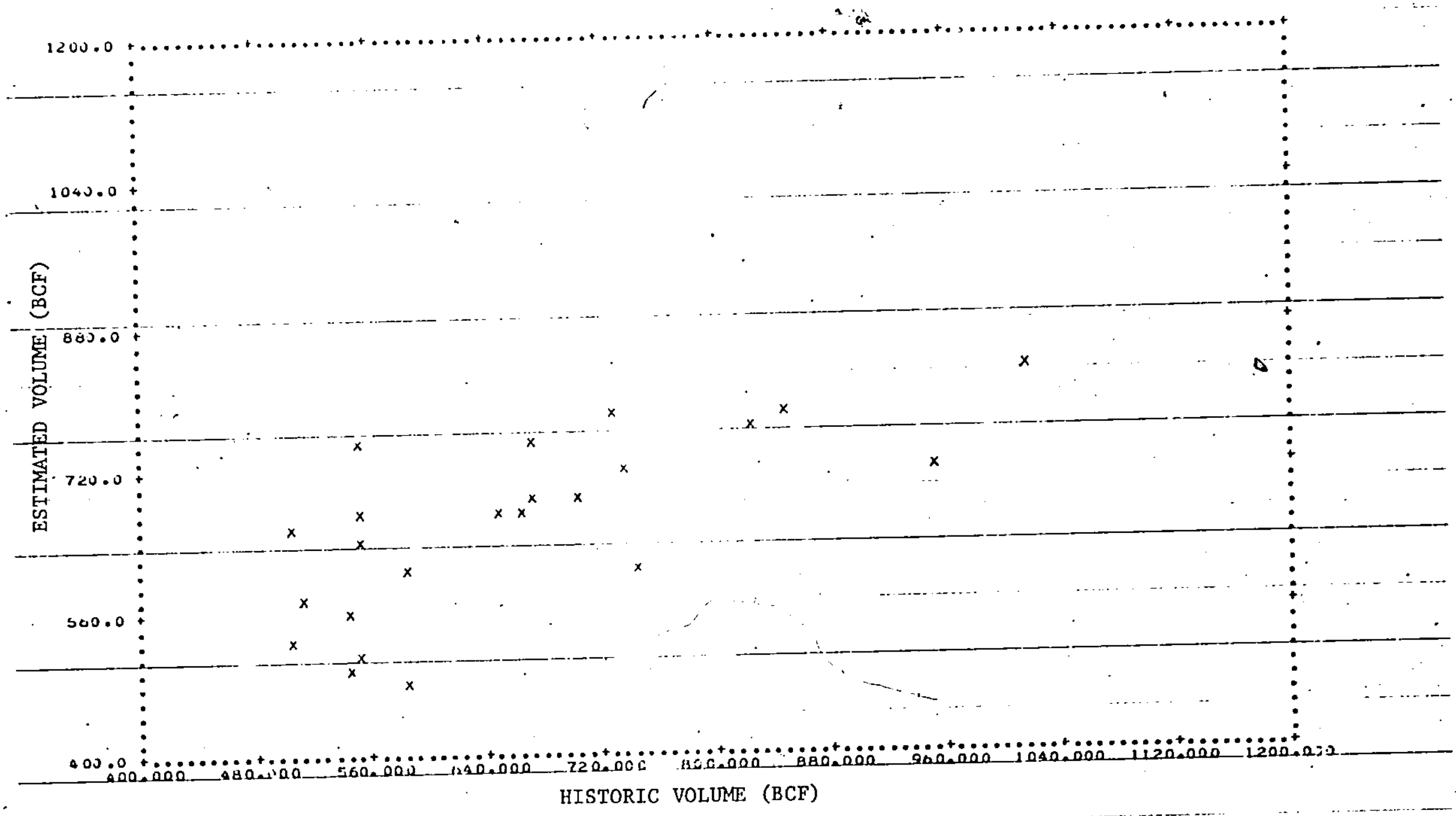


Figure D2: Scattergram for Spring Water Yield (BCF)- Model 2.

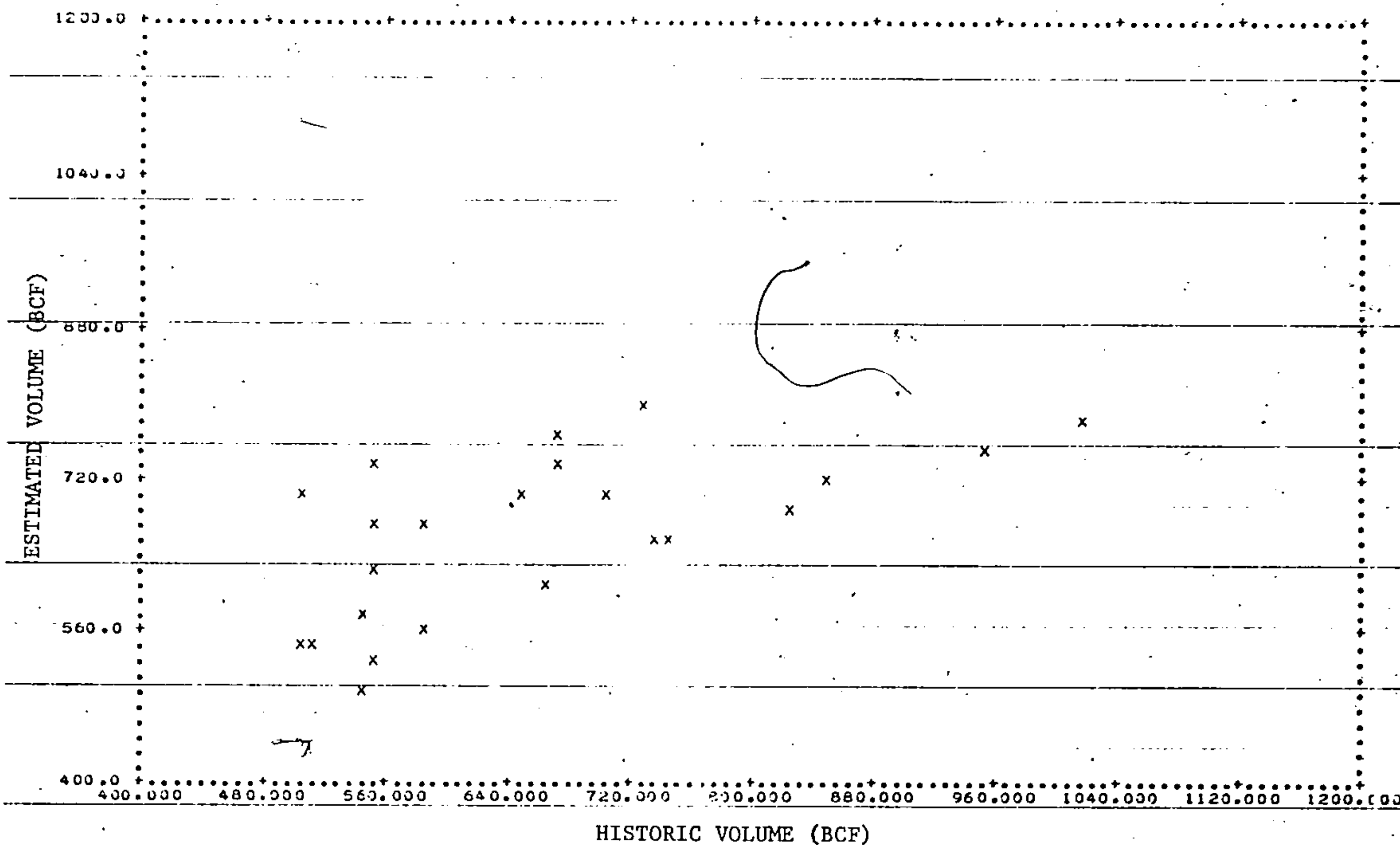


Figure D3: Scattergram for Spring Water Yield (BCF)- Model 3.

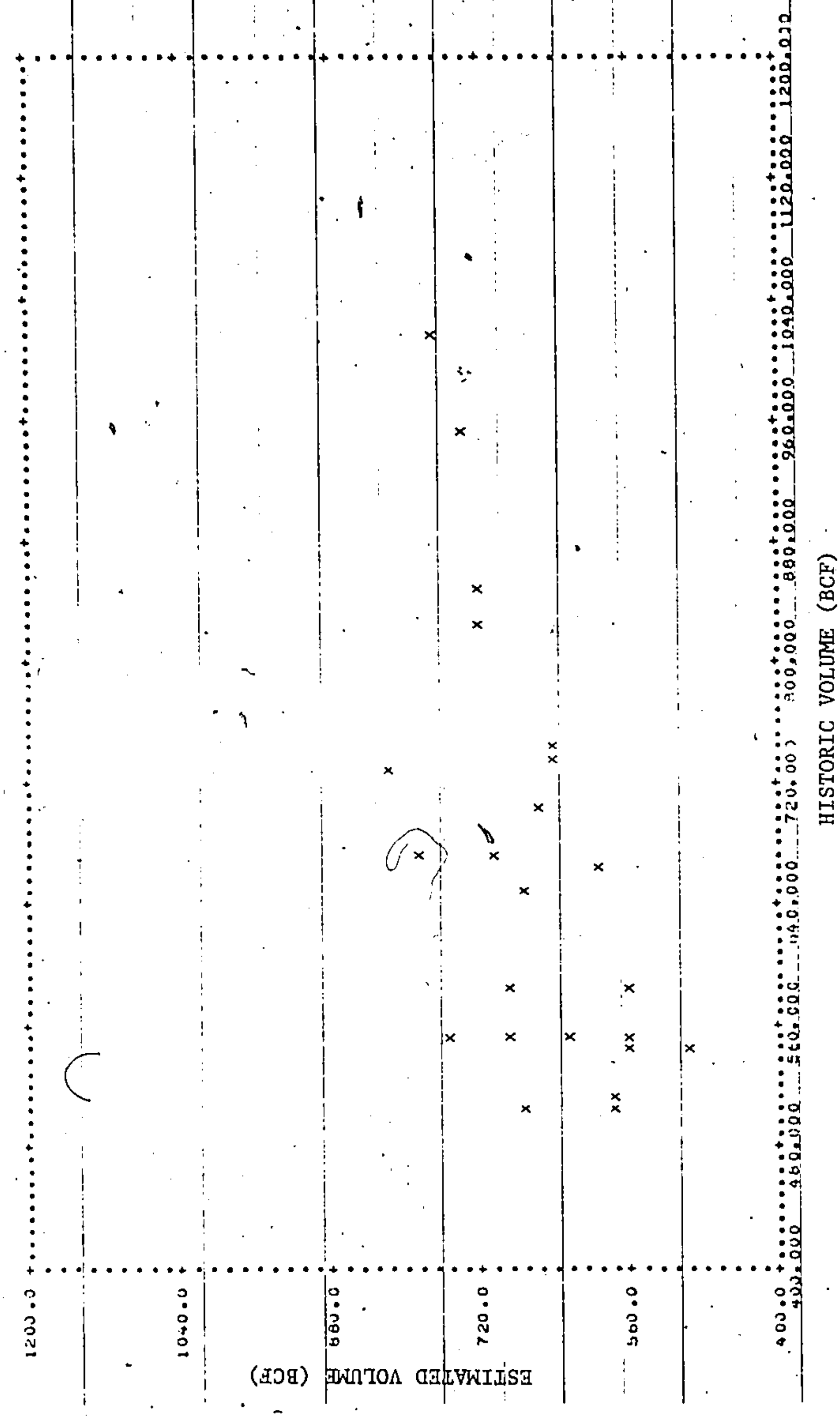


Figure D4: Scattergram for Spring Water Yield (BCF)- Model 4.

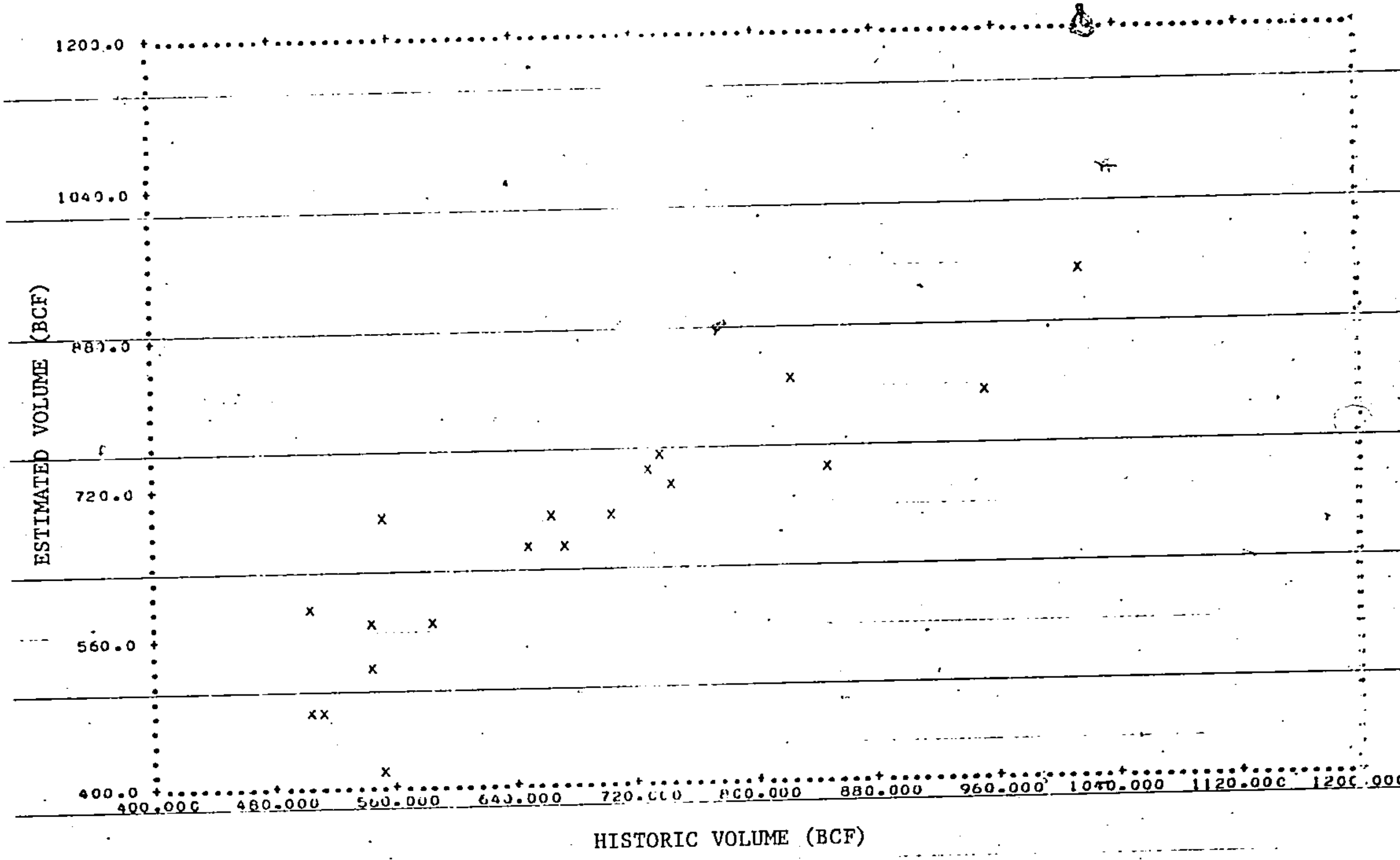


Figure D5: Scattergram for Spring Water Yield (BCF)- Model 5.

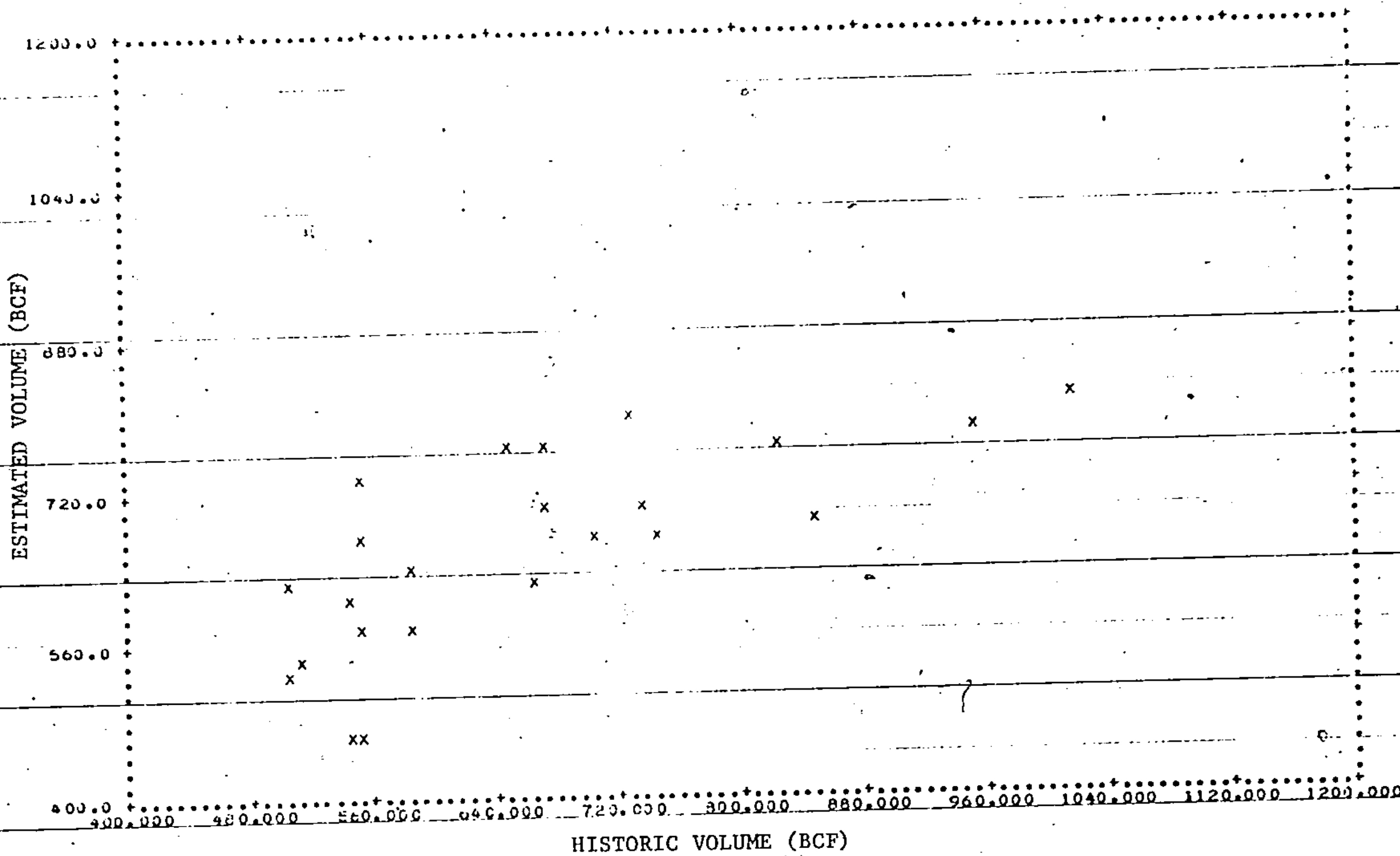


Figure D6: Scattergram for Spring Water Yield (BCF)- Model 6.

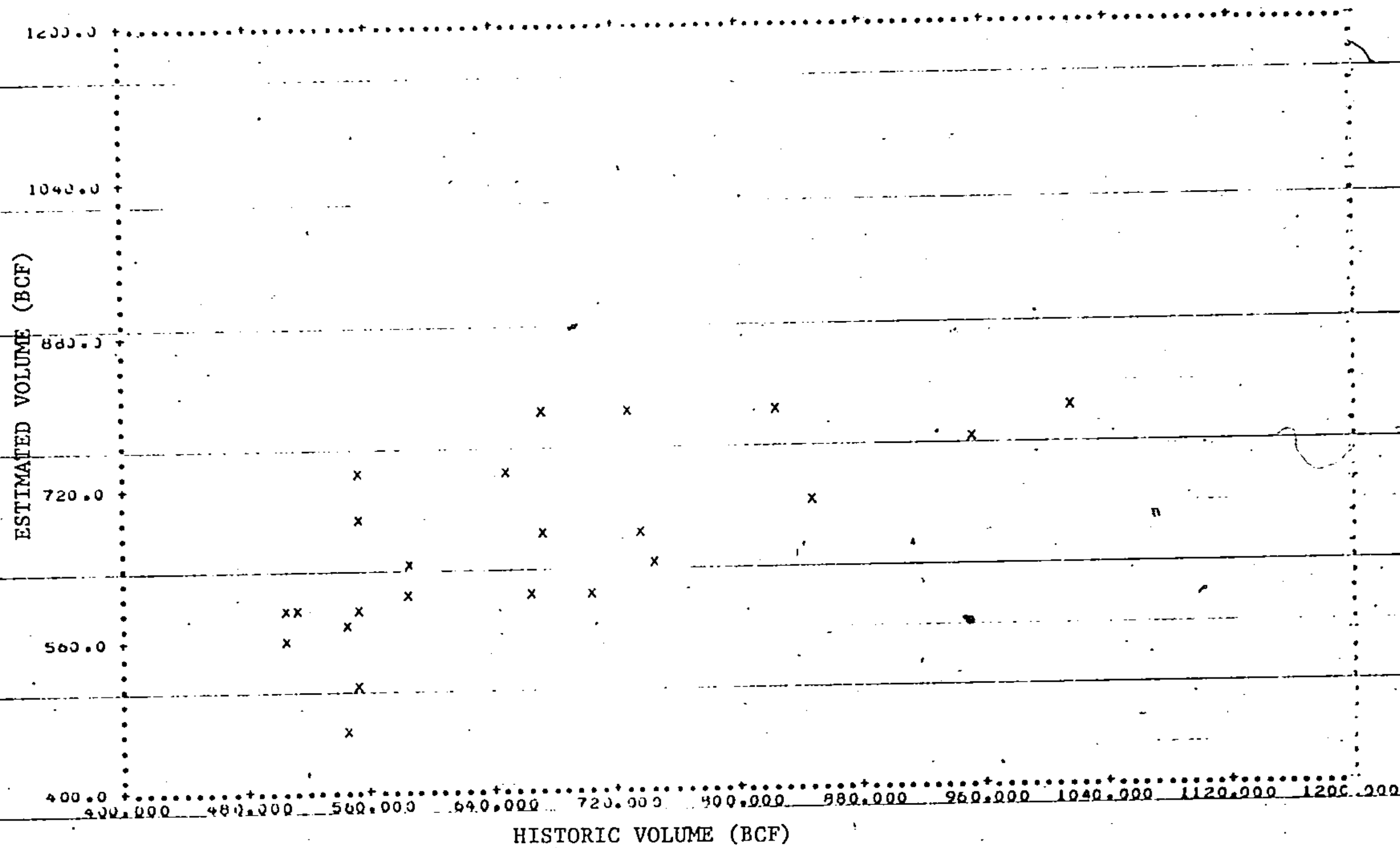


Figure D7: Scattergram for Spring Water Yield (BCF)- Model 7.

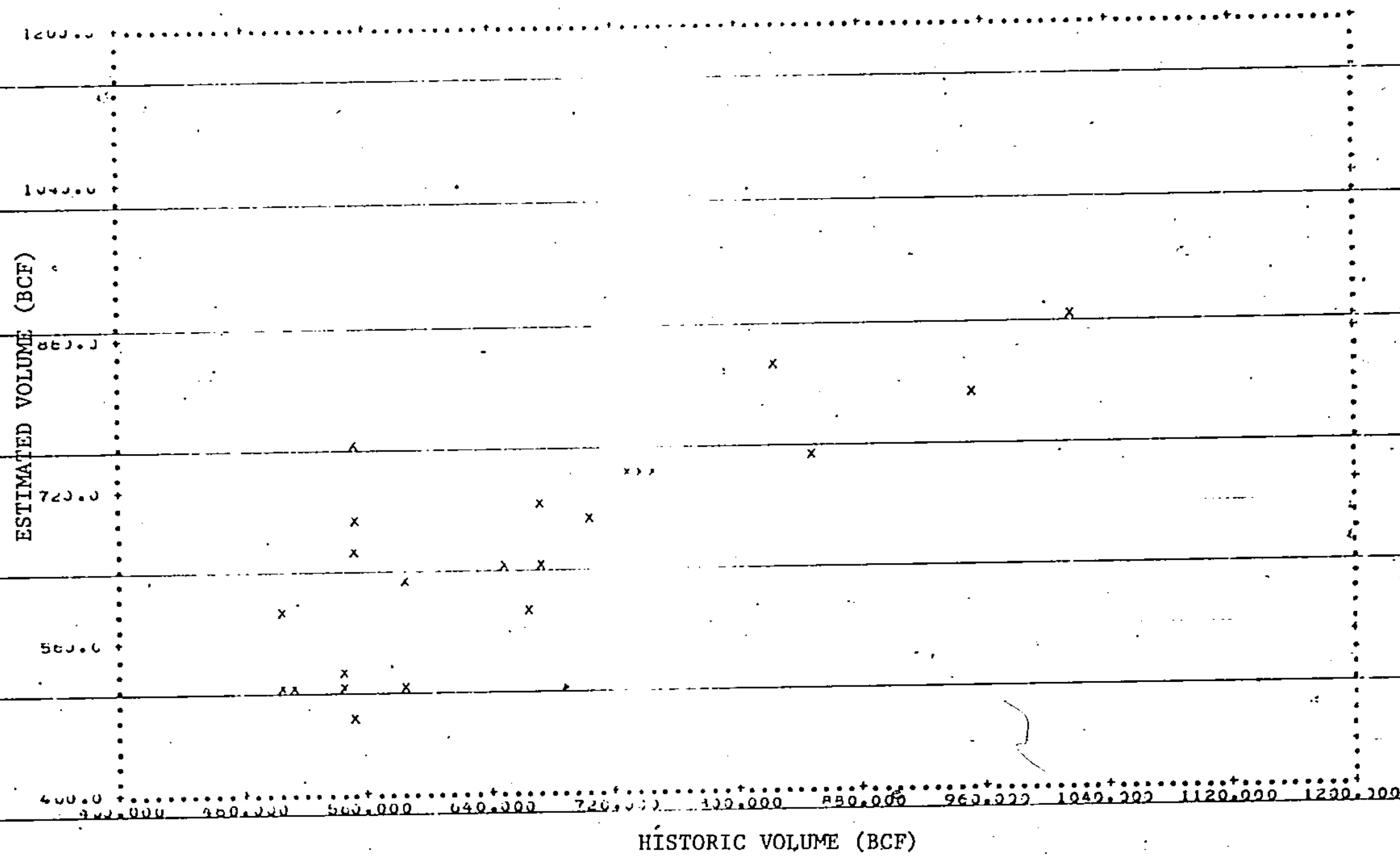


Figure D8: Scattergram for Spring Water Yield (BCF)- Model 8.

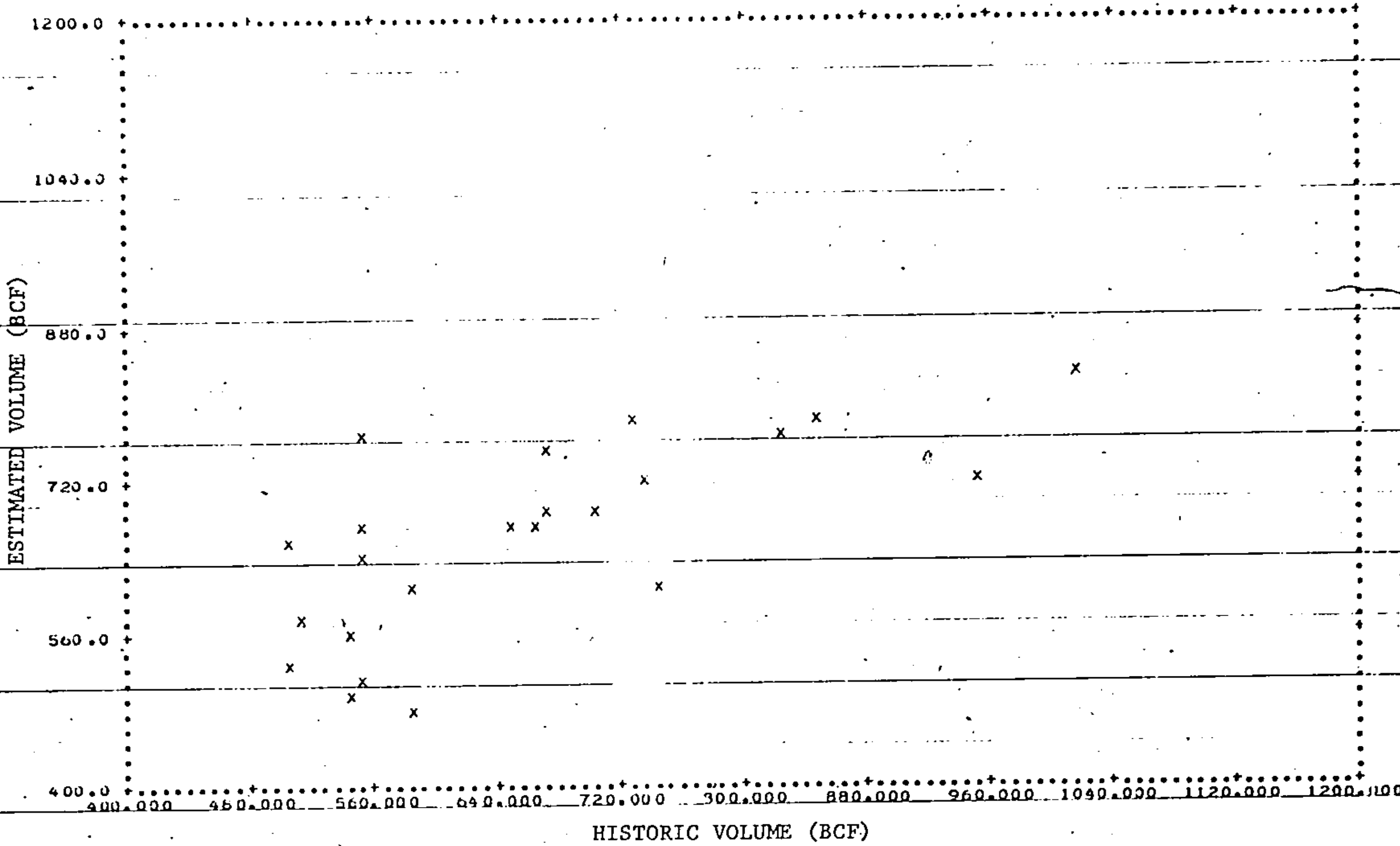


Figure D9: Scattergram for Spring Water Yield (BCF)- Model 9.

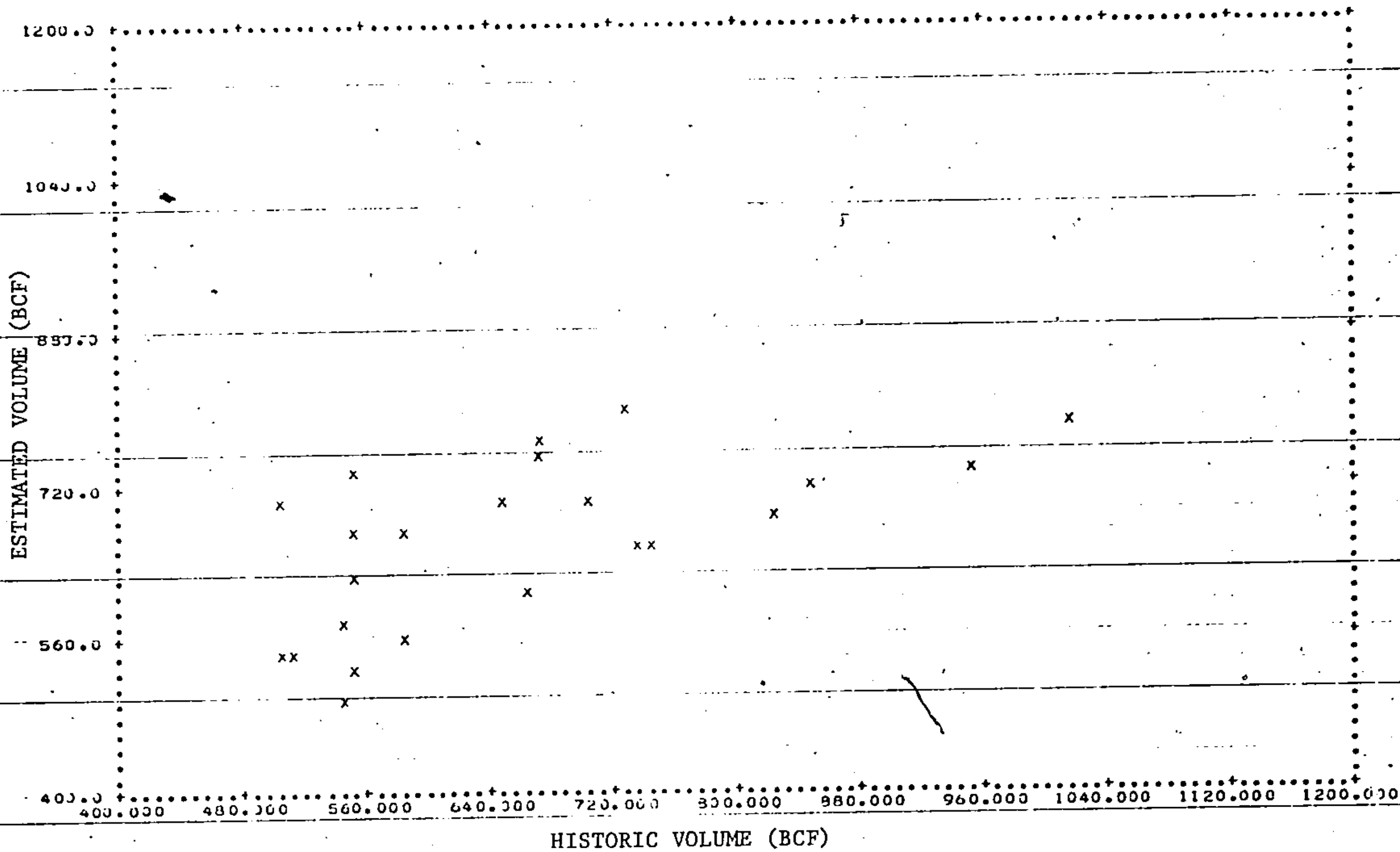


Figure D10: Scattergram for Spring Water Yield (BCF)- Model 10.

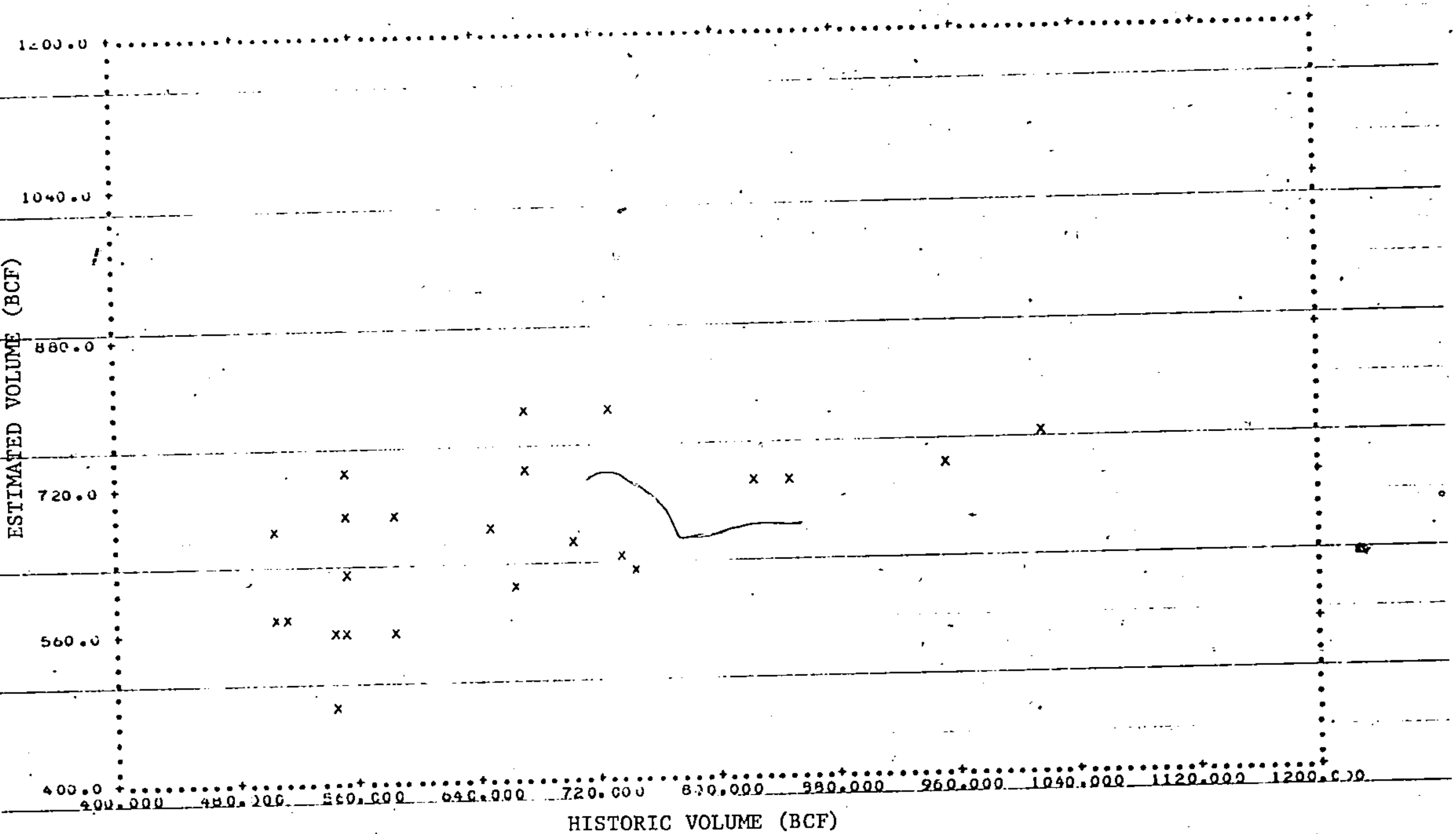


Figure D11: Scattergram for Spring Water Yield (BCF)- Model 11.

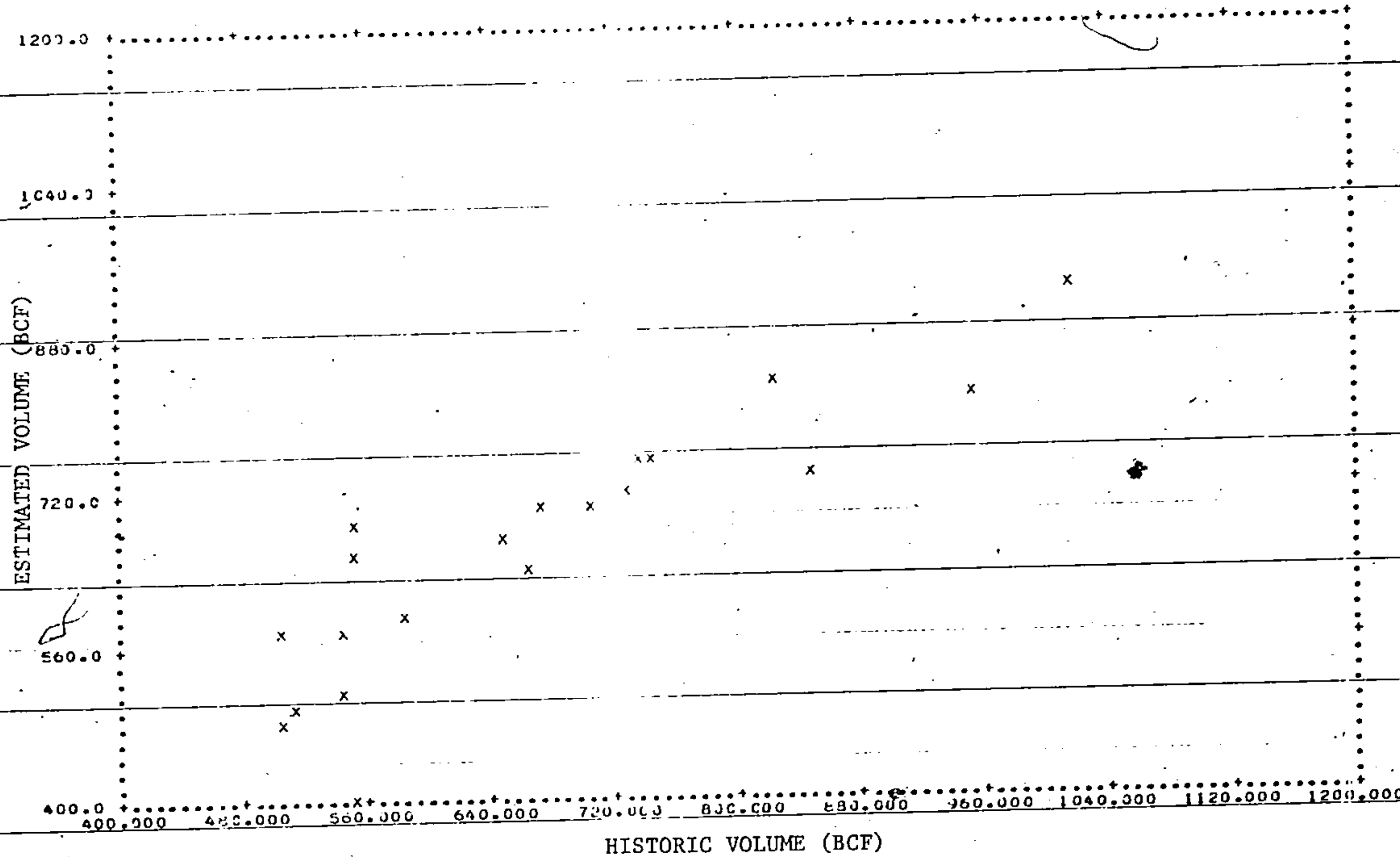


Figure D12: Scattergram for Spring Water Yield (BCF)- Model 12.

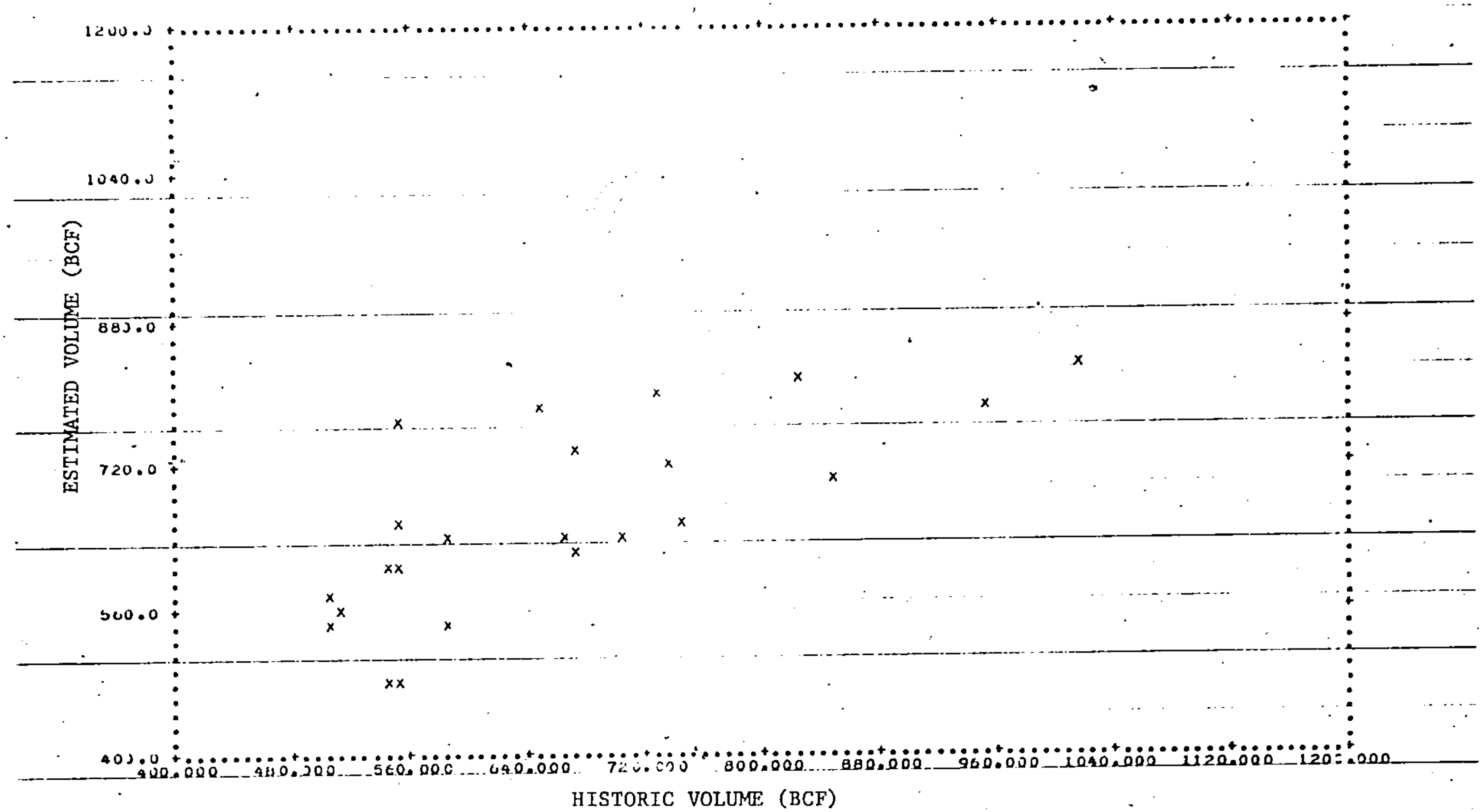


Figure D13: Scattergram for Spring Water Yield (BCF)- Model 13.

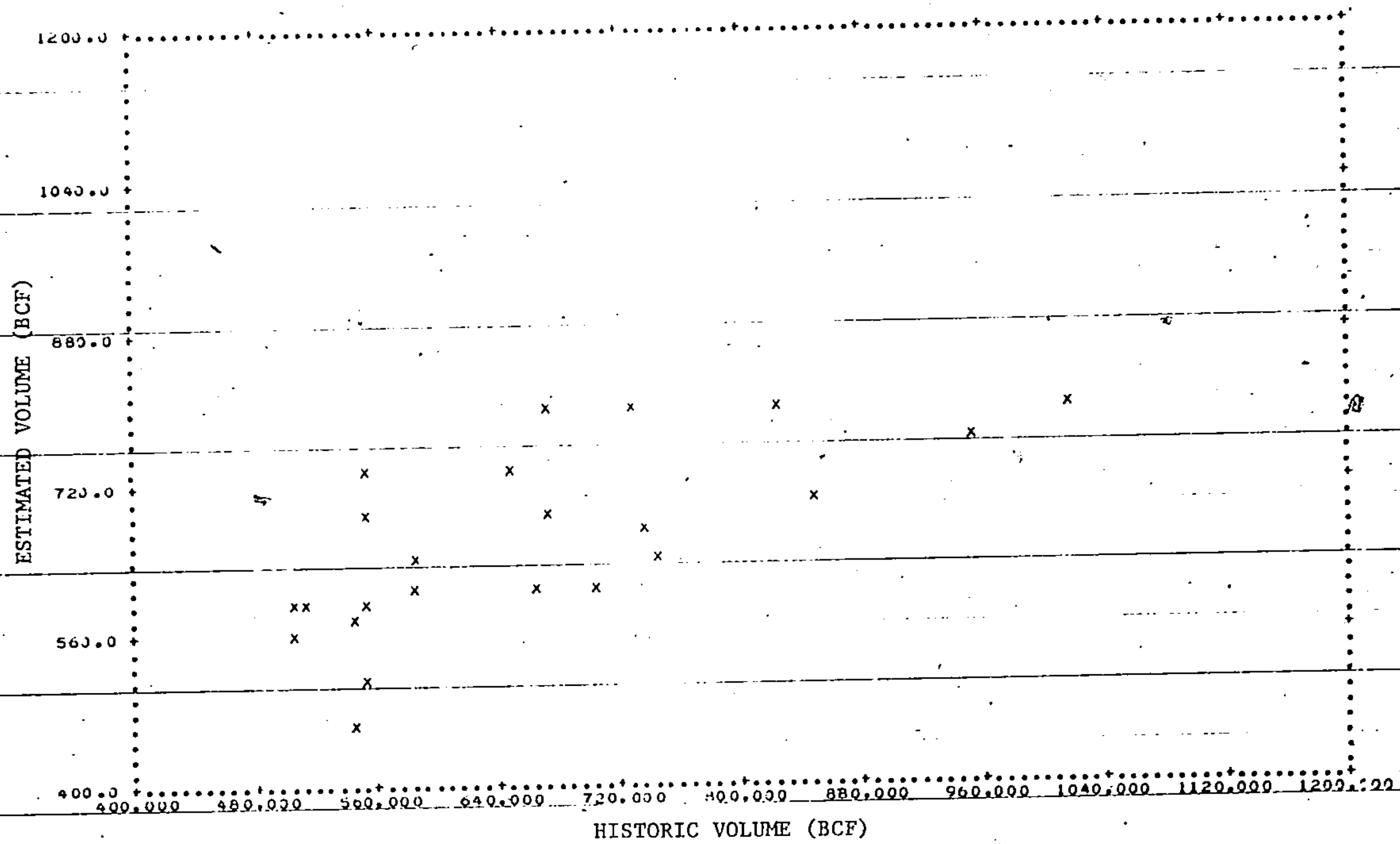


Figure D14: Scattergram for Spring Water Yield (BCF)- Model 14.

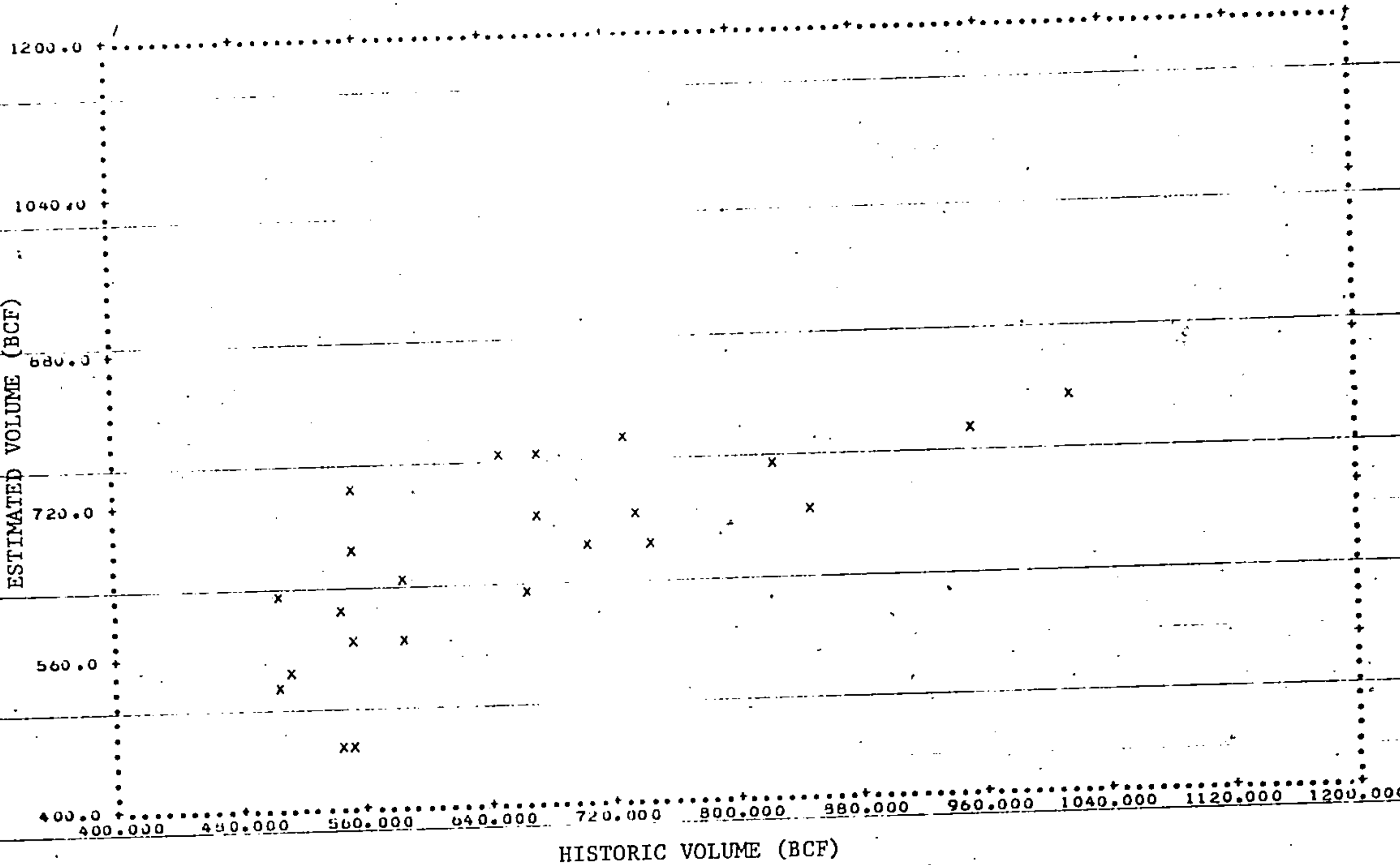


Figure D15: Scattergram for Spring Water Yield (BCF)- Model 15.

Appendix E

PLOT OF RESIDUALS OF MODELS 16 THROUGH 35

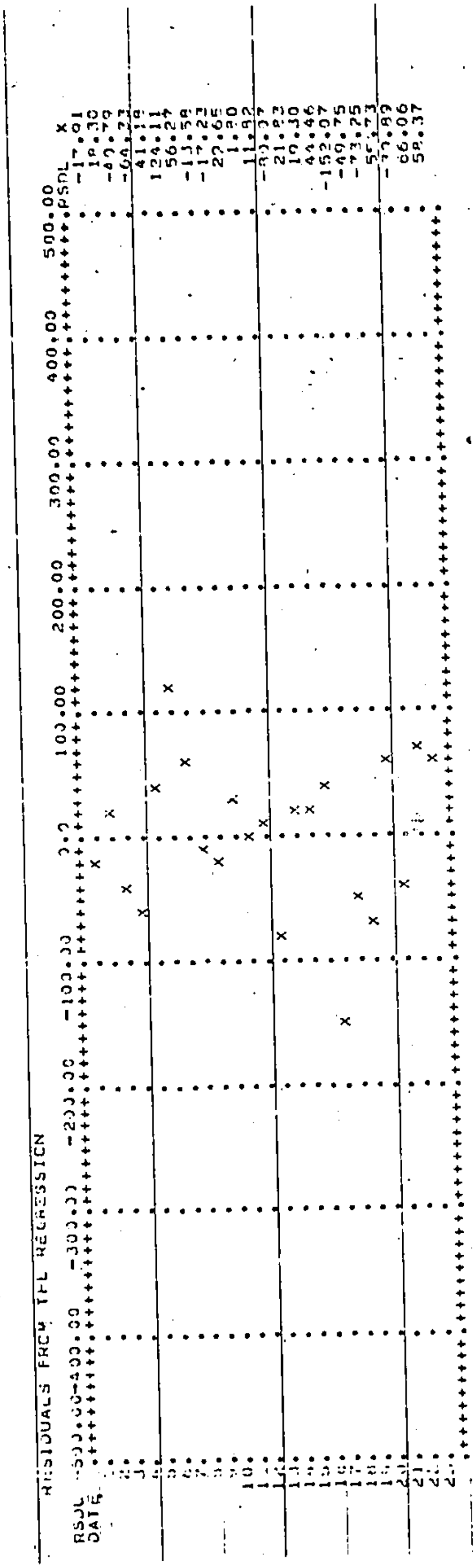


Figure E1: Residuals (BCF) of Model 16.

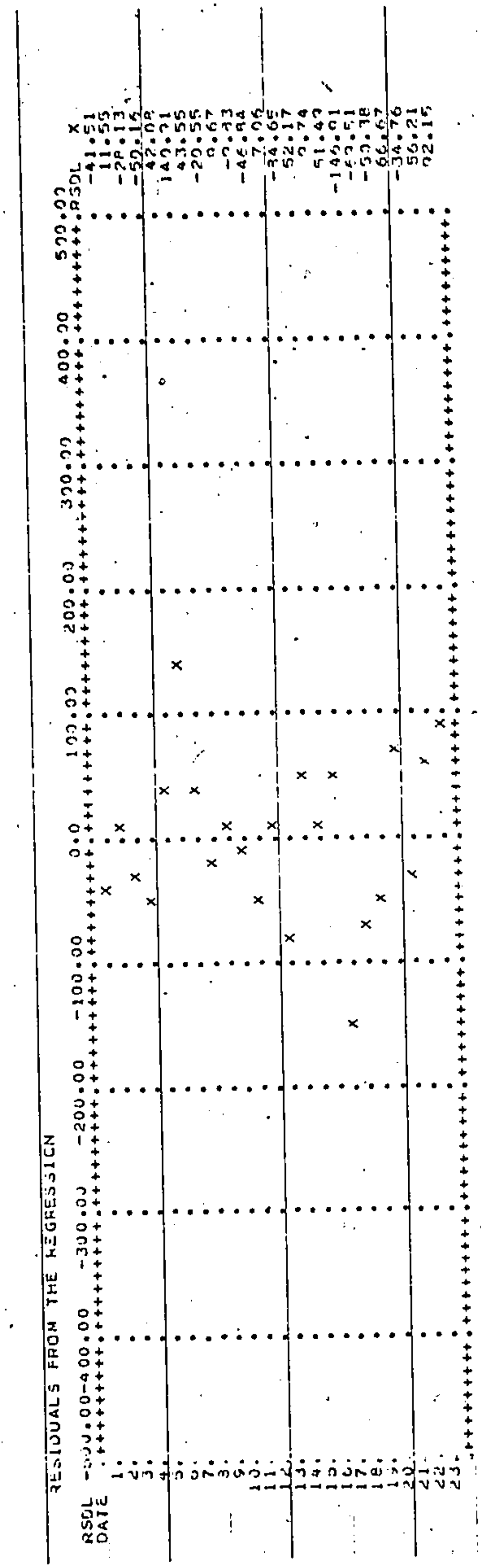


Figure E2: Residuals (BCF) of Model 17.

POOR COPY
COPIE DE QUALITEE INFERIEURE

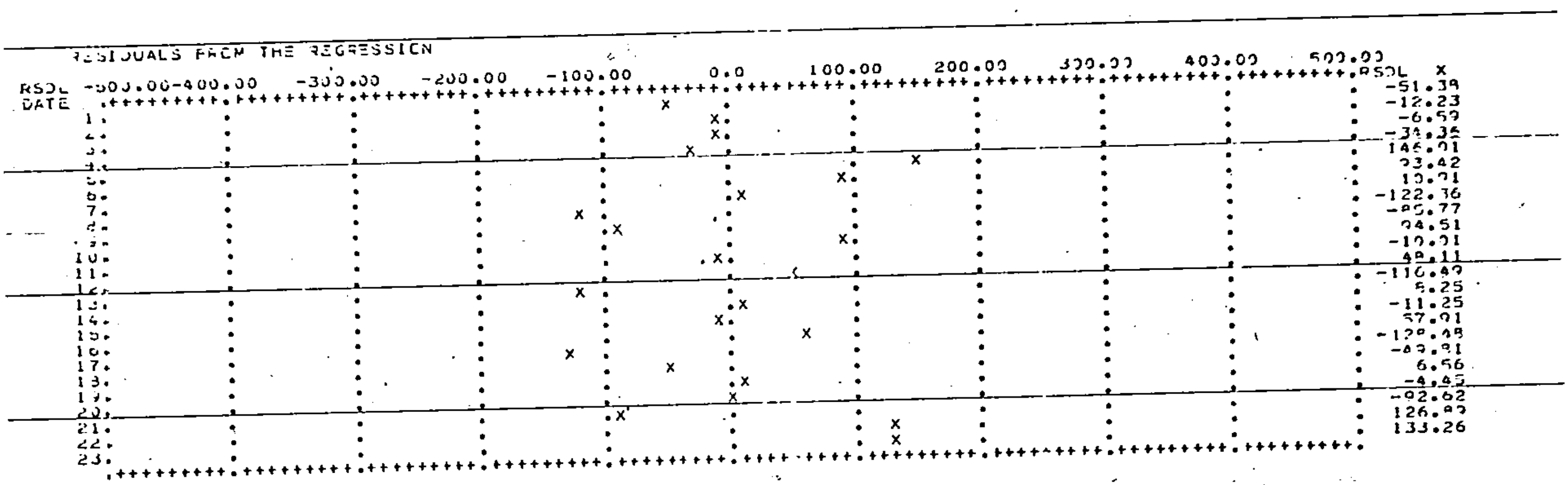


Figure E3: Residuals (BCF) of Model 18.

POOR COPY
COPIE DE QUALITEE INFERIEURE

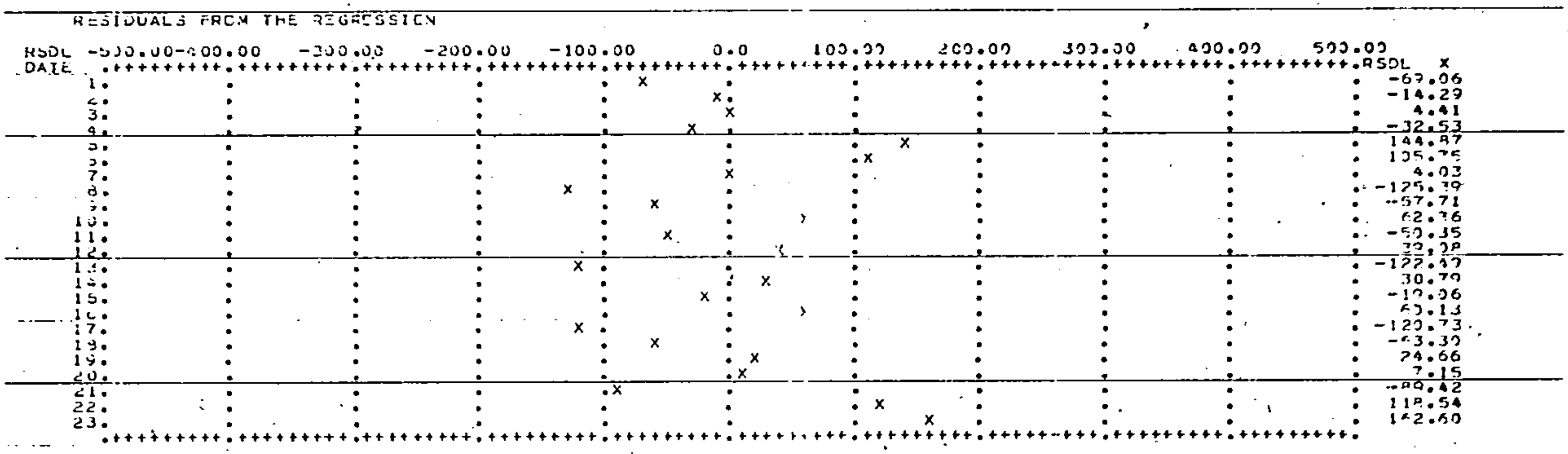


Figure E4: Residuals (BCF) of Model 19.

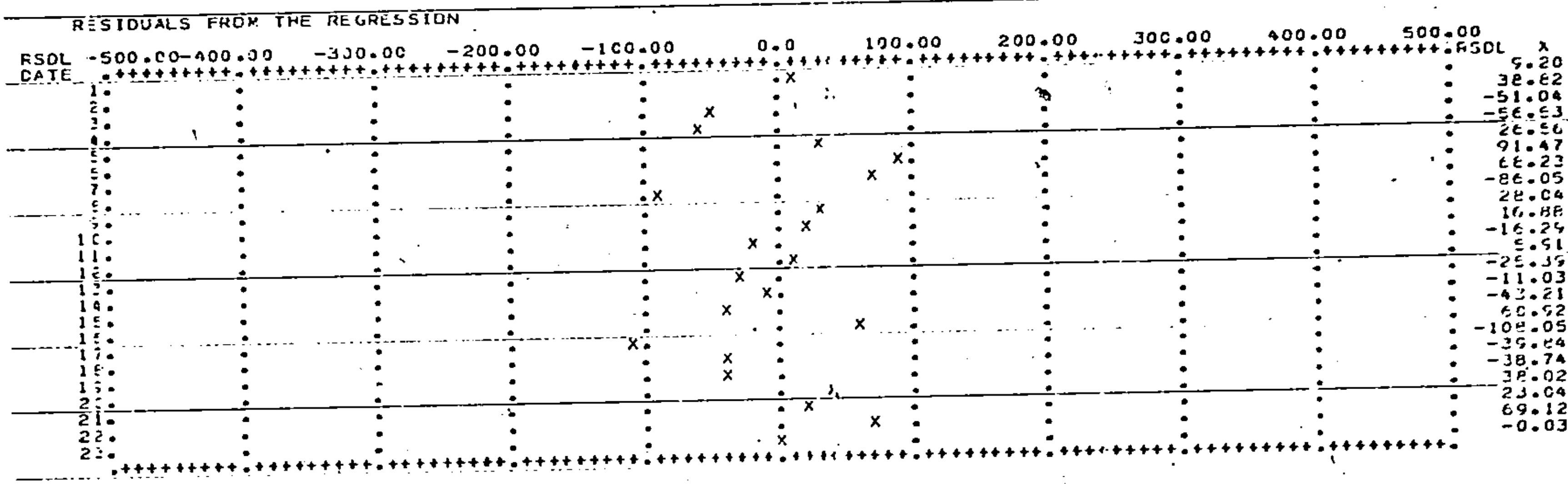


Figure E5: Residuals (BCF) of Model 20.

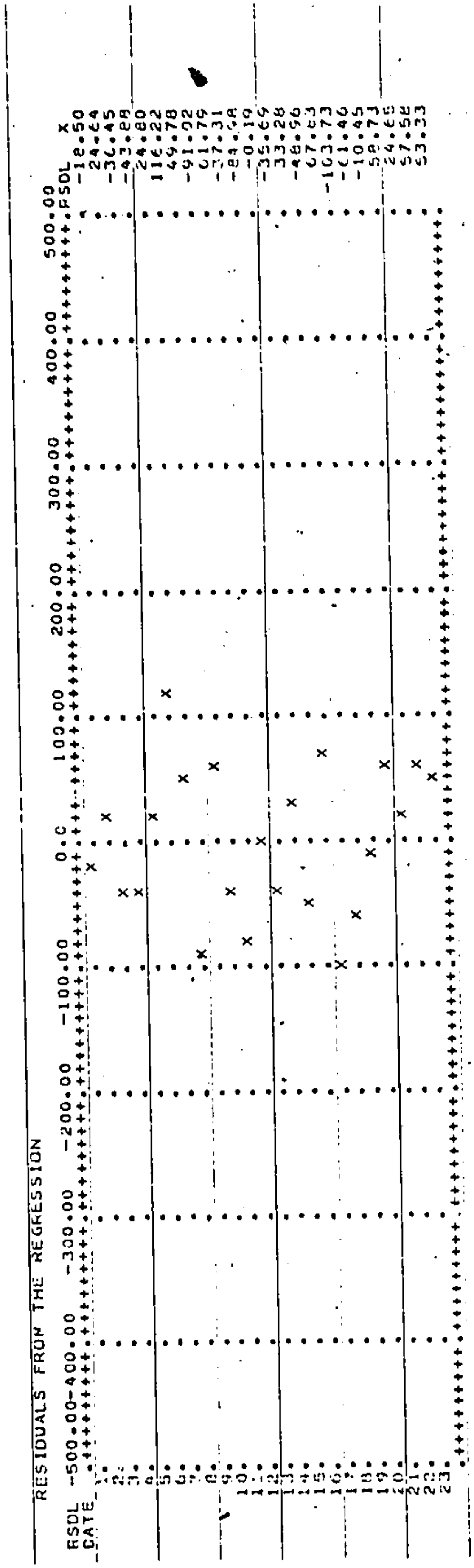


Figure E6: Residuals (BCF) of Model 21.

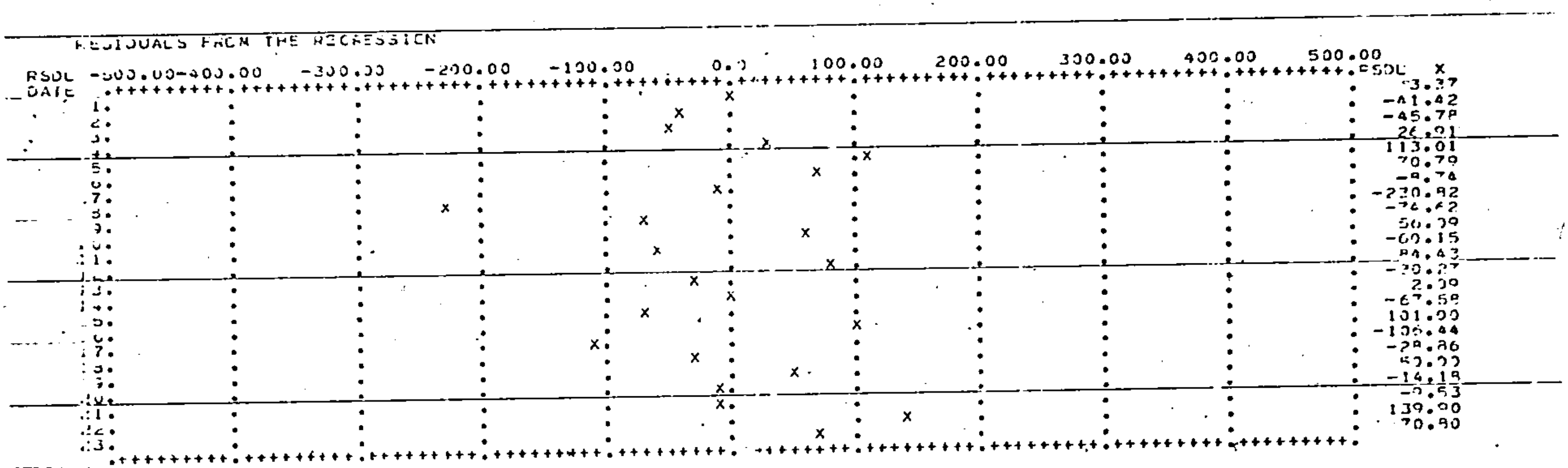


Figure E7: Residuals (BCF) of Model 22.

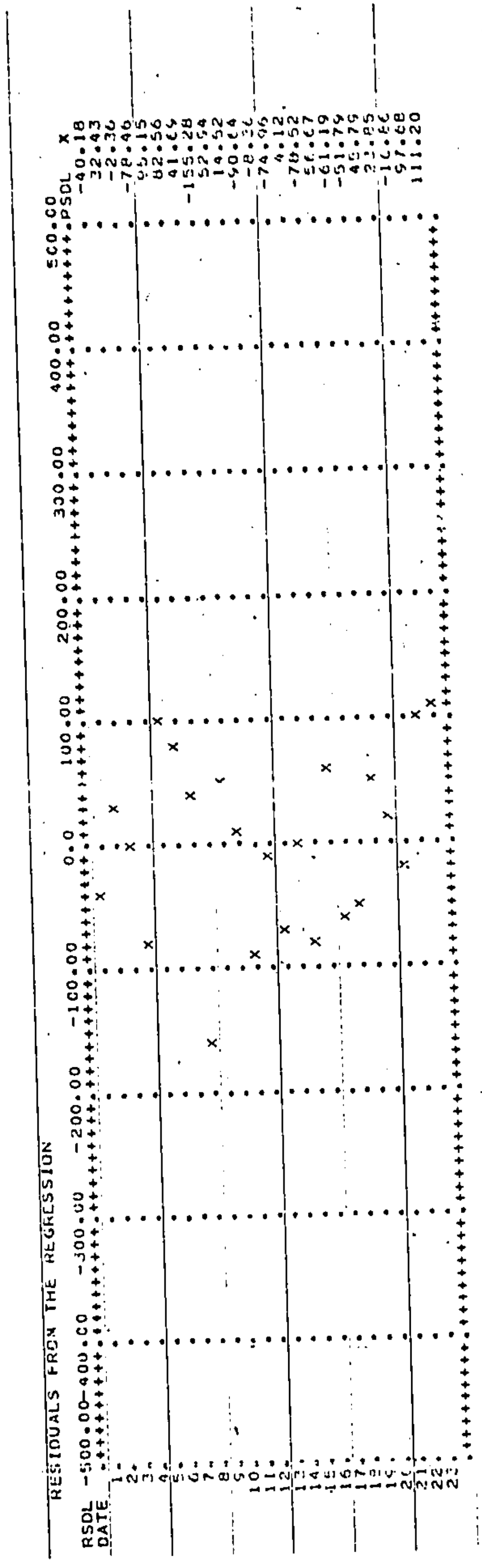


Figure E8: Residuals (BCF) of Model 23.

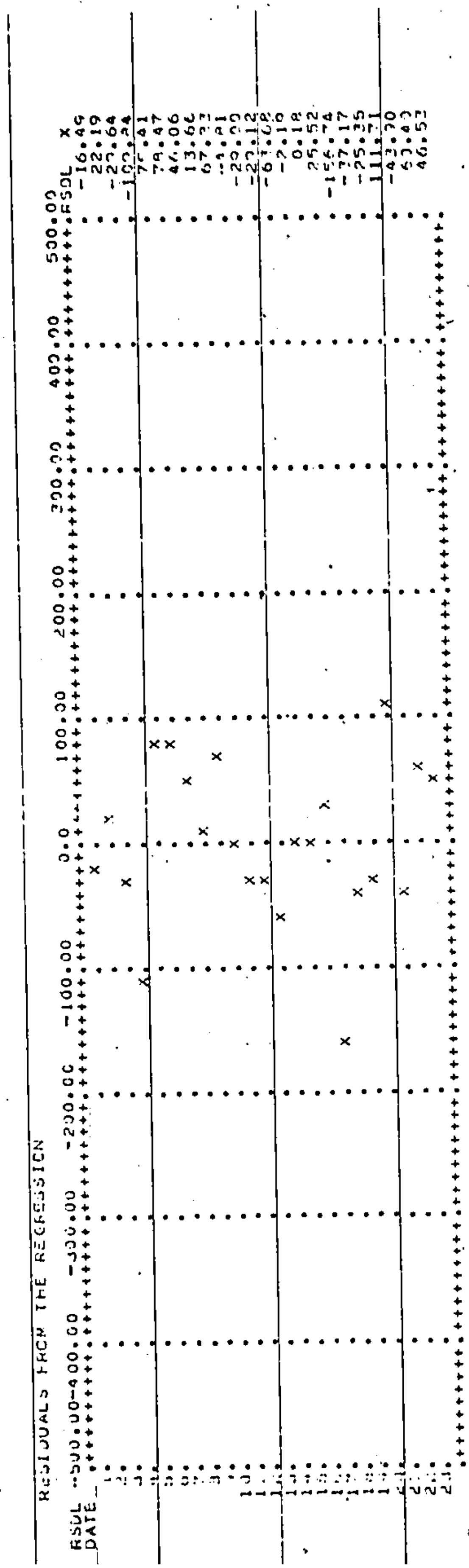


Figure E9: Residuals (BCF) of Model 24.

POOR COPY
COPIE DE QUALITEE INFERIEURE

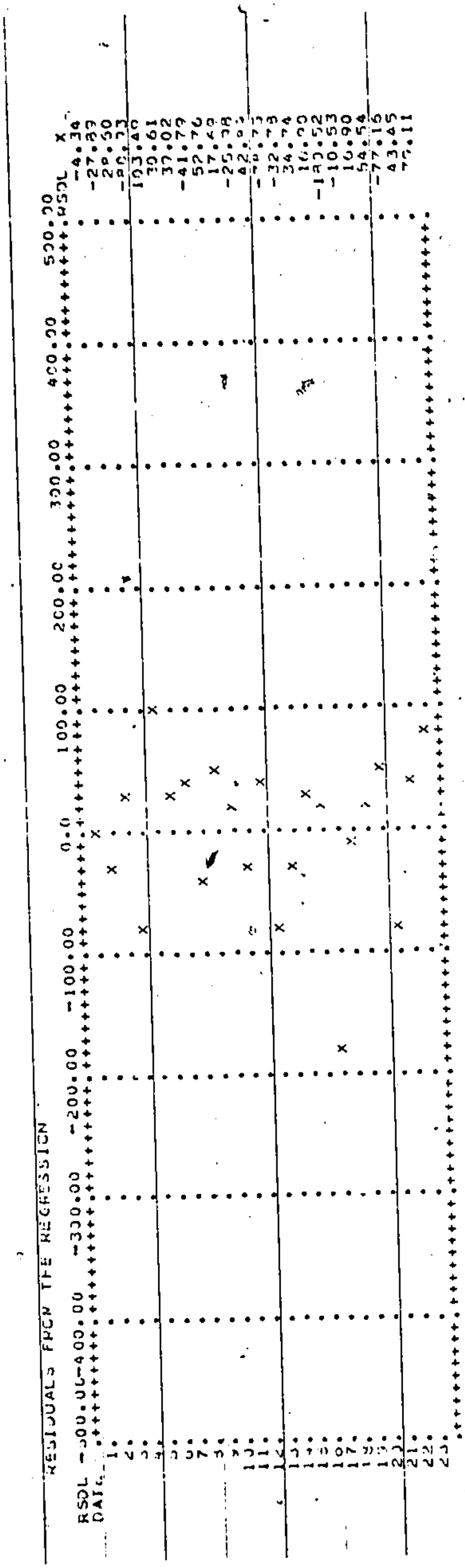


Figure E10: Residuals (BCF) of Model 25.

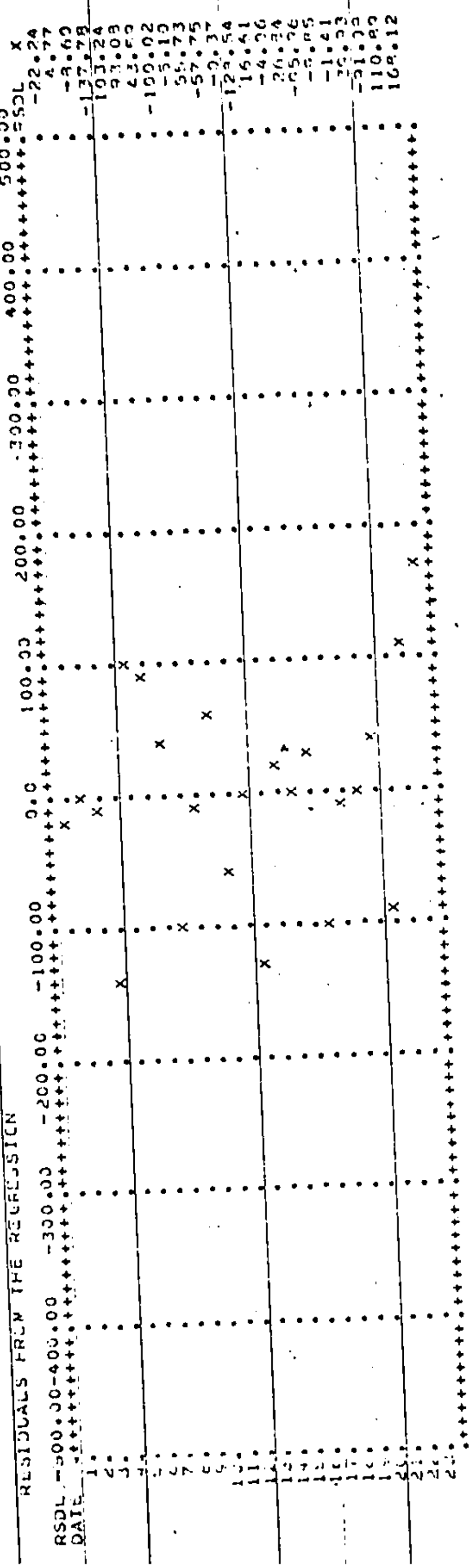


Figure E11: Residuals (BCF) of Model 26.

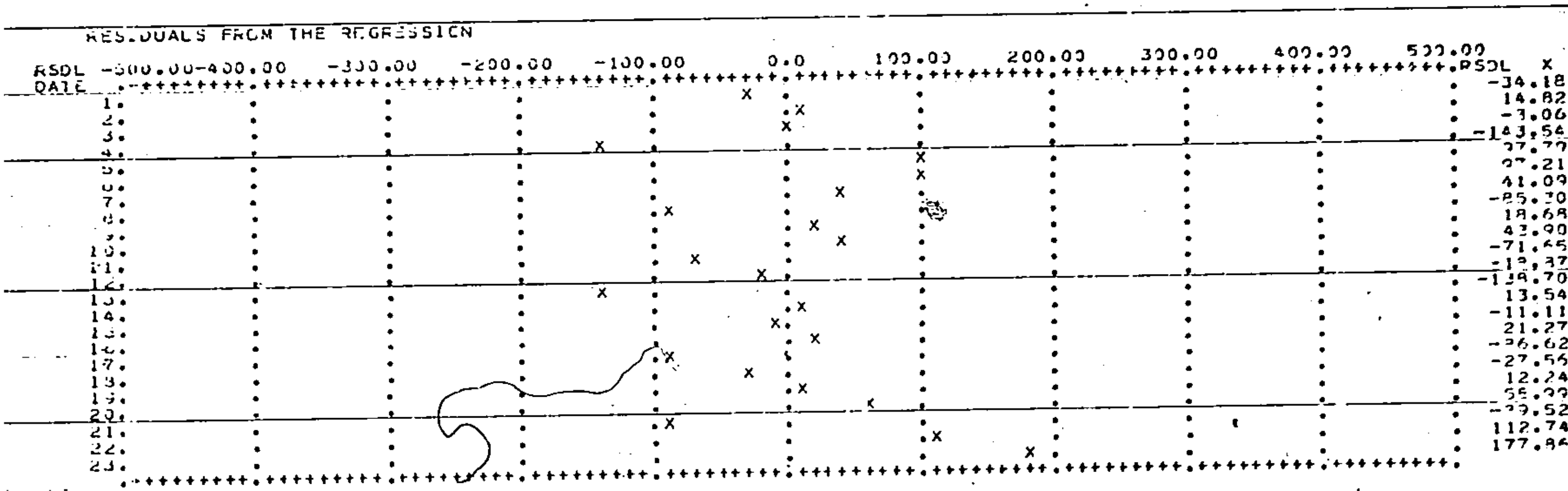


Figure E12: Residuals (BCF) of Model 27.

POOR COPY
COPIE DE QUALITEE INFERIEURE

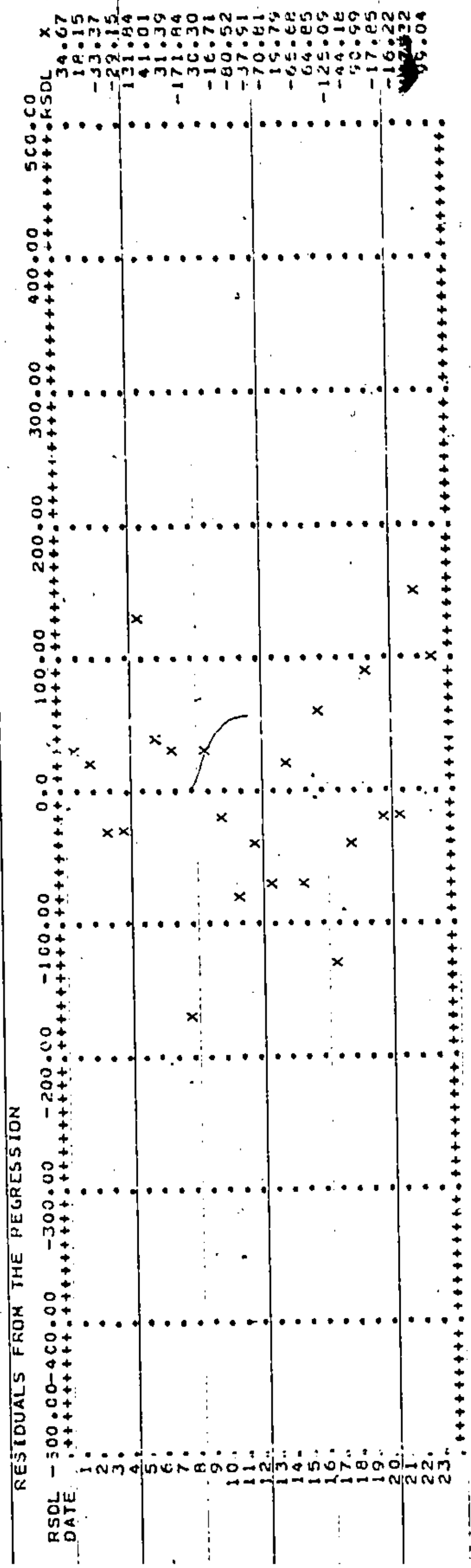


Figure E13: Residuals (BCF) of Model 28.

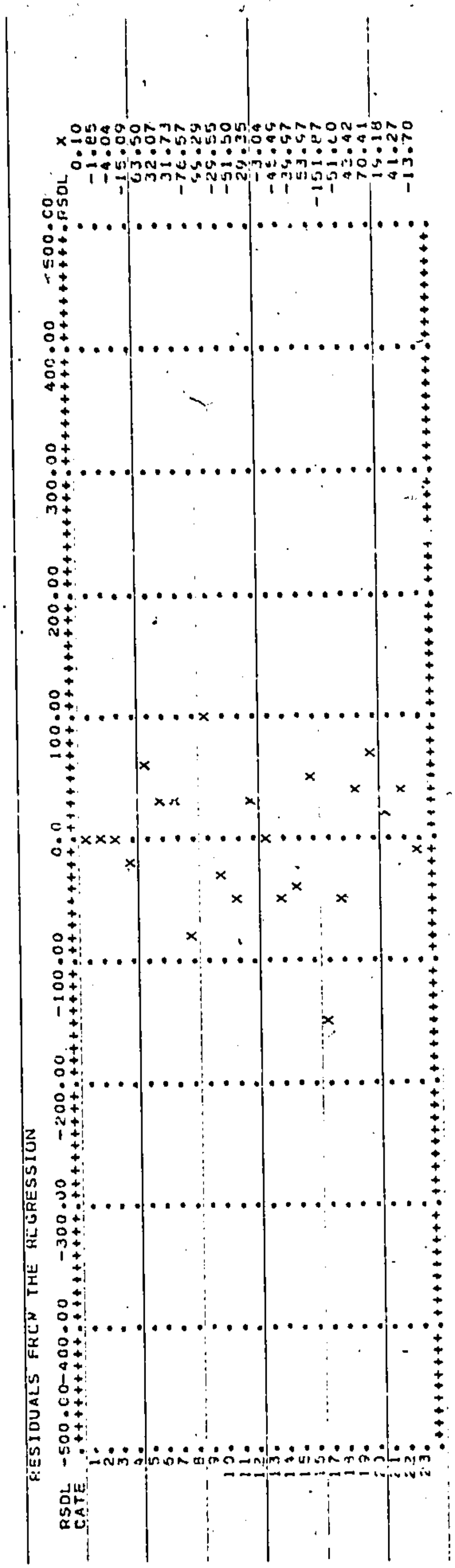


Figure E14: Residuals (BCF) of Model 29.

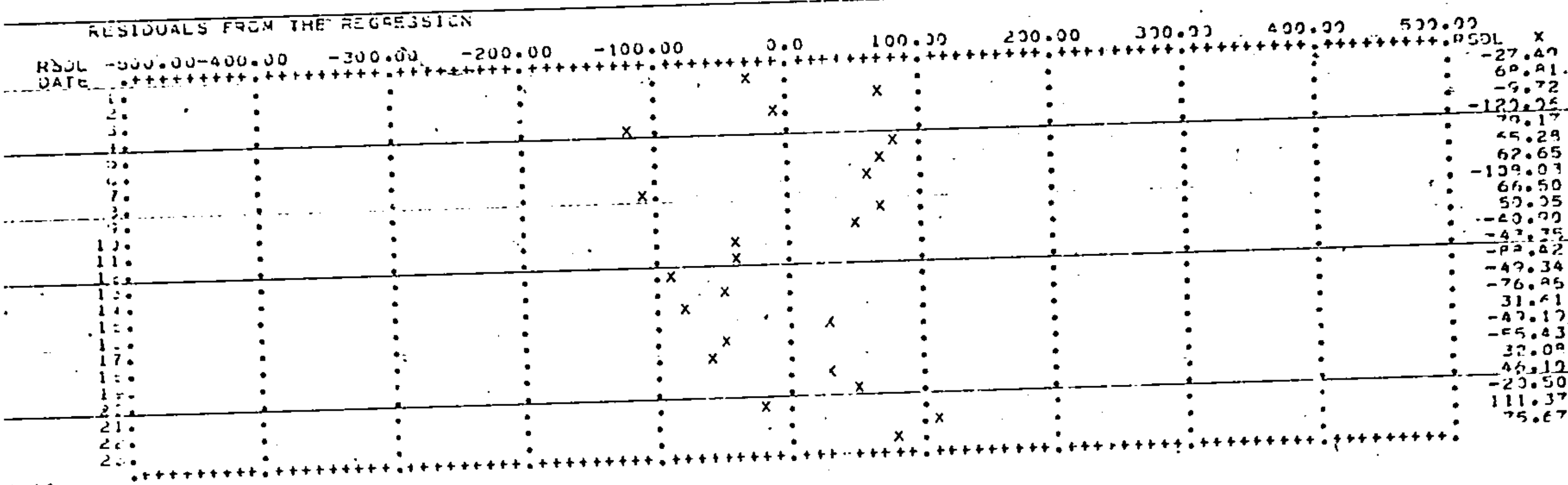


Figure E15: Residuals (BCF) of Model 30.

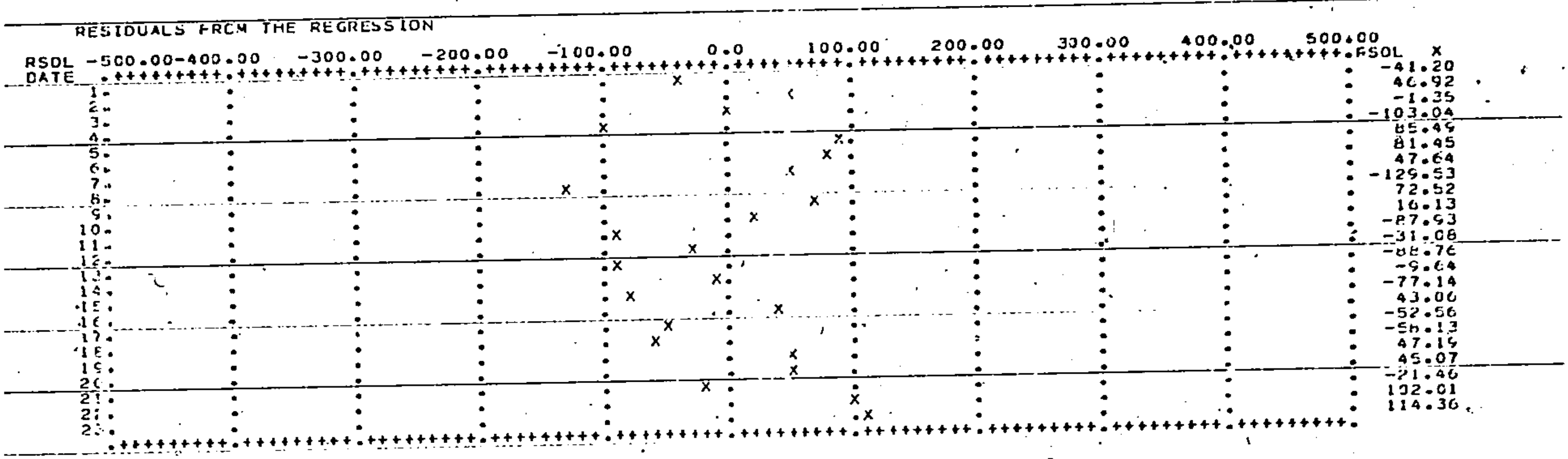


Figure E16: Residuals (BCF) of Model 31.

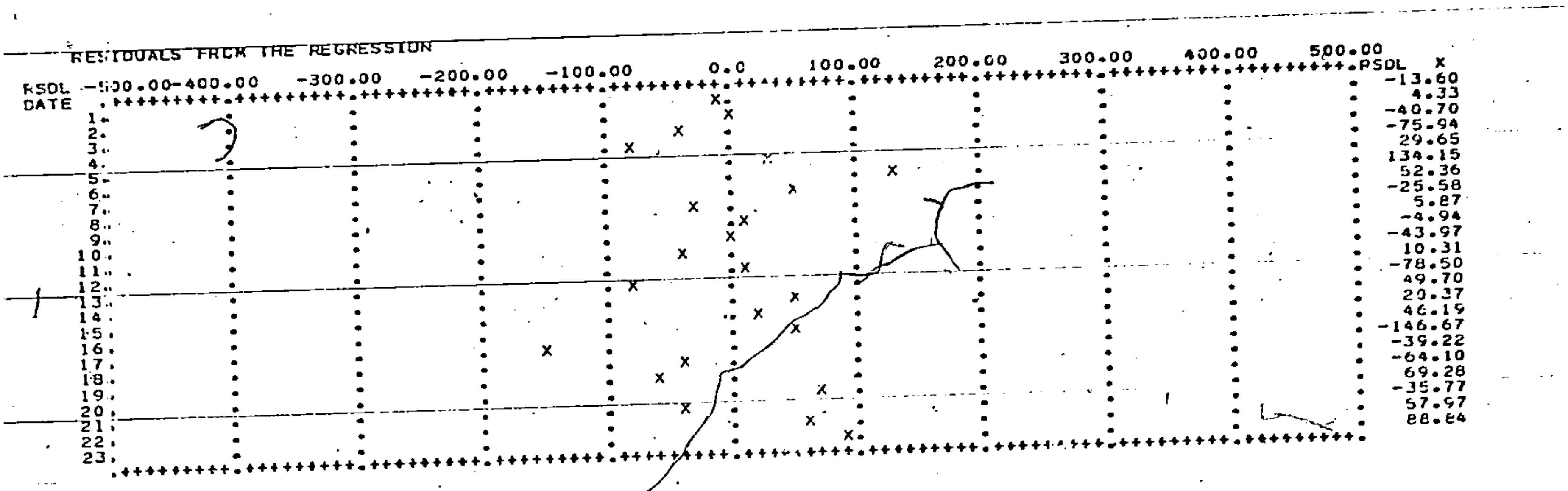


Figure E17: Residuals (BCF) of Model 32.

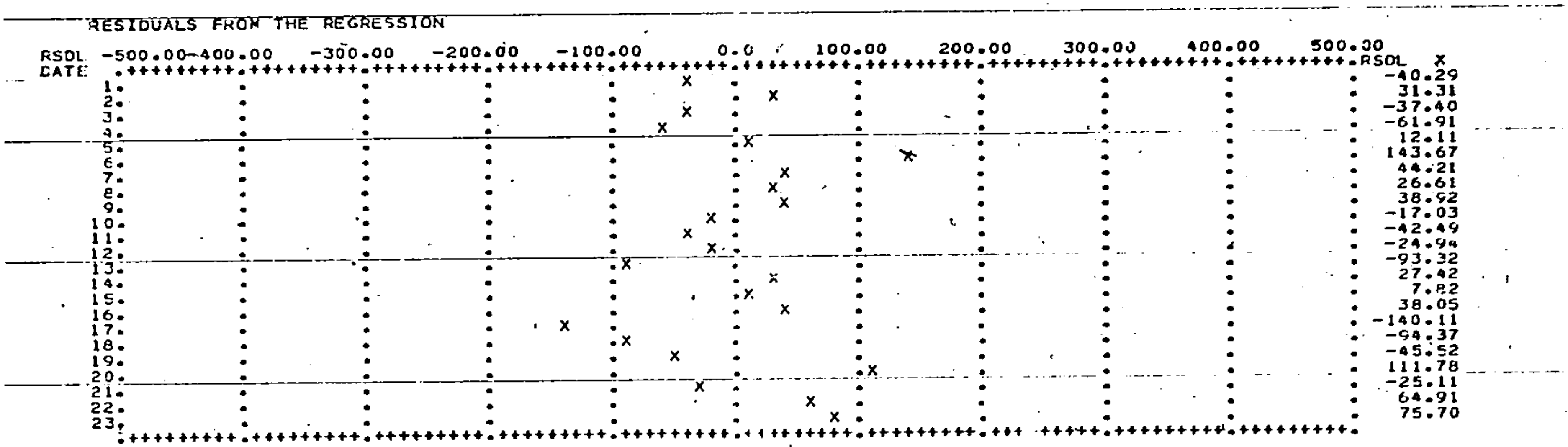


Figure E18: Residuals (BCF) of Model 33.

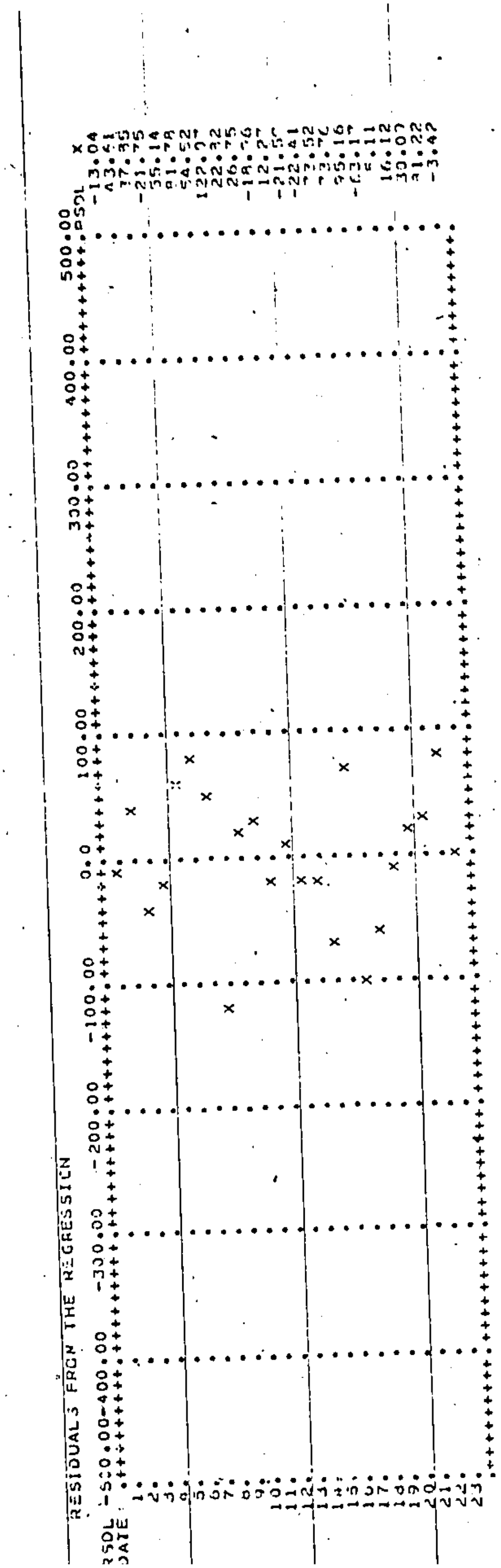


Figure E19 : Residuals (BCF) of Model 34.

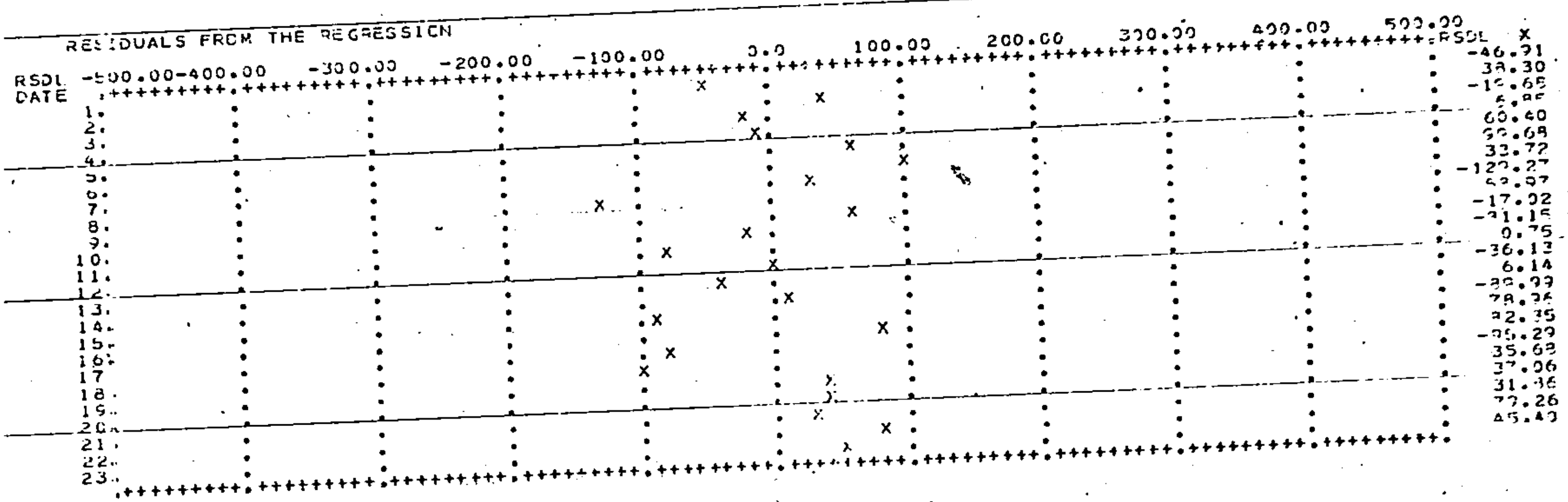


Figure E20: Residuals (BCF) of Model 35.

Appendix F

SCATTERGRAMS OF MODELS 16 THROUGH 35

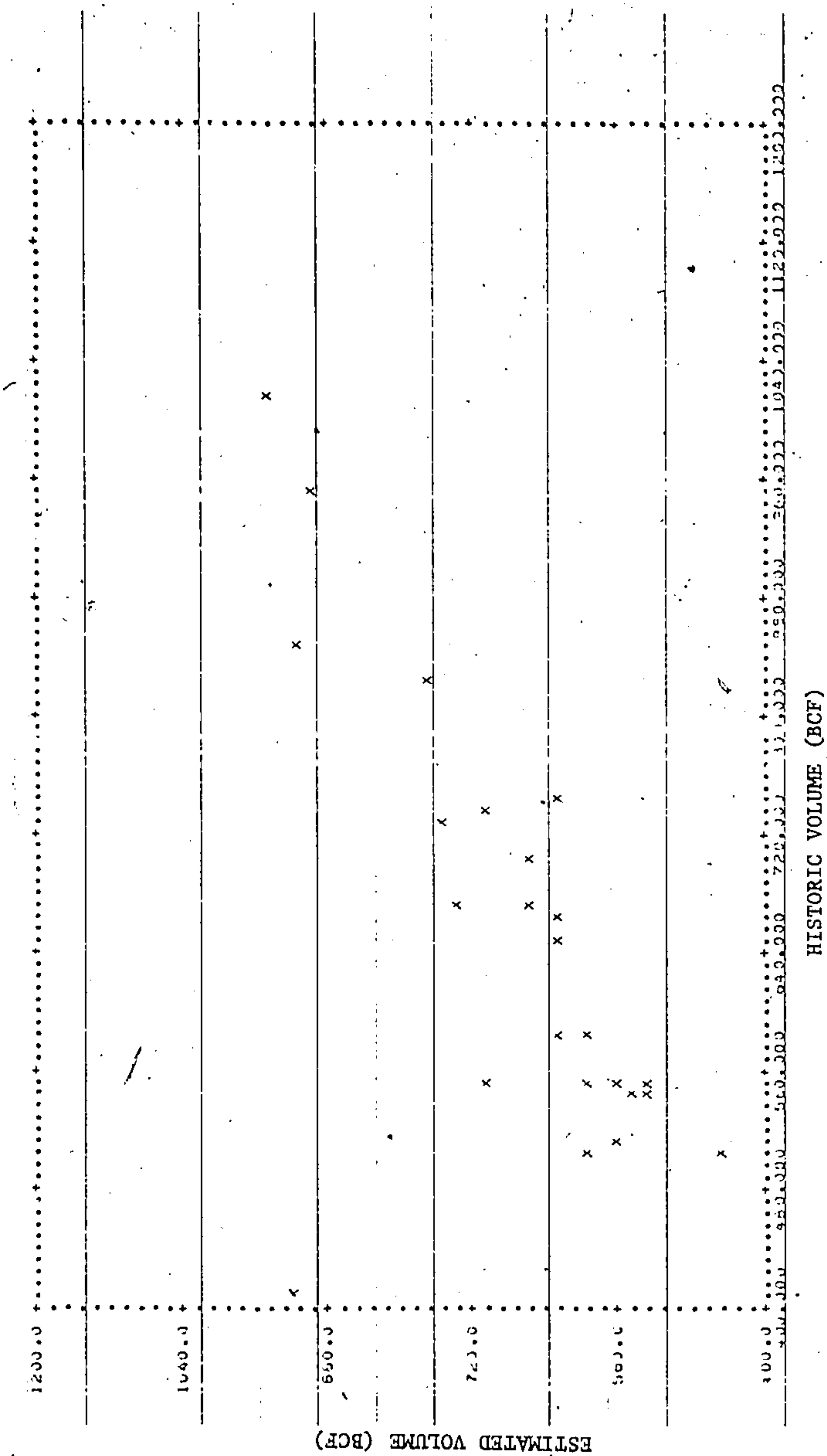


Figure F1: Scattergram for Spring Water Yield (BCF) - Model 16

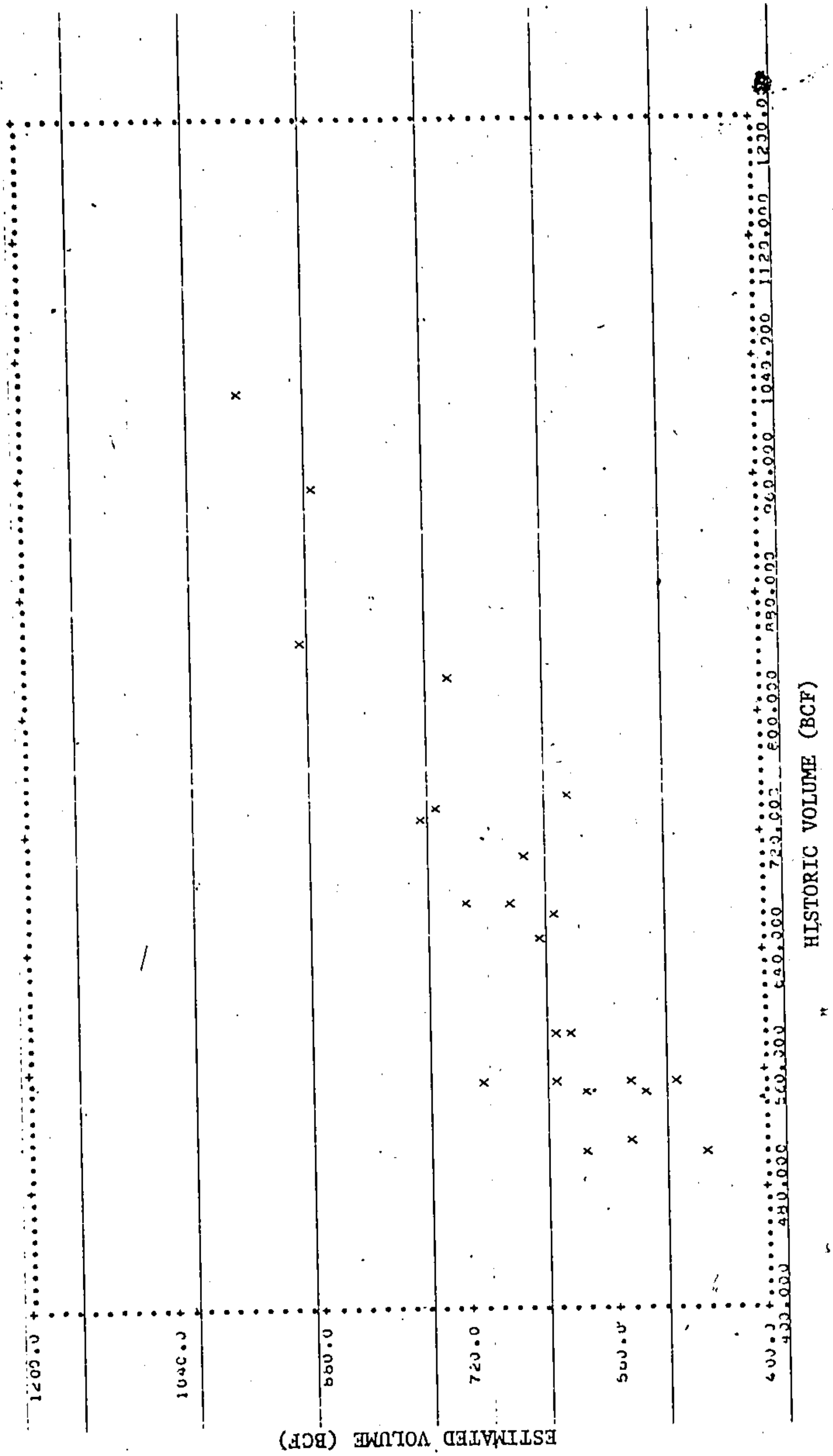


Figure F2: Scattergram for Spring Water Yiel (BCF) - Model 17.

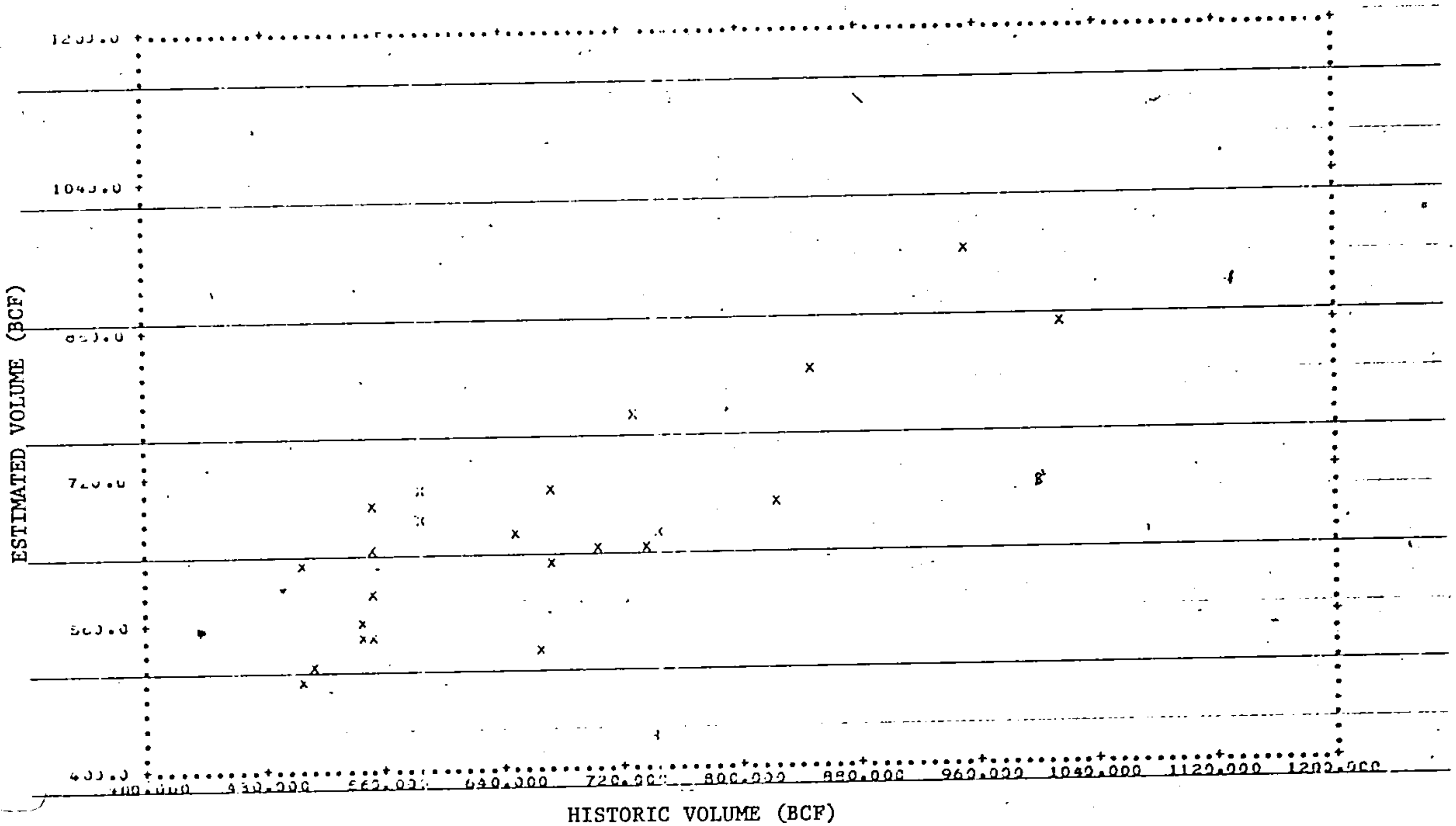


Figure F3: Scattergram for Spring Water Yield (BCF) - Model 18.

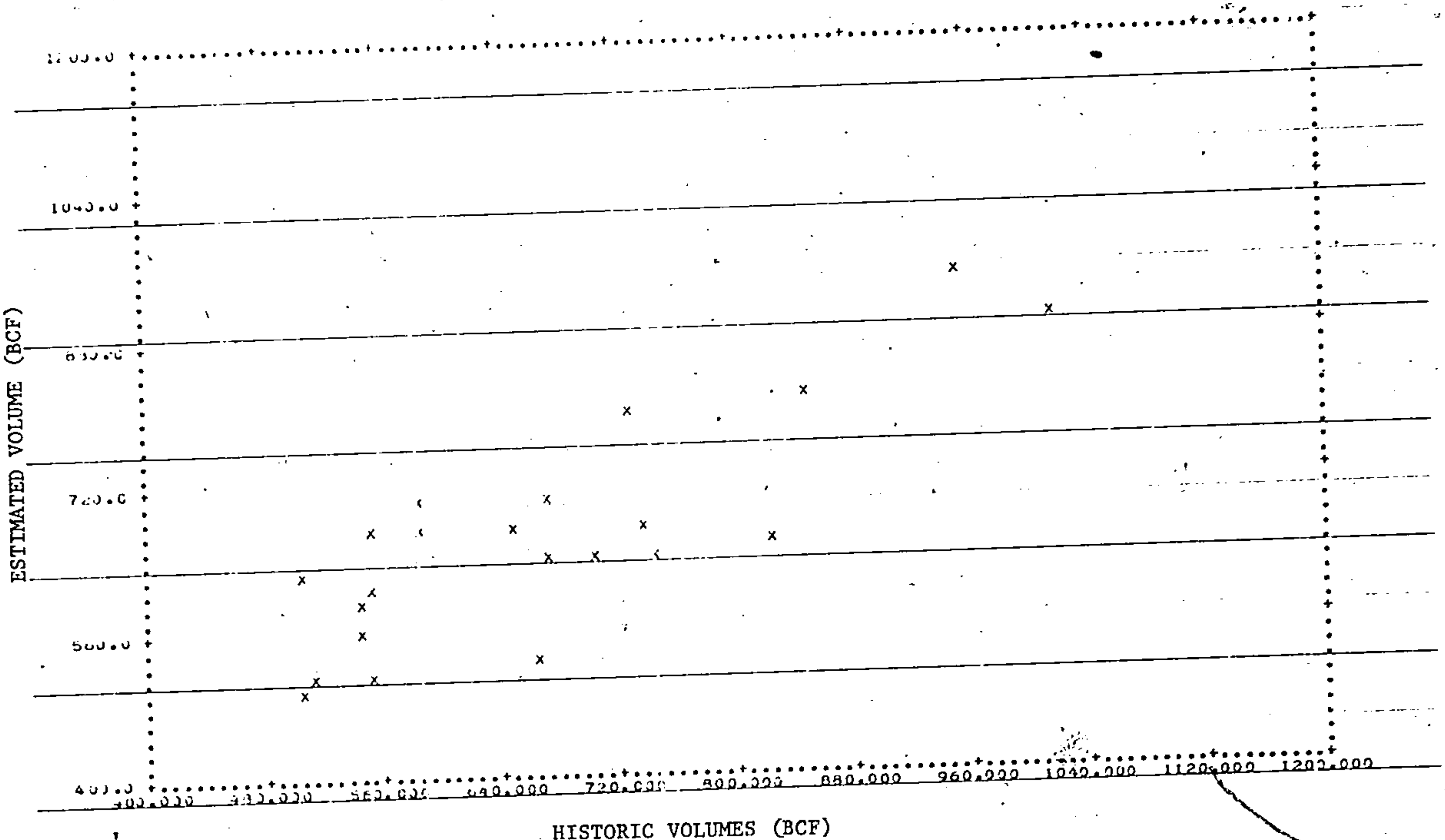


Figure F4: Scattergram for Spring Water Yield (BCF) - Model 19.

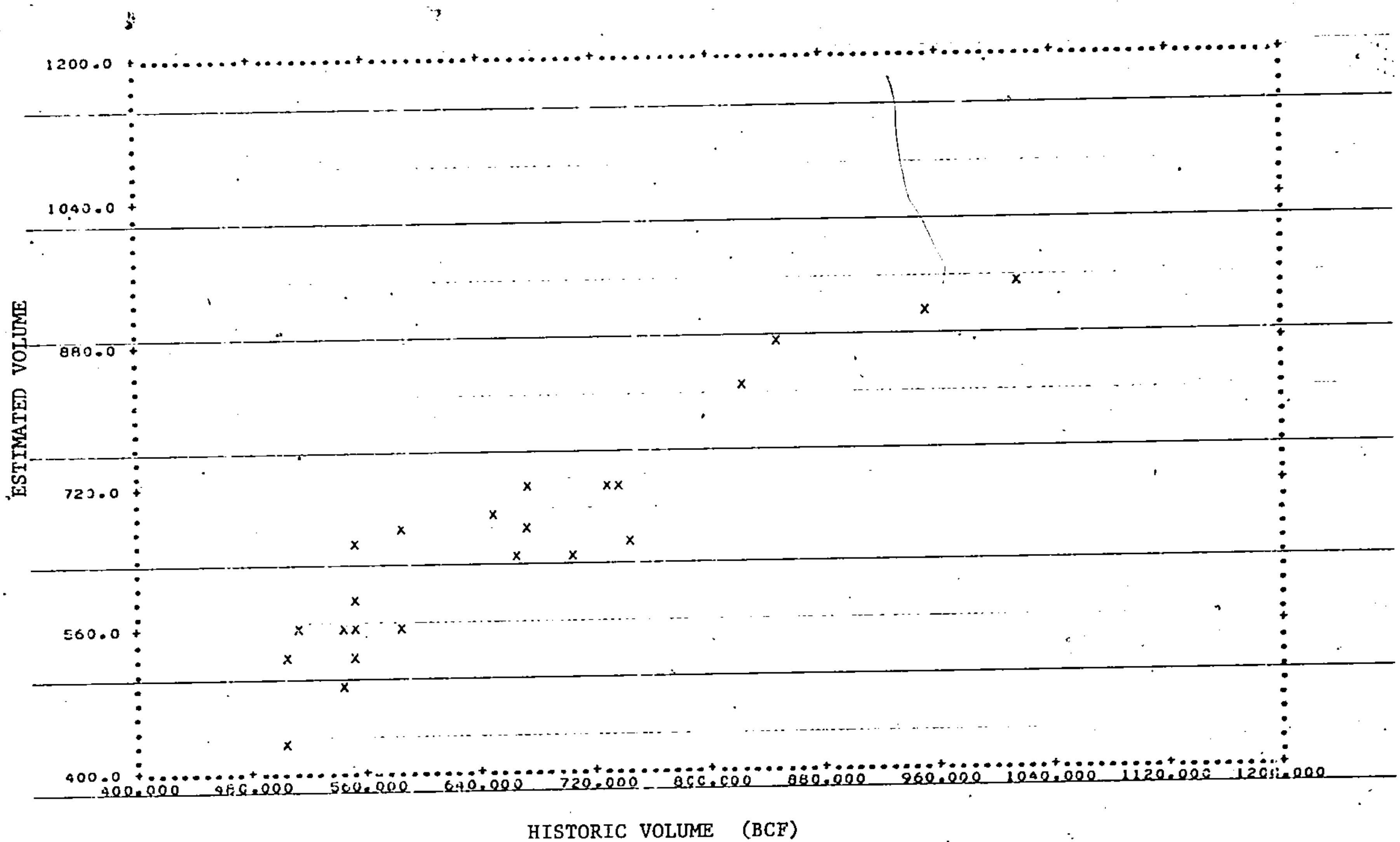


Figure F5: Scattergram for Spring Water Yield (BCF) - Model 20.

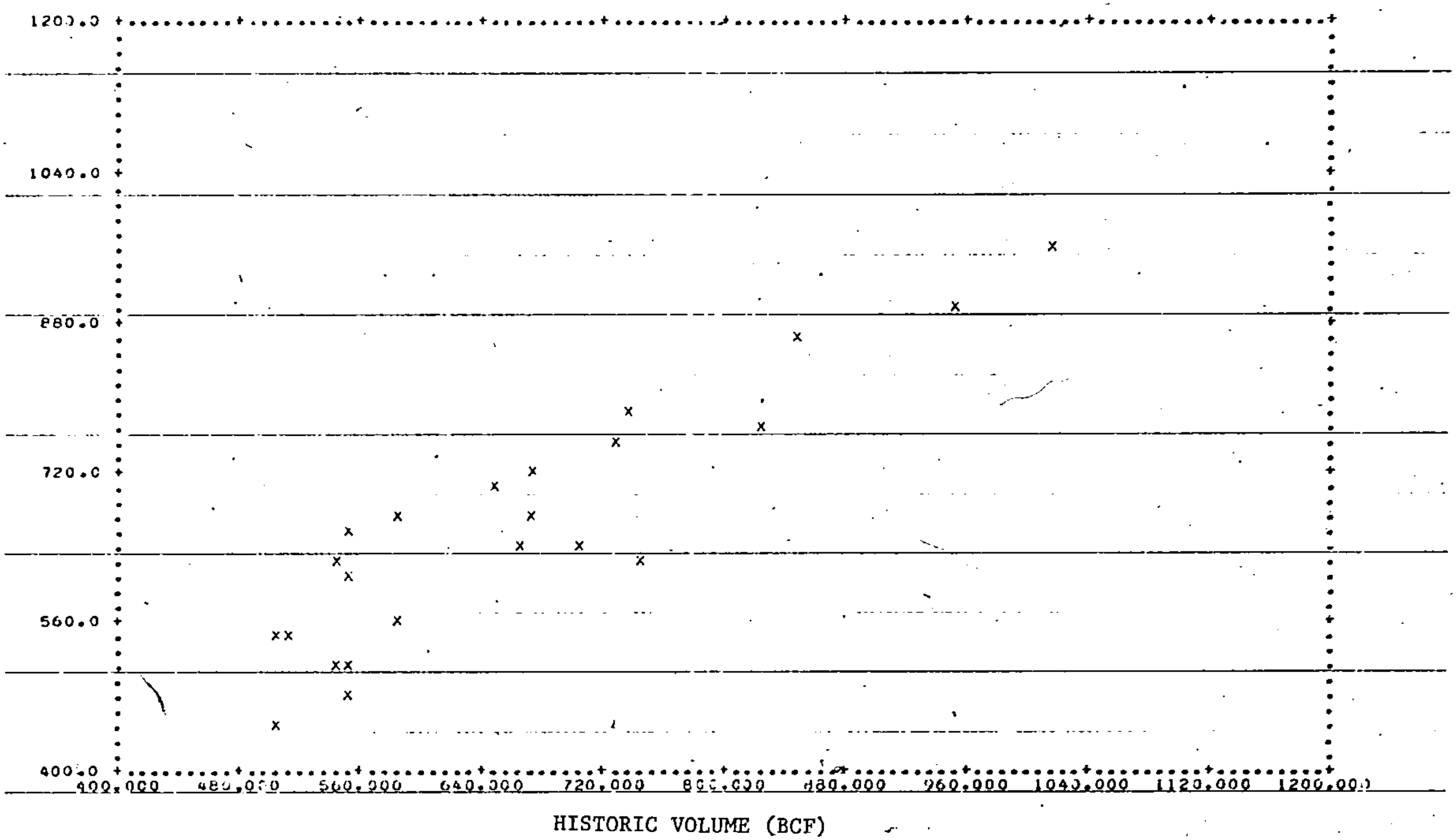


Figure F6: Scattergram for Spring Water Yield (BCF) - Model 21.

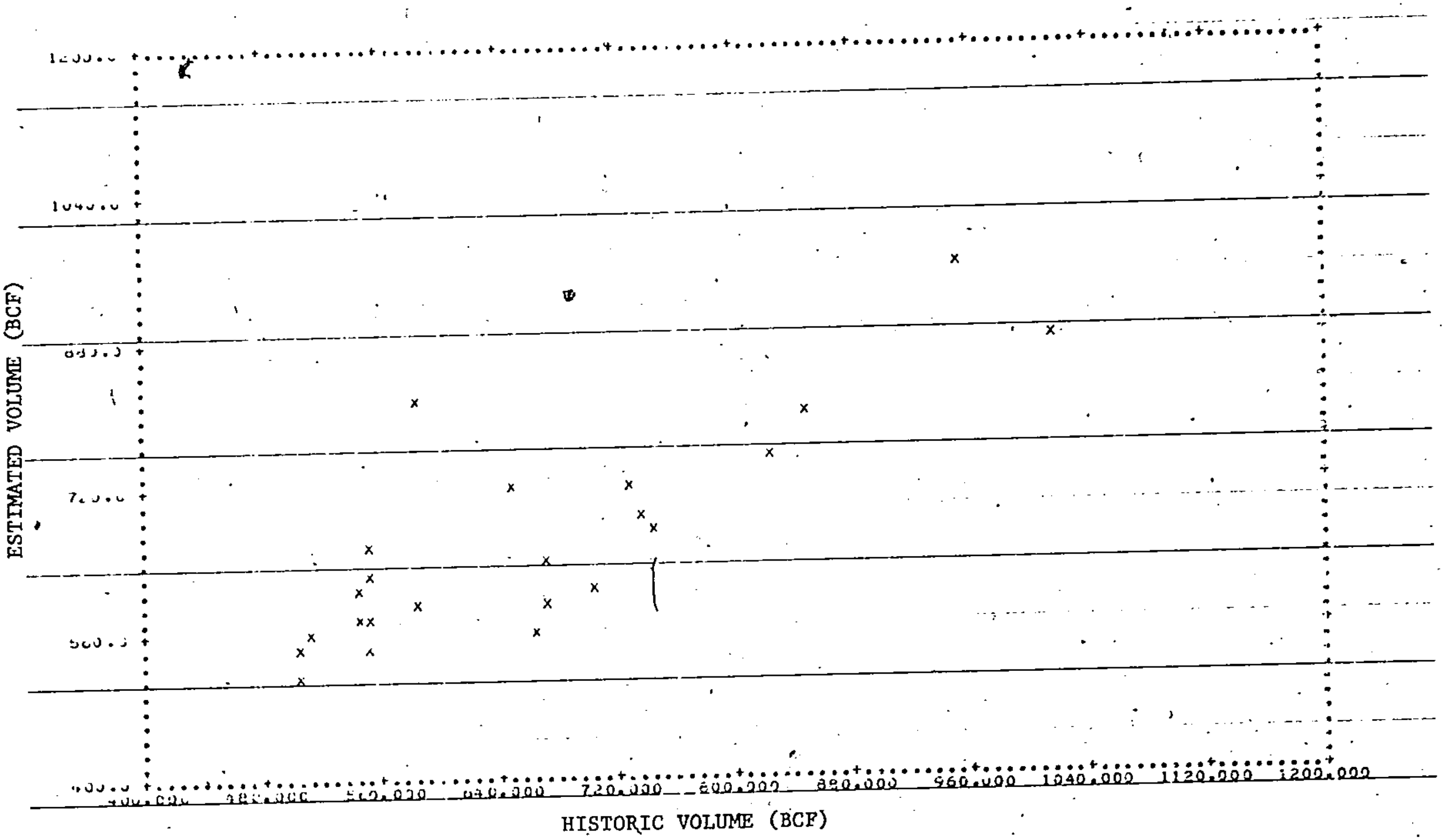


Figure F7: Scattergram for Spring Water Yield (BCF) - Model 22.

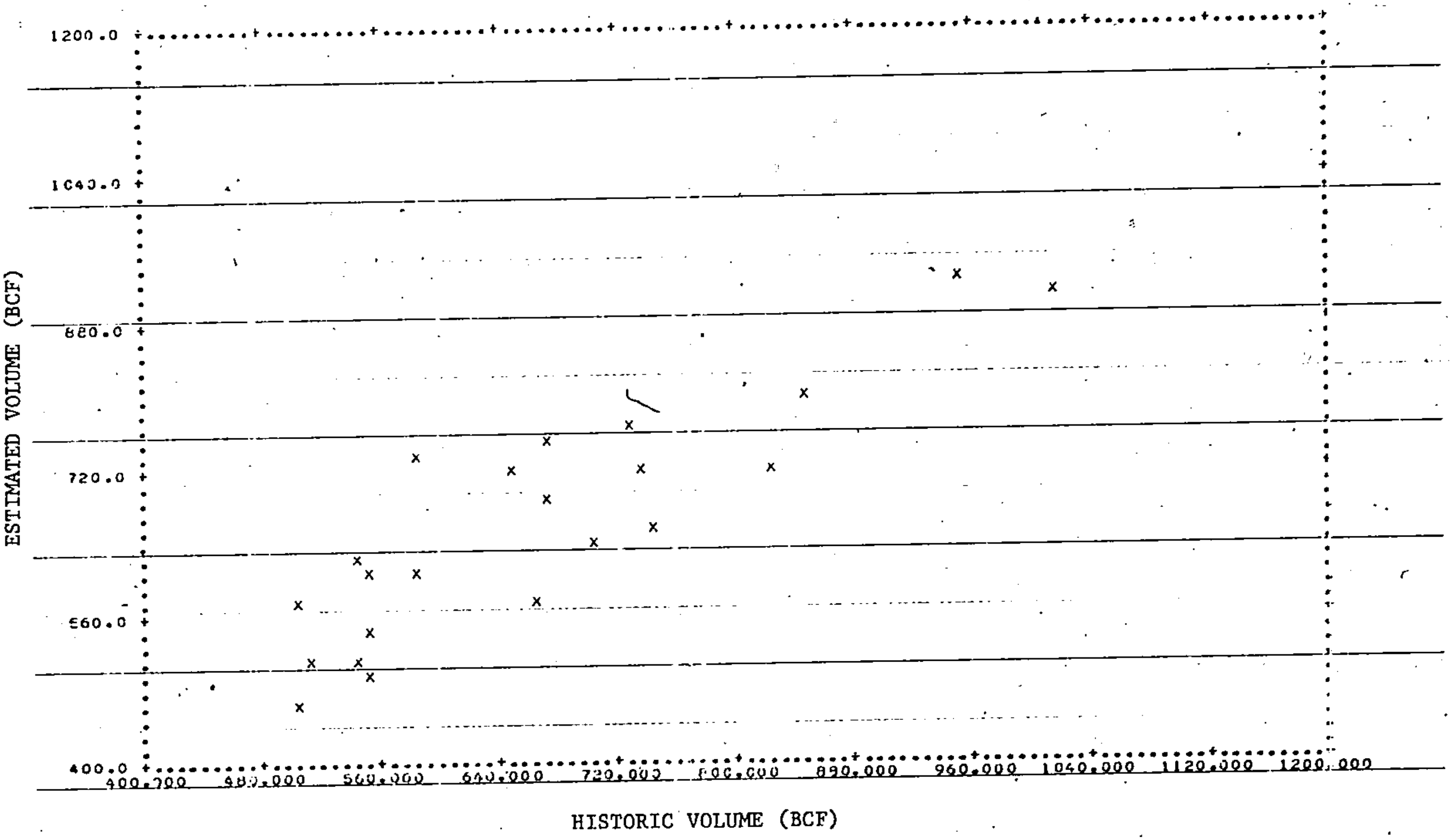


Figure F8: Scattergram for Spring Water Yield (BCF) - Model 23.

POOR COPY
COPIE DE QUALITEE INFERIEURE

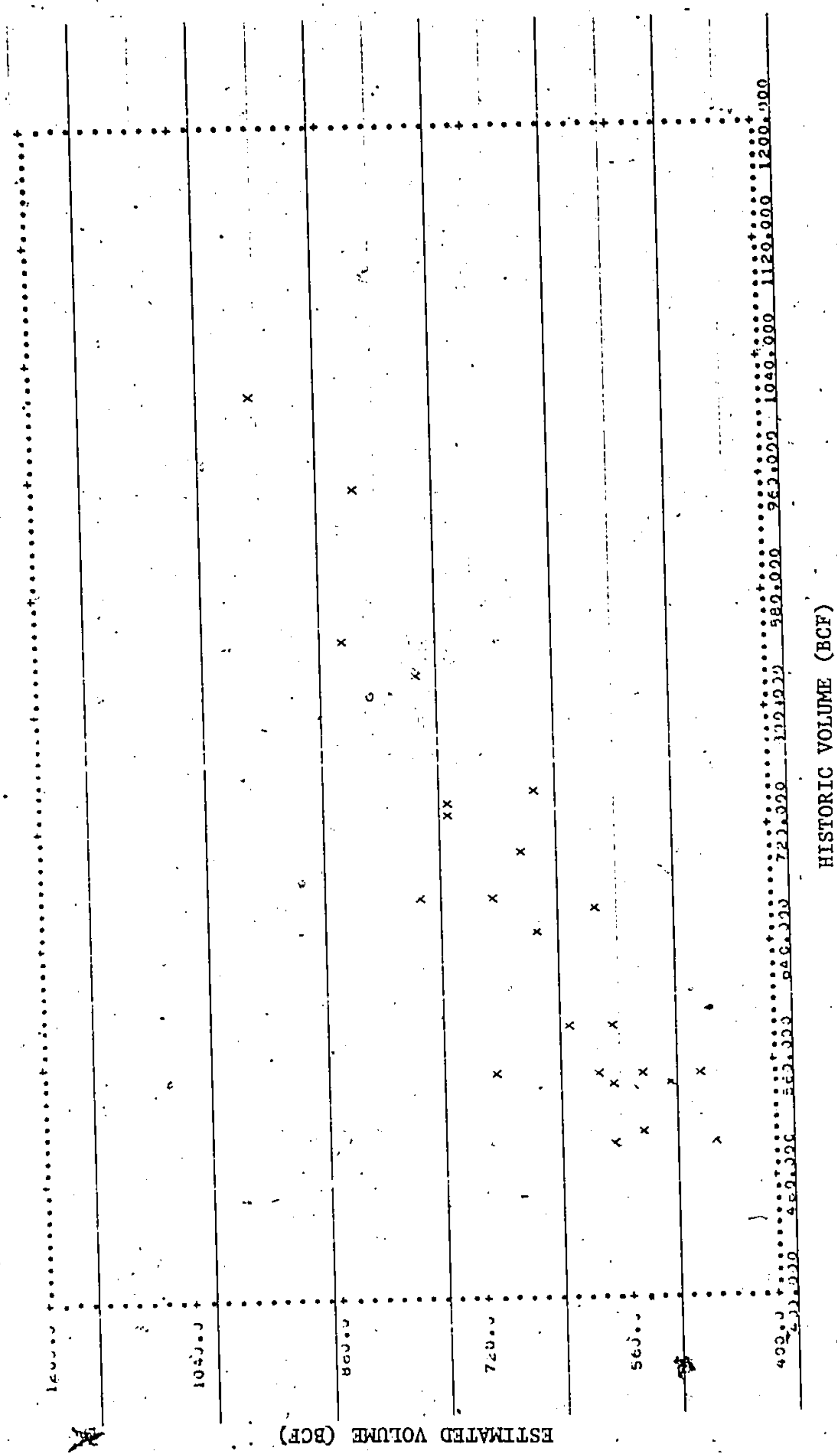


Figure F9: Scattergram for Spring Water Yield (BCF) - Model 24.

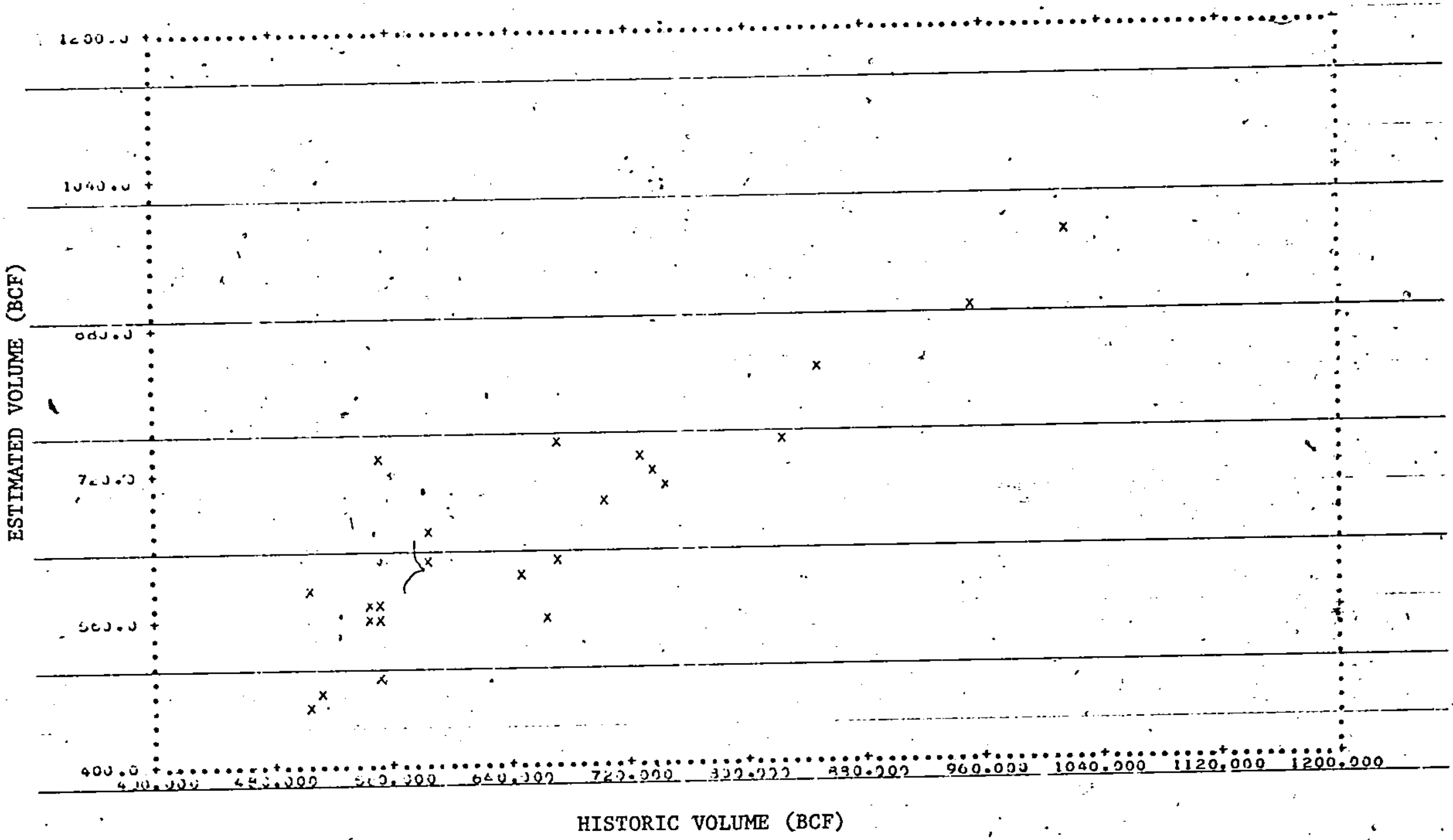


Figure F10: Scattergram for Spring Water Yield (BCF) - Model 25.

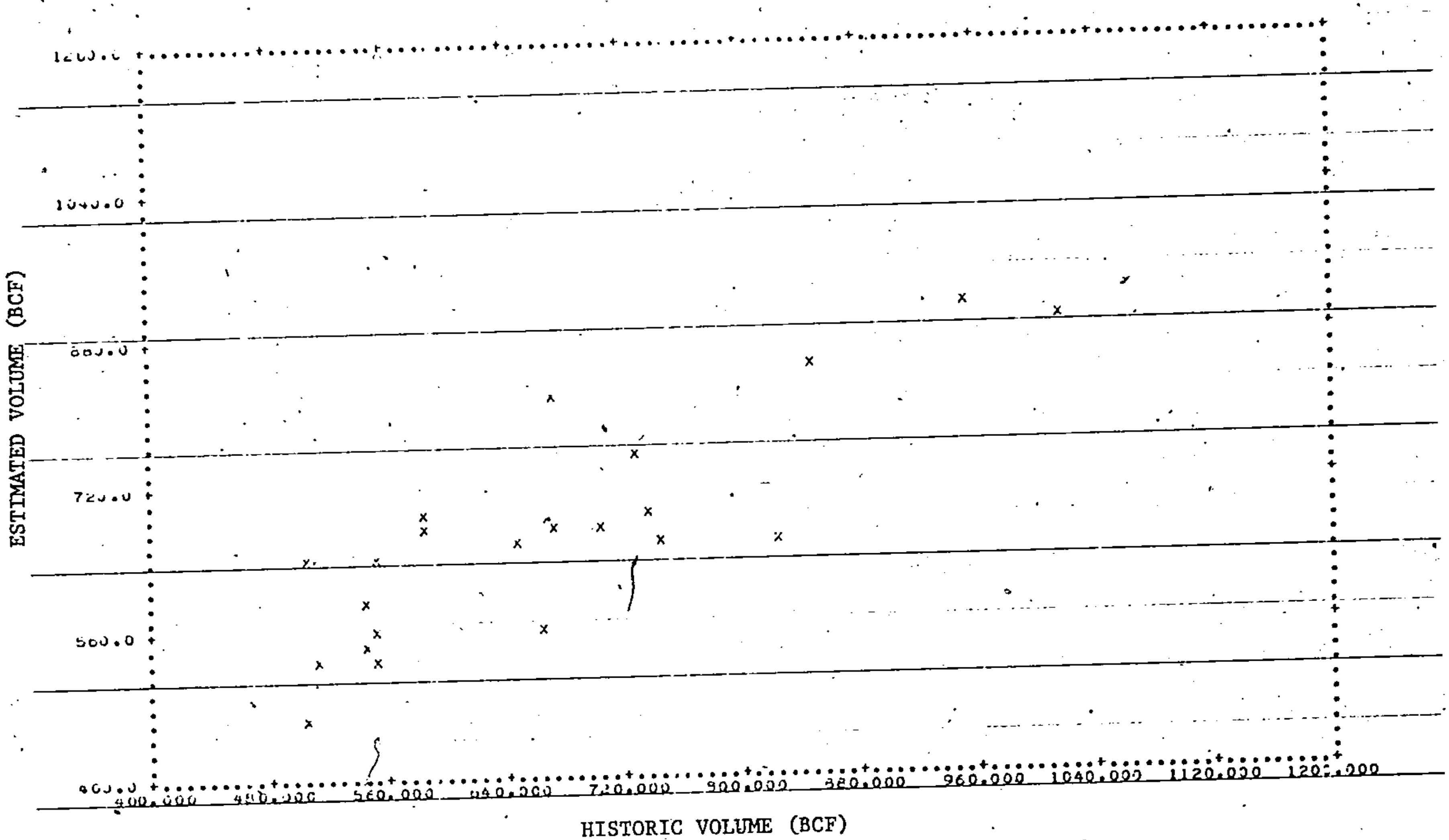


Figure F11: Scattergram for Spring Water Yield (BCF) - Model 26.

POOR COPY
COPIE DE QUALITEE INFERIEURE

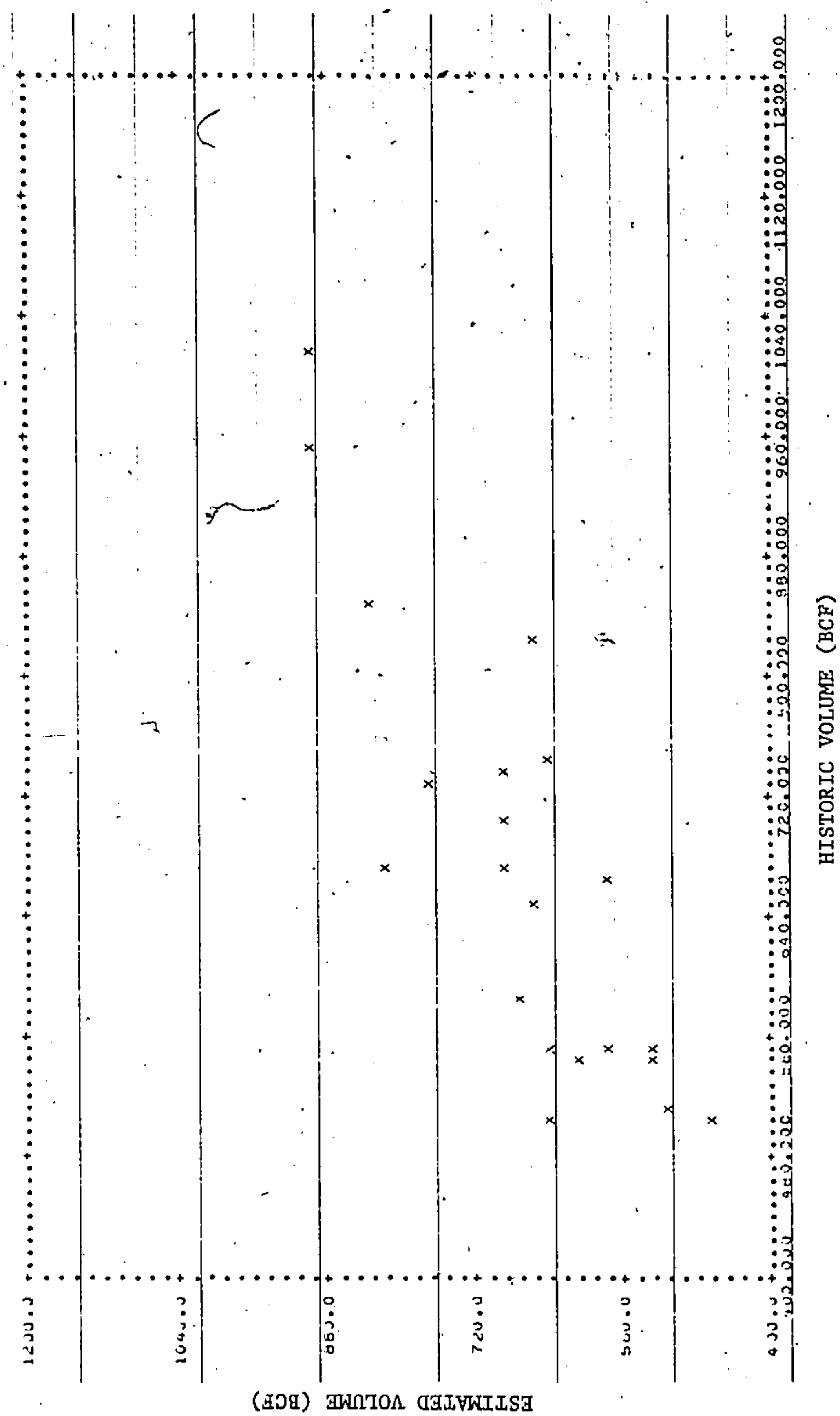


Figure F12: Scattergram for Spring Water Yield (BCF) - Model 27.

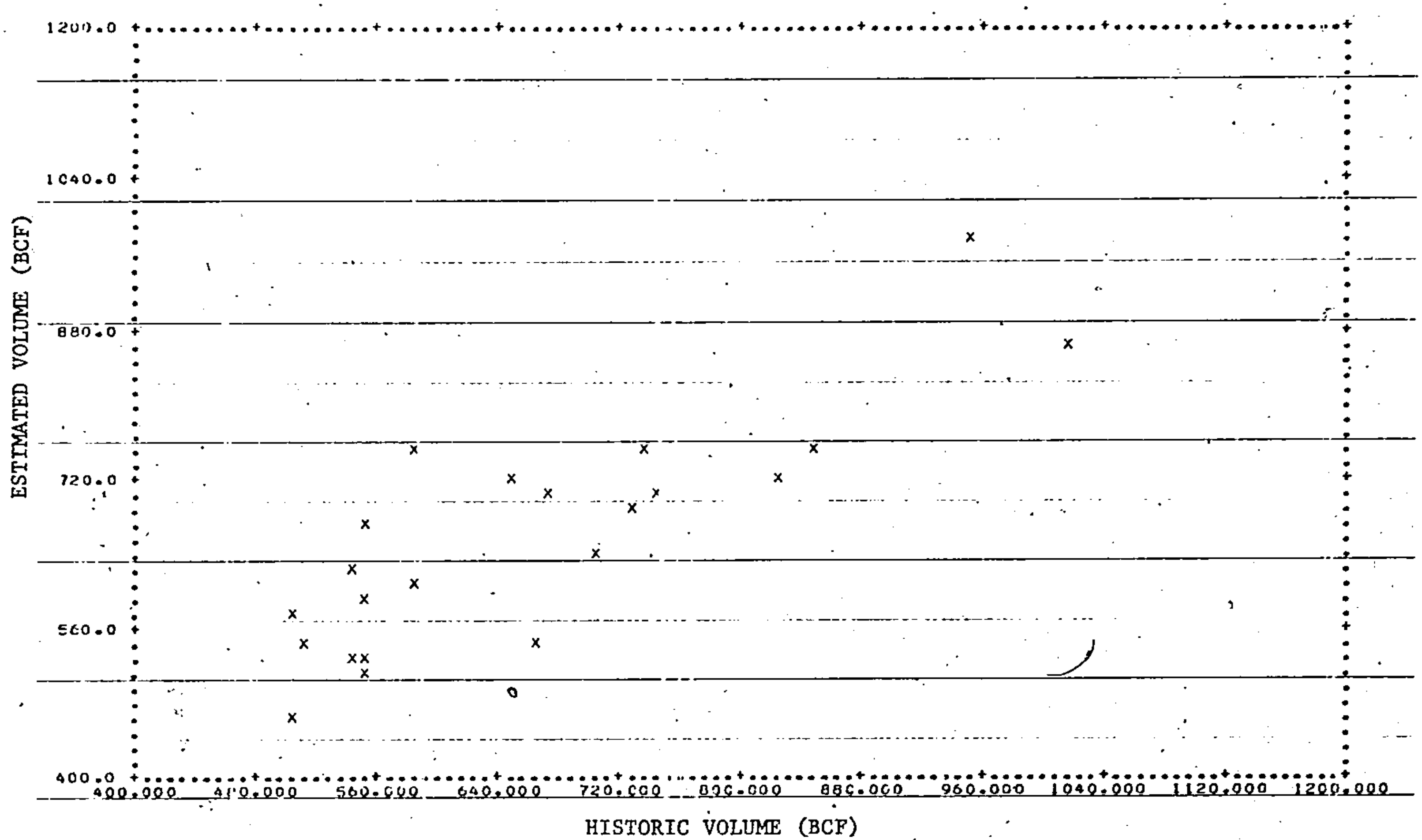


Figure F13: Scattergram for Spring Water Yield (BCF) - Model 28.

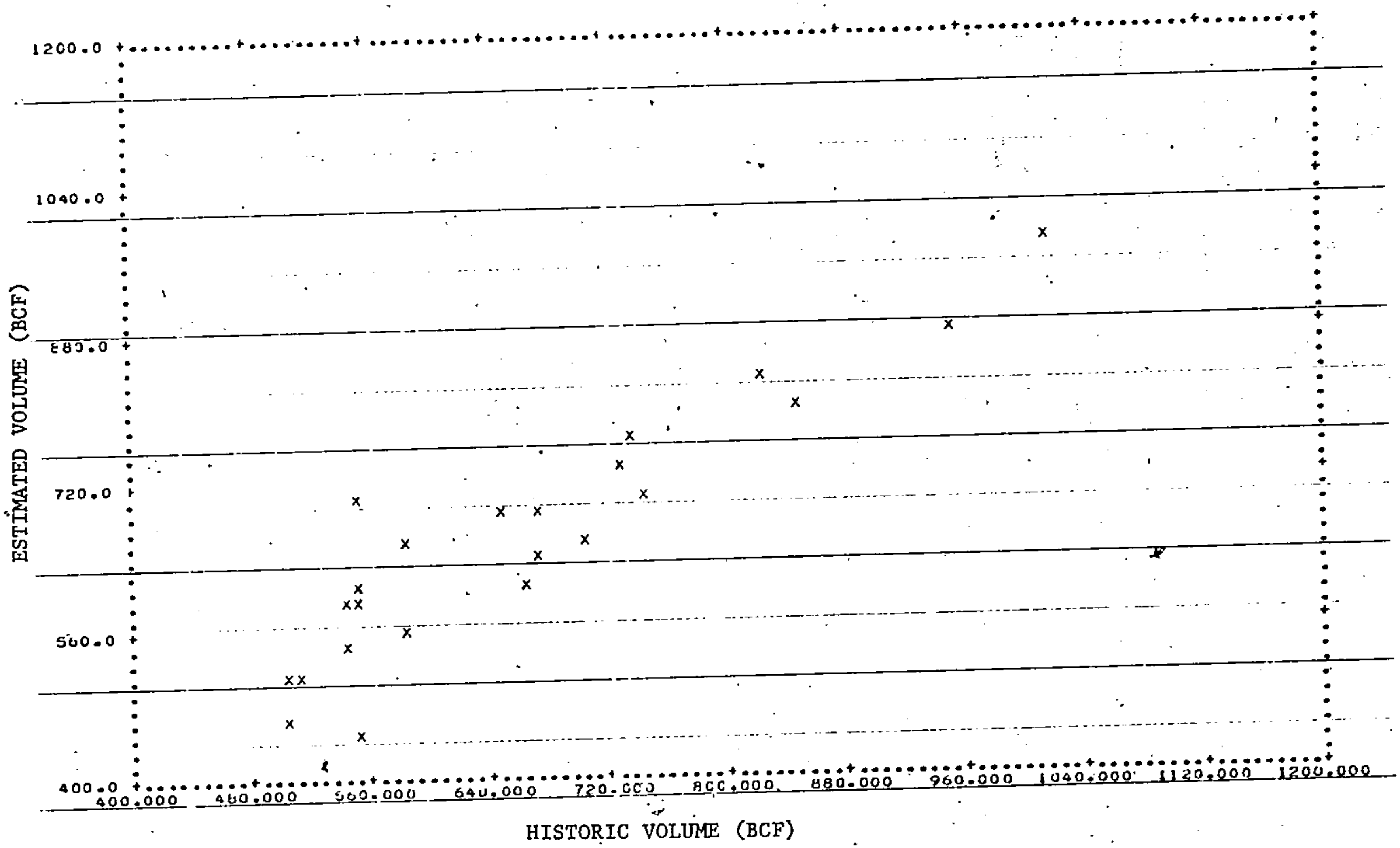


Figure F14: Scattergram for Spring Water Yield (BCF) - Model 29.

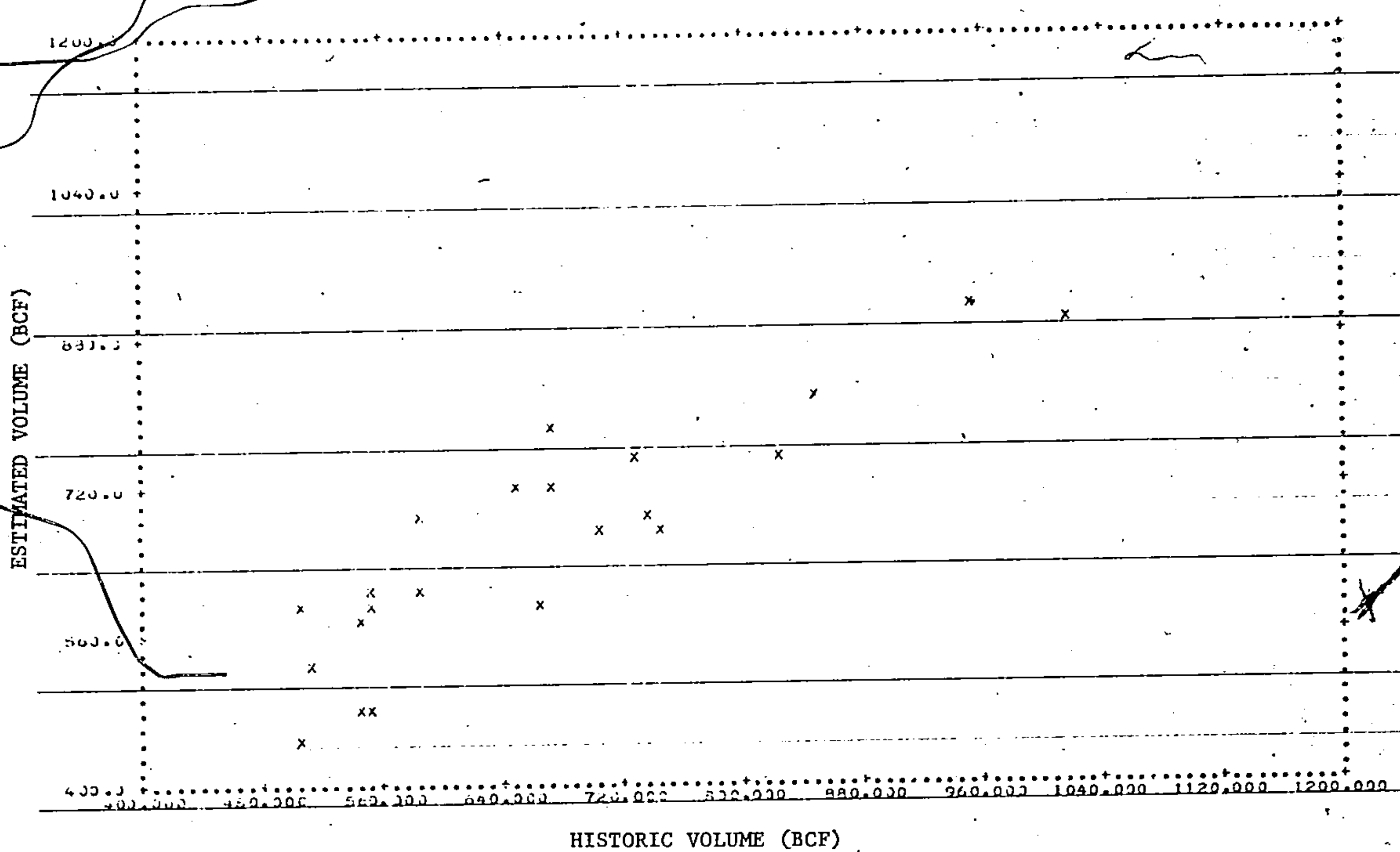


Figure F15: Scattergram for Spring Water Yield (BCF) - Model 30.

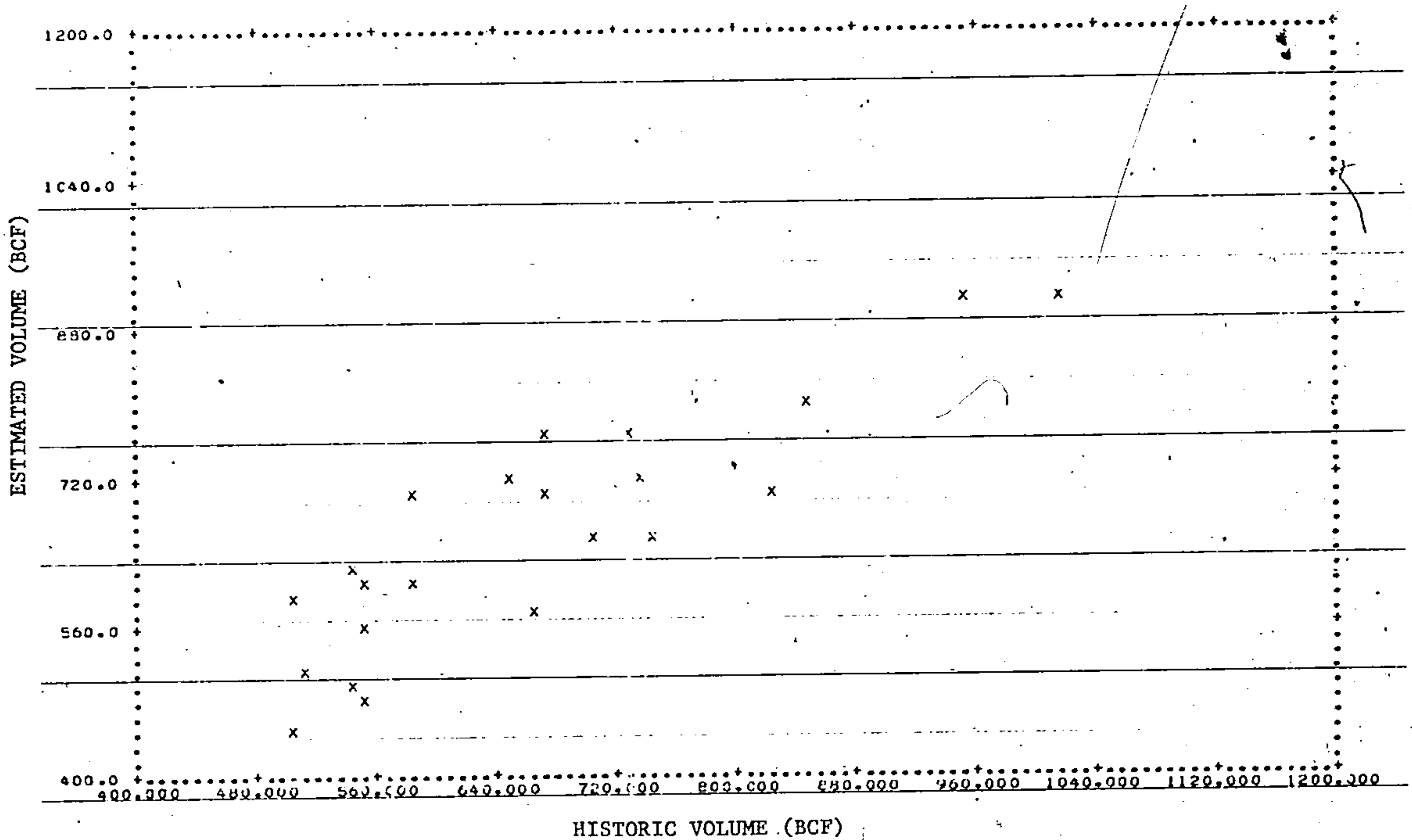


Figure F16: Scattergram for Spring Water Yield (BCF) - Model 31.

POOR COPY
COPIE DE QUALITEE INFERIEURE

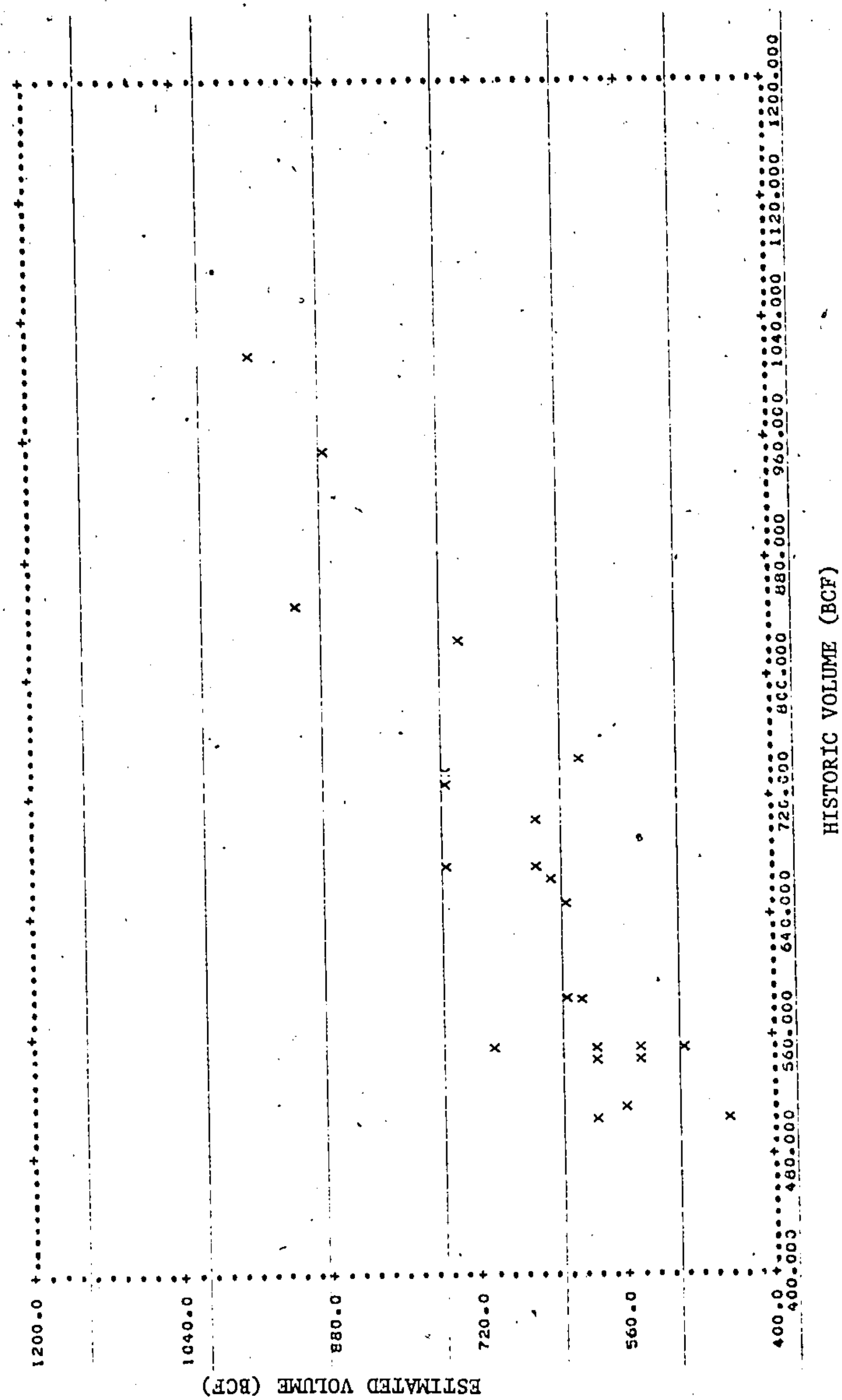


Figure F17: Scattergram for Spring Water Yield (BCF) - Model 32.

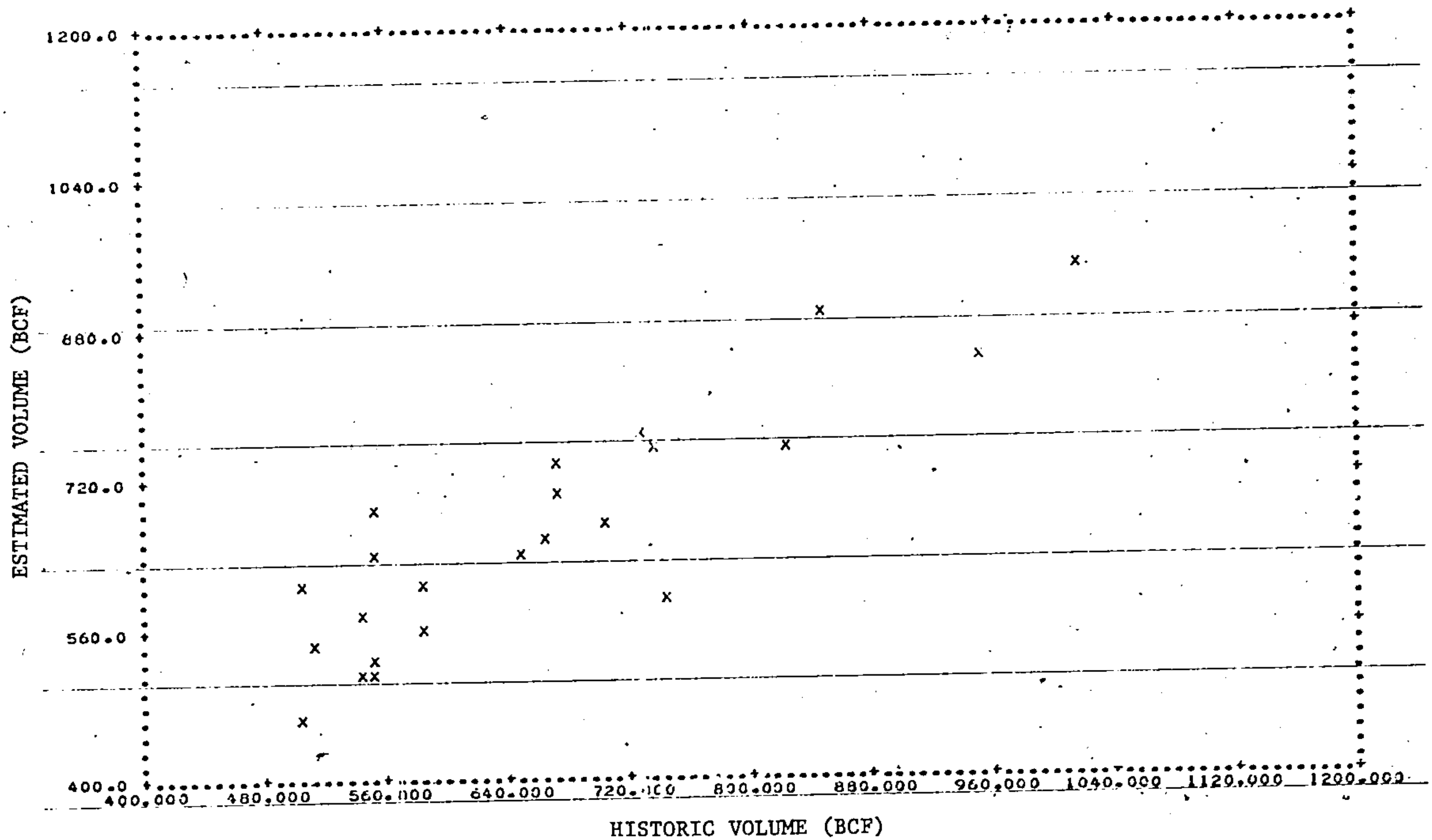


Figure F18: Scattergram for Spring Water Yield (BCF) - Model 33.

POOR COPY
COPIE DE QUALITEE INFERIEURE

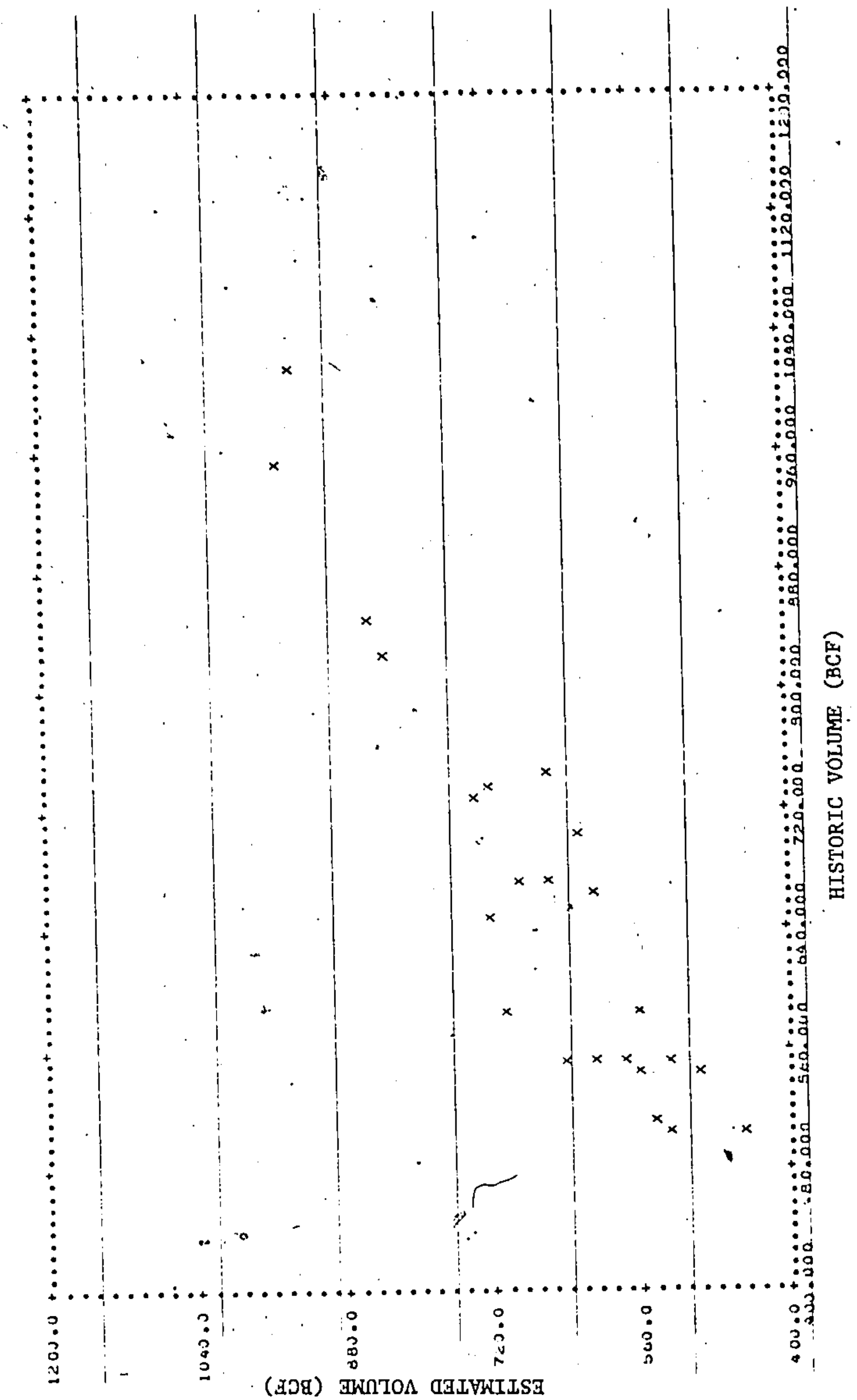


Figure F19: Scattergram for Spring Water Yield (BCF) - Model 34.

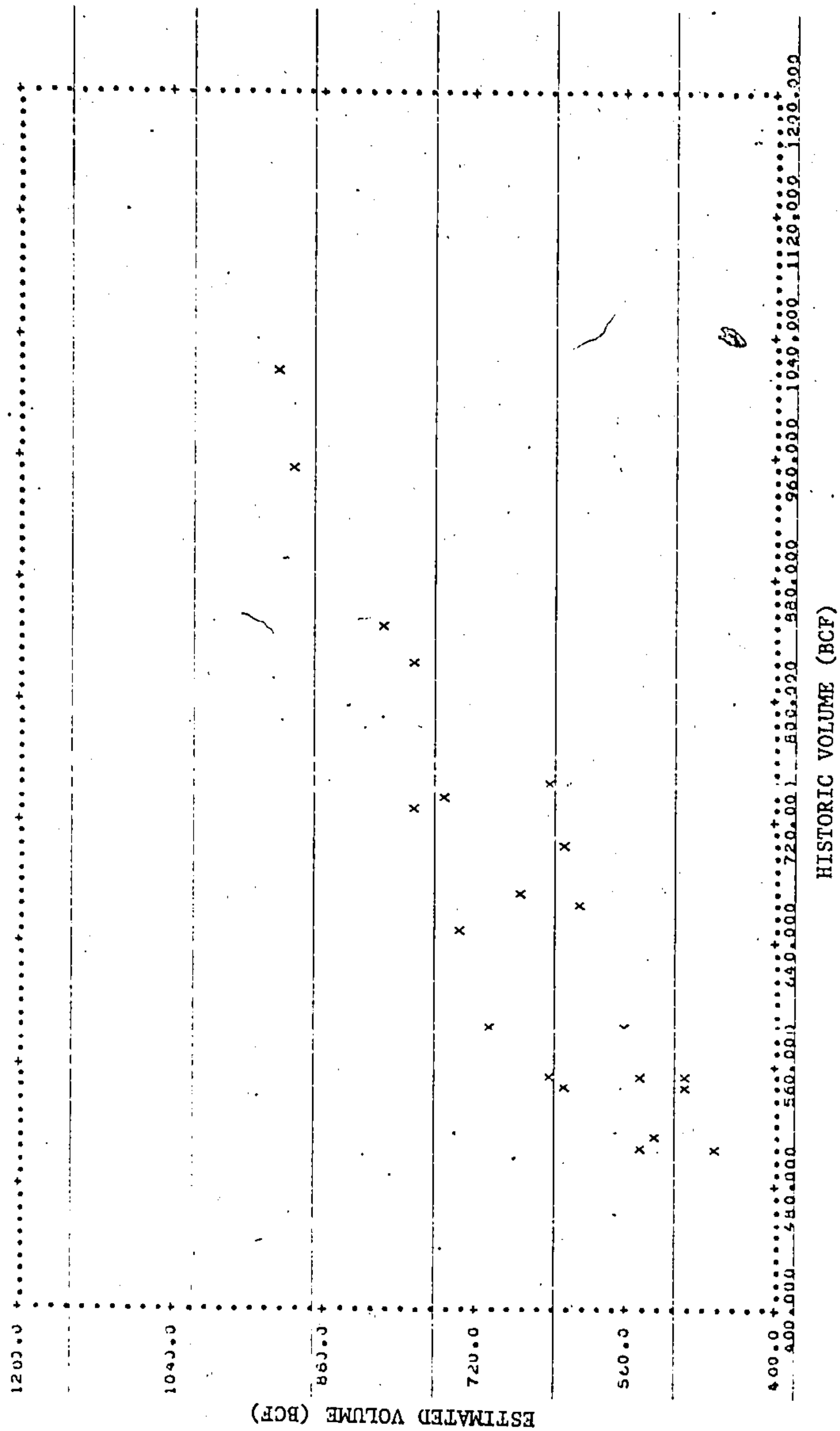


Figure F20: Scattergram for Spring Water Yield (BCF) - Model 35.

Appendix G

CORRELOGRAM OF SPRING FLOOD VOLUMES (1913-1977)

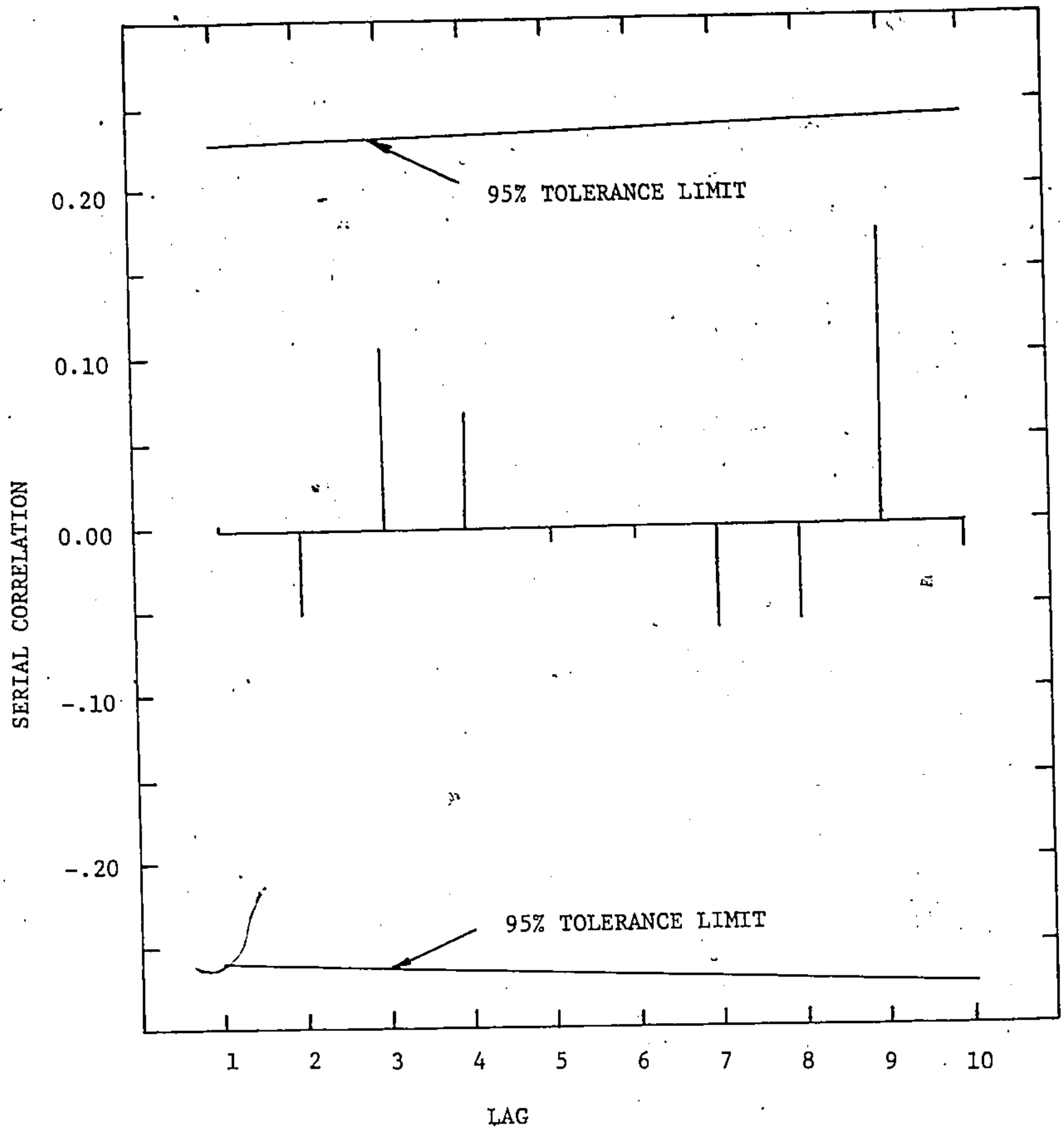


Figure G1: Correlogram of spring flood season volumes for the total Lac St-Jean watershed (1913-1977).