DOCUMENT RESUME

ED 112 896 IR 002 607

AUTHOR Christensen, John O.
TITLE A Comparison of Arrival Algorithms for Automated Serial Claiming Operations.
INSTITUTION Brigham Young Univ., Provo, Utah. Graduate Dept. of Library and Information Sciences.
PUB DATE 21 Mar 75
NOTE 76p.

EDRS PRICE MF-$0.76 HC-$4.43 Plus Postage
DESCRIPTORS *Algorithms; Computer Oriented Programs; *Library Automation; Library Research; Library Technical Processes; Program Descriptions; *Serials
IDENTIFIERS *Serials Claiming

ABSTRACT

Heretofore, methods of claiming missing issues of serials publications have been ad hoc. The study describes four methods of claiming and analyzes the performance of three of the methods on 16 arrival patterns of serials. The arrival patterns, selected as workable examples from a random sample of the Kardex files at the University of Utah, were graphed against a statistically-based claiming algorithm and two algorithms based on the arrival frequency plus a lag factor. The statistically-based algorithm produced more false claims and usually claimed issues sooner than the algorithm based on the lag factor. The statistically-based algorithm worked best with the medium frequency (monthly, quarterly, etc.) serials and the lag factor algorithm appeared to work best with the long- and short-frequency (annuals, weekly, etc.) serials. (Author)

****
* Documents acquired by ERIC include many infcrmal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED 122696

A COMPARISON OF ARRIVAL ALGORITHMS FOR AUTOMATED SERIAL CLAIMING OPERATIONS

A Research Paper

Submitted to The

Graduate Department of Library and Information Sciences

Brigham Young University

Provo, Utah

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

In Partial Fulfillment

of the Requirements of

L. I. S. 697

by

John O. Christensen

March 21, 1975

ED 002 607

ABSTRACT

Heretofore, methods of claiming missing issues of serials publications have been ad hoc. The study describes four methods of claiming and analyzes the performance of three of the methods on sixteen arrival patterns of serials. The arrival patterns, selected as workable examples from a random sample of the Kardex files at the University of Utah, were graphed against a statistically-based claiming algorithm and two algorithms based on the arrival frequency plus a lag factor. The statistically-based algorithm produced more false claims and usually claimed issues sooner than the algorithm based on the lag factor. The statistically-based algorithm worked best with the medium frequency (monthly, quarterly, etc.) serials and the lag factor algorithm appeared to work best with the long and short frequency (annuals, weekly, etc.) serials.

TABLE OF CONTENTS

LIST OF GRAPHS

PREFACE

Originally, I became interested in serials when I did a project for the Serials Department of the Harold B. Lee Library for a library practicum. Very little is taught about serials control in library school. Serials problems are difficult but answers are possible.

My appreciation goes first to H. Kirk Memmott, Head of the Serials Department of the Harold B. Lee Library. This project would not have been possible without his help and his belief that answers do exist to serials problems. I would also like to thank Keith H. Stirling, Assistant Professor of Library and Information Science at Brigham Young University, whose many questions, suggestions, and ideas on this and other subjects were invaluable to me. I would also like to thank Victor W. Purdy and Nathan M. Smith for their reviews and criticisms of the paper as the teachers of L.I.S. 696 and L.I.S. 697. Kathleen Done of the serials Department of the Harold B. Lee Library and Deana L. Astle, Head of the Serials Department of the Willard J. Marriott Library, deserve thanks for their cooperation and help on the data gathering. The administrations of the Harold B. Lee Library and the Willard J. Marriott Library were very kind in arranging for and allowing me to use the

Kardex files of the Willard J. Marriott Library. Finally, thanks go to my fellow students who reviewed the paper, my typist, and others for their help and inspiration.

CHAPTER I

THE PROBLEM

Introduction

The Harold B. Lee Library at Brigham Young University (hereafter BYU) has decided to automate its serials claiming operation. Automation of serials control functions is probably the most difficult problem facing libraries trying to mechanize their operations. Automation programs are multiplying rapidly in other areas of library control, but a LARC Survey in 1970 showed that out of three hundred libraries running various automated systems, only one had an automated claiming system for serials.[1]

Most of the operations involved in a manual serials control system are mechanical and effective automation of these operations could improve library service. This could be especially beneficial to large libraries such as the library at BYU. Some large libraries have reported that three-fourths of their accessions are serials.[2] The

---

[1]Barbara Evans Markuson, ed., Application of Automation in American Libraries; An Analysis of the LARC Survey Returns, The LARC Reports, vol. 3, no. 1 (Tempe, Arizona: The LARC Association, Inc., 1970), p. 2.

[2]International Business Machines Corporation, Library Automation--Computerized Serials Control, Library Automation Series (White Plains, N.Y.: International Business Machines Corporation, 1971), p. 1.

complexity of the serials control problem is due to the complexity of the serials themselves. Some of the causes are:

1. The unpredictability of publication of some serials
2. Bibliographical changes in the serials
3. Shipping delays, damage, and loss
4. Publishers' claiming deadlines
5. The shortness of the 'in-print' life of each issue of a serial publication

To operate both efficiently and effectively, a serials department desires to minimize costs and duplicated effort (e.g., claiming items not yet published) and maximize service to the library patrons. Gaps in the serial holdings greatly decrease the service value of the serials in the library. Samuel Lazerow, Head of the Serials Section of the Library of Congress, has said:

> If all serials arrived on schedule there would be no need to develop and maintain a follow-up and claiming activity. But because of the many reasons for non-receipt of current serials, there must be some system for insuring the receipt of all issues due and available before they become out-of-print.
>
> My personal philosophy of the serials record function has always included the belief that one of the essential purposes of a serials record is to facilitate the claiming of missing issues. Indeed the necessity for claiming can be viewed as a major reason for establishing and maintaining a serial record.[1]

---

[1] Samuel Lazerow, "Serial Records: A Mechanism for Control," in Serial Publications in Large Libraries, ed. Walter C. Allen, Allerton Park Institute, no. 16 (Urbana, Illinois: University of Illinois, Graduate School of Library Science, 1970), p. 113.

Application of data processing to serials control, and in particular to claiming, promises great rewards to the libraries who are able to design successful systems. The automated check-in system at the University of California at San Diego reduced the required time spent on each item by 50 percent.[1] A study of the on-line claiming operation at the Biomedical Library of the University of California at Los Angeles estimated an improvement over the previous manual system of thirty additional receipts per week. Before the advent of the automated system, the claiming function was so disorganized that very few claim letters resulted in the receipt of items. After automating the serials approximately 50 percent of the claims produced and sent resulted in the receipt of a journal which would not have arrived otherwise.[2]

Serials claiming can be divided into three areas of study. First, the organization of the serials section must be examined and redesigned to fit the needs of the automated system. Second, the costs of operation also require study--both past operational costs and future predicted costs. And third, the method of determining the arrival date, central to the operation of claiming, must be investigated.

---

[1]Don L. Bosseau, "The University of California at San Diego Serials System--Revisited," Program 4 (January 1970): 9.

[2]James Fayollat, "On-Line Serials Control System in a Large Biomedical Library, Part III: Comparison of On-Line and Batch Operations and Cost Analysis," Journal of the American Society for Information Science 24 (March-April 1973): 84.

Definitions

Algorithm. An algorithm is a rule for determining the date past which an issue of a serial publication will be claimed.

Arrival pattern. An arrival pattern is a plot of the arrival dates of issues of a serial publication.

Claim. A claim is a determination that a particular issue of a serial publication which has not been received by the library is past due and a claim letter should be sent to the publisher or distributor.

Flagging. Flagging is a device used in computer files (magnetic tape, disc, etc.) to bring the attention of the computer to a certain feature in the file.

Lag time. Lag time is a period of time added to the claiming algorithm of a serial to account for normal delays in receipt of an issue (e.g., publication delays, mailing time, etc.).

Hole. A hole is a missing issue with no indication on the kardex card that a claim was made, or with the incication that the issue is out-of-print, has not yet been printed, or never was printed.

Arrival period. The arrival period is the period of time within which an issue is designated on the kardex card as expected to arrive (e.g., the arrival period of the first

issue of a quarterly for 1975 is January 1, 1975 to March 31, 1975). In most cases this may be equated to the frequency of publication of the serial.

Automatic claiming. Recommendation to claim an issue of a serial made by the computer through the application of some claiming algorithm programmed for the use of the computer.

The Problem Statement and Delimitations

This study involved describing and comparing three algorithms for claiming serials based on predictions of the serials' arrival dates. All of the major functions of a serials department depend on when its serials issues arrive. In the analysis of the data, the arrival patterns were assumed to be normally distributed. The merits of a batch system versus an on-line system were excluded.

CHAPTER II

LITERATURE SEARCH

To claim a serial issue one must have some idea of when it should arrive at the library in order to know when it can be considered late. Prediction of the arrival date of a serial is necessary to all claiming systems, whether automated or manual.[1] The crudeness of the prediction of arrival dates has been one of the basic problems of manual claiming systems. The claiming procedures of manual systems have at best been erratic and at worst chaotic. However, a review of the literature revealed no one predicting the development of a completely automated claiming system.

There are numerous ways to predict the arrival date of a serial issue. However, three general concepts were identified in the literature:

1. Prediction of arrival by publication date plus a lag time.

2. Prediction of arrival using the history of arrival dates and a statistical method to extrapolate from

---

[1] Five Associated University Libraries, Joint Serials Control System Project for the Libraries of Cornell University, University of Rochester and the State University of New York at Buffalo, Phase I Feasibility Study Final Report, (Bethesda, Md.: ERIC Document Reproduction Service, ED 051 827, 1971), p. 13.

the past history of arrival dates

3. A combination of 1 and 2

A fourth concept also appeared viable to the author: use no prediction of arrival--merely set a claiming date (e.g., claim just in time to catch the publisher's deadline for claiming).

The first concept based the arrival frequency on the publisher's stated arrival date--most often a date found printed on the publication itself. This was the most common practice in libraries with automated serials sections. Various lag factors were added to compensate for shipping and publication delays. Lag factors may be individualized for particular serials titles.[1] This means of predicting the arrival date depended on the librarian's knowledge of the particular problems with each serial title.

The keys to this concept were the prediction of arrival information (date, issue, number, etc.) and the librarian's knowledge of publication and shipping patterns. In a study of the UCLA Biomedical Library, Fayollat and Luck[2] estimated that all arrival information can be predicted for 70 percent of their serials titles. It was also estimated that arrival information can be predicted for 80 percent of

---

[1]Martin D. Fried and Ruth Dunham, California State Library: Processing Center Design and Specifications, vol. 4: Serials Control System (Berkeley, Calif.: Institute of Library Research, University of California, 1969), p. 116.

[2]James Fayollat and Don Luck, "Computer Based Serials Control System, Biomedical Library, UCLA," American Documentation 20 (October 1969): 385.

the serial titles at the University of California at San Diego Library.[1]

The second concept[2] was based on the simple statistical principles of mean, standard deviation and confidence interval. The theory behind this concept was that the issues of a serial arrive at a library in a normally distributed pattern. Therefore, a confidence interval could be computed around the mean of the arrival dates or arrival intervals of the issues of a serial so that the serials librarian could expect to have a certain percentage of confidence (single-sided normal distribution confidence level) that an issue of that serial would arrive within a certain period of time (97.7 percent confidence level at two standard deviations and 99.9 percent confidence level at three standard deviations). Based on these premises the librarian could claim an issue of a serial with a theoretically reasonable amount of accuracy. No such system was found to be in actual operation.

---

[1] Don L. Bosseau, "Case Study of the Computer Assisted Serials System at the University of California, San Diego," in Proceedings of the LARC Institute on Automated Serials Systems, ed. H. William Axford (Tempe, Arizona: The LARC Association, 1973), p. 108.

[2] Interview with H. Kirk Memmott, Harold B. Lee Library, Provo, Utah, October 1974; and Robert W. Burns, Jr., The Design and Testing of a Computerized Method of Handling Library Periodicals (Title III) (Bethesda, Md.: ERIC Document Reproduction Service, ED 050 753, 1970).

One possible combination of the first and second concepts was the Ohio College Library Center System (hereafter O.C.L.C.).[1] It was based on the arrival date of the last issue of a serial received at the library. The frequency of publication and a lag factor are added to the old arrival date to compute the new expected arrival date. The Ohio College Library Center proposed to use this system in their forthcoming automated serials control system, but it had not yet reached the final design stage. No claiming system similar to the proposed O.C.L.C. System was known to be in operation.

No precedent was found for the fourth concept in the literature (see page 15 for a fuller explanation of the concept).

Systems for each of these four concepts are developed into models in chapter 3 and will be called (1) the Jordan-Larsen System,[2] (2) the Memmott System,[3] (3) the Ohio College Library Center System,[4] and (4) the Deadline

---

[1]Meg Sarver, Ohio College Library Center, to John O. Christensen, 28 January 1975, Personal Files of John O. Christensen, Provo, Utah.

[2]So called because K. Paul Jordan and A. Dean Larsen of the Acquisitions Department of the Harold B. Lee Library have proposed such a system for the serials section.

[3]This system has been proposed by H. Kirk Memmott, Head of the serials section of the Harold B. Lee Library.

[4]Sarver, Letter to author.

System.[1] All four systems can compute a date on which the expected issue will be claimed.

There are some serials for which little hope exists for their control under a claiming algorithm. They are very irregular in their publication pattern. The same title may vary from many issues during one year to many years for one issue.[2] These cannot, at present, be claimed and necessitate a stop-claim flag in the serial file.

It is appropriate here to clarify the use of the term "automatic claiming." The four claiming algorithms described above produce claiming dates, however, a human interface is still necessary. The claiming dates or lists of items to be claimed are the recommendations of the computer to the librarian. The librarian makes the final decision--to claim or not to claim.

Versions of the first concept were the only ones known to be in operation. The other three concepts are only proposals. One of the purposes of this study was to compare these proposals to see if they would be viable solutions to the problem of serials claiming.

---

[1] Referring to the publisher's deadline for claiming.

[2] David Bishop, Arnold L. Milner, and Fred W. Roper, "Publication Patterns of Scientific Serials," American Documentation 16 (April 1965): 118.

CHAPTER III

DATA COLLECTION

Approach to the Problem

The four claiming algorithms proposed in chapter 2 were compared by using actual records of arrival dates. Three other procedures were considered: (1) analysis with contrived arrival dates, (2) variation of several parameters (e.g., mean, standard deviation, etc.) considered in the four algorithms as plotted against the performance of the algorithms, and (3) analysis with a presently running serials system to check the performance of the algorithms with current data. The first method was rejected as being synthetic--the human design of the dates would throw bias into the results. The second method was rejected as not being related closely enough to the working problem to permit easy application to the specific situations (at hand). The third method, although possibly better than the one used, was rejected because of lack of time and money.

Survey Plan

The four concepts described in chapter 2 were developed into specific claiming algorithms. Arrival dates have not been kept in the Kardex files at the Harold B. Lee

Library, therefore a random sample of the serials at the Willard J. Marriott Library at the University of Utah was taken. These dates were used to test the performance of the claiming algorithms. Sixteen arrival patterns of serials were chosen as representatives of the random sample and were graphed to check what kind of patterns they follow and how they would be claimed under the different algorithms. Conclusions were then made from general observations on the random sample and specific observations on the sixteen chosen for checking the algorithms.

Preparation of the Instrument

The four concepts described in chapter 2 were developed into claiming algorithms and explanations of each follow.

Jordan-Larsen system

When an issue is checked in with the Jordan-Larsen System (see appendix C for flowcharts), if the issue is the expected one, then the new claim date becomes the end of the publication period for the next issue plus a lag factor (or grace period). Some of the suggested lag factors are: (a) one week for weekly publications, (b) one month for monthly publications, (c) three months for quarterly publications, and (d) six months for annuals. There are many possible lag factors.

Several problems are immediately noticed. The accuracy of the claiming is dependent on the librarian's

knowledge of individual titles. For large libraries with thousands of current subscriptions past experience has proved that this method is inefficient.[1] Also, several issues of a serial title would have to arrive before a proper lag factor could be determined. New personnel would require many months, if not years, to become competent in determining the lag factors and making periodic adjustments to the algorithm.

Memmott system

The Memmott System (see appendix C for flowcharts) uses the standard deviation and mean of past arrival patterns on the premise that serial arrivals act like a normal distribution. It appears as if it should work best for serials with periodic intervals of arrival, but several difficulties prevent it from being comprehensive. First, to compute a standard deviation, at least three data points (arrival dates) are needed which necessitates an initial waiting period until three issues of the serial have arrived. Second, if the publication pattern of a serial is not periodic (e.g., one example of an irregular pattern would be a publication which is published monthly--eleven months of the year), the standard deviation is biased and loses much of its value of prediction. With the aid of a publication

[1] Robert W. Burns, Jr., The Design and Testing of a Computerized Method of Handling Library Periodicals (Title III) (Bethesda, Md.: ERIC Document Reproduction Service, ED 050 753, 1970), pp. 3-6.

pattern recorded in the serial file, the intervals of equal length between consecutive issues are used to determine a standard deviation. Thus, in the example of the serial published eleven months of the year, the ten one-month intervals are used to compute the standard deviation and the one two-month interval is thrown out.[1]

There is a further problem. When an issue is claimed, the date of arrival of the claimed issue must be discarded because it will not reflect the true arrival pattern. The discarded data points will be larger than those used to compute the standard deviation, causing the standard deviation to be low. Because of this problem, further refinements are necessary to avoid claiming an item too soon. The more data points (arrival dates) available, the finer tuned the standard deviation becomes. Some unpredictable occurrences will have to be dealt with (e.g., labor strikes, financial problems, natural calamities, etc.) and will continue to require a human interface in claiming with the Memmott System as well as any other automated system as indicated in the literature search.

O.C.L.C. system.

The O.C.L.C. algorithm (flowcharts would be very similar to those used for the Jordan-Larsen System which are found in appendix C) is based on the last arrival date of an issue of a serial. The numerical sequence of issues (if one

[1]Fried and Dunham, Serials Control System, pp. 113-19.

has been determined) is conserved. When an issue of a serial arrives, the new claiming date is computed from the arrival date by adding the frequency of publication plus a lag factor to the arrival date to compute a new expected arrival date for the next issue of the serial title.

Deadline System[1]

A fourth method for claiming uses only the publisher's deadline for claiming. In conjunction with the expected issue information, the known publisher's deadline is used to allow as much time as possible for the publication to arrive without defaulting on the deadline set by some publishers for claiming issues of their serials. Due to the lack of information of the deadlines of many publishers, this algorithm can be used only selectively.

Notation

The notation below is used hereafter in this paper:

d_i = the day of the month (week, period, etc.) on which the issue arrives

T_i = today's date (arrival date)

I_i = the number of days in the i^{th} interval between arrival dates of a serial $I_i = T_i - T_{i-1}$

n = the number of intervals used to compute the standard deviation of the arrival pattern (in number of days)

[1]This is the author's own idea.

$$I = \text{Mean} = \frac{\sum_{i=1}^{n} I_i}{n} \quad \text{or} \quad \bar{d} = \frac{\sum_{i=1}^{n} d_i}{n}$$

$$S_i = \text{Standard Deviation} = \left[\frac{\sum_{i=1}^{n} (I_i - \bar{I})^2}{n-1}\right]^{\frac{1}{2}}$$

$$= \left[\frac{\sum_{i=1}^{n} (d_i - \bar{d})^2}{n-1}\right]^{\frac{1}{2}}$$

C_i = claim date computed after i^{th} interval =

97.7% confidence[1]	or	99.9% confidence[2]
$T_i + \bar{I} + 2S_i$		$T_i + \bar{I} + 3S_i$

Collection Techniques

Arrangements were made by H. Kirk Memmott with Deana Astle, Head of the Serials Department at the Willard J. Marriott Library, to microfilm a random sample of the kardex cards of the University of Utah. The Kardex file of the University of Utah contains approximately 18,000 entries including some newspapers and 'see' references. An Olympus half-frame 35 mm camera on a tripod with Kodak Tri-X film was used to photograph the sample.

All of the titles were considered currently received items. The number eight was randomly picked to begin the

---

[1]This represents a single-sided normal distribution confidence level. Owen L. Davies and Peter L. Goldsmith, Statistical Methods in Research and Production (New York: Hafner, 1972), p. 434.

[2]Ibid., p. 435.

sampling, and thereafter a systematic sample was taken of every twentieth item. The newspapers and 'see' references were excluded. In this manner kardex cards for 758 titles were photographed.

The developed Tri-X film was spliced to form one reel of film in alphabetical order for easier analysis. A record of each title, its frequency of publication, years of receipt, claims and holes, along with certain descriptive comments was made as an index to the roll of film. Of the 758 records, 358 were found to be too irregular in their arrival patterns or to have records of too few arrival dates (usually less than six) to be useful for analysis. Thirty-five titles were noted by the author as appearing to be workable examples for analysis. Many others could have also been chosen. Sixteen titles, representative of the thirty-five noted titles, were chosen for final analysis. This subjective means of choosing certain "examples" for analysis was introduced due to a lack of time for a comprehensive analysis of all the arrival patterns.

It was decided that the plotting of graphs of the arrival patterns versus the claiming algorithms would be the clearest means of analysis. The arrival patterns of fifteen of the sixteen serials titles were plotted on two different graphs representing two different ways of determining the arrival patterns of regular serials (to be explained in chapter 4). The ninth pattern was an irregular pattern and could be analyzed on only one such graph.

Compilation and Preparation of the Graphs

The arrival pattern of a regular (designated frequency of publication) serial title was represented two different ways: (a) the arrival dates were represented as the number of days elapsed between the beginning date of the expected arrival period of an issue (as calculated from the frequency of arrival from the kardex card) and the actual arrival date of an issue (as indicated on the kardex card), and (b) the arrival dates were represented as the number of days in the interval of time between the arrival of consecutively expected issues. Only the 'b' method above was feasible for the one irregular serial analyzed. With the arrival of each succeeding issue of each serial title, the mean ($\bar{I}$) and standard deviation (S_i) were computed (see appendix B for computer program), and from them the claiming date for the next expected issue (C_i) was computed at the 97.7 percent confidence level as the date when the next issue would be claimed.

Two graphs were made for each regular serial. The 'a' graphs contained plots of the arrival patterns of the regular serials by the 'a' method defined above, plots of C_i (Memmott algorithm), and plots of the Jordan-Larsen claiming algorithm.

In most cases, the Jordan-Larsen claiming algorithm was computed as twice the length of the arrival period (e.g., graphed as two months for a monthly), however, in the case of annuals and bi-annuals the Jordan-Larsen claiming

algorithm was computed as one and one-half times the issue's arrival period (i.e., one and one-half years and nine months, respectively). The lag periods are the choice of the author. This represents the 'frequency plus a lag factor' concept mentioned earlier.

The 'b' graphs contained plots of the arrival patterns of regular and irregular serials by the 'b' method defined above and plots of C_i. Those graphs representing the analysis of regular serials also contained plots of the O.C.L.C. method of claiming serials. The periods used for the O.C.L.C. method were the same as those used for the Jordan-Larsen method on the 'a' graphs.

CHAPTER IV

PREPARATION AND ANALYSIS OF THE DATA

The following is an explanation of the symbols on the graphs of the publication patterns and claiming algorithms:

'a' graphs

1. Numbers on the ordinate (vertical axis) represent the number of days in the interval from the beginning date of the period in which an issue is expected and the date of its actual arrival

2. Numbers on the abscissa (horizontal axis) represent the years (chronologically) in which the issues are expected with subdivisions for the periods within the years (again taken from the frequency of arrival and the information recorded on the kardex card)

3. ——— represents a plot of the serial arrival pattern

4. — · — · · represents a plot of the Memmott claiming algorithm ($T_i + \bar{I} + 2S_i$)

5. · · · · · represents a plot of the Jordan-Larsen claiming algorithm

6. X represents claims of issues which would have been made by the Memmott claiming algorithm on issues

actually missed by the University of Utah (whether eventually received by claim or not)

7. ○ represents claims of issues which would have been made by the Memmott claiming algorithm on issues actually received by the University of Utah (and apparently not claimed)

The caption with each graph gives the title of the serial and its frequency of publication.

'b' graphs

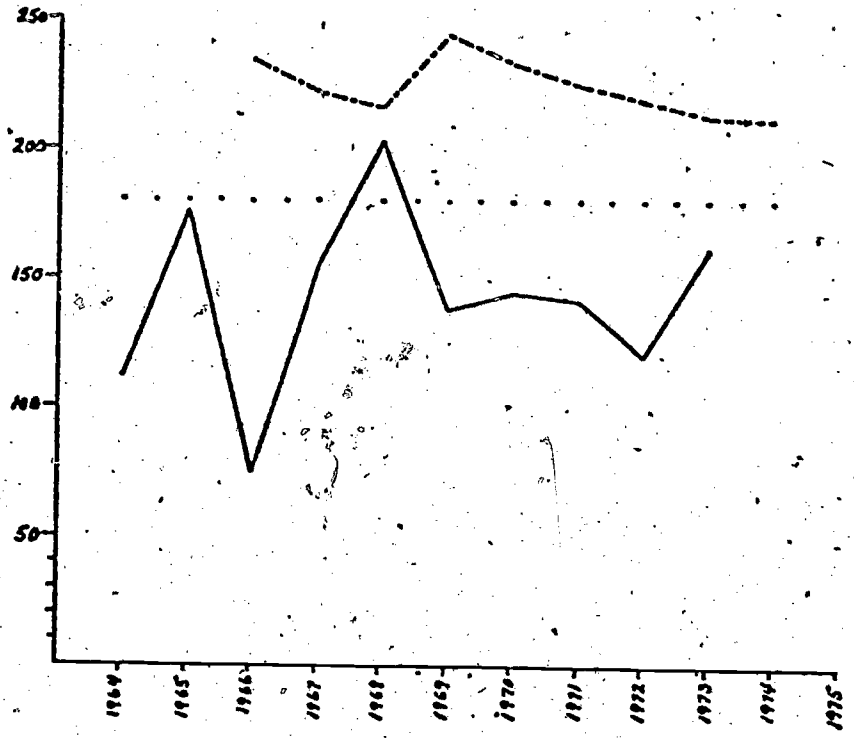
The same symbols as above apply to the 'b' graphs except:

1. ····· represents a plot of the O.C.L.C. claiming algorithm, and

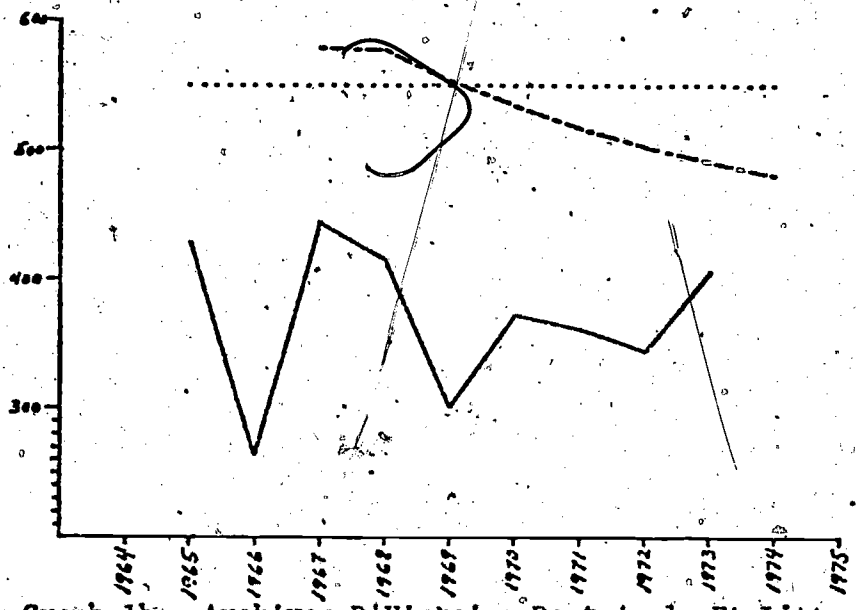
2. The O.C.L.C. algorithm was not included in graph 9 because this is the graph of an irregular serial and the O.C.L.C. claiming algorithm could not be applied to irregular serials.

Not enough data were found to evaluate the fourth method of claiming, the Deadline system. Therefore, it was left out of the analysis of the data. The graphs of the data follow in graphs 1a to 16b.

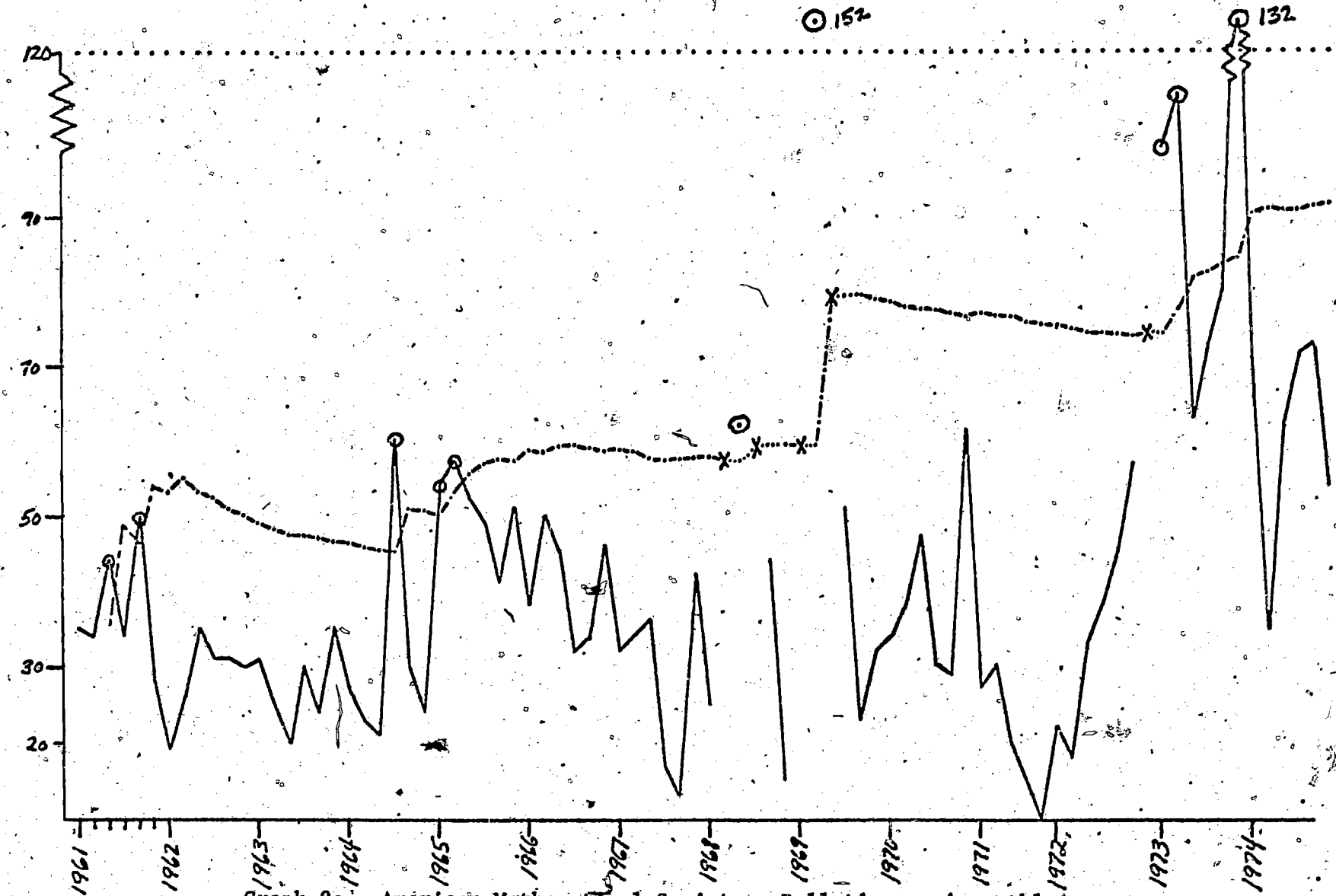
Some numerical compilations of the results of the graphs and the random sample are compiled below. Ten basic classes of arrival patterns were used by the author with an eleventh for those patterns which did not fit in any of the ten. These definitions are the result of the indicated



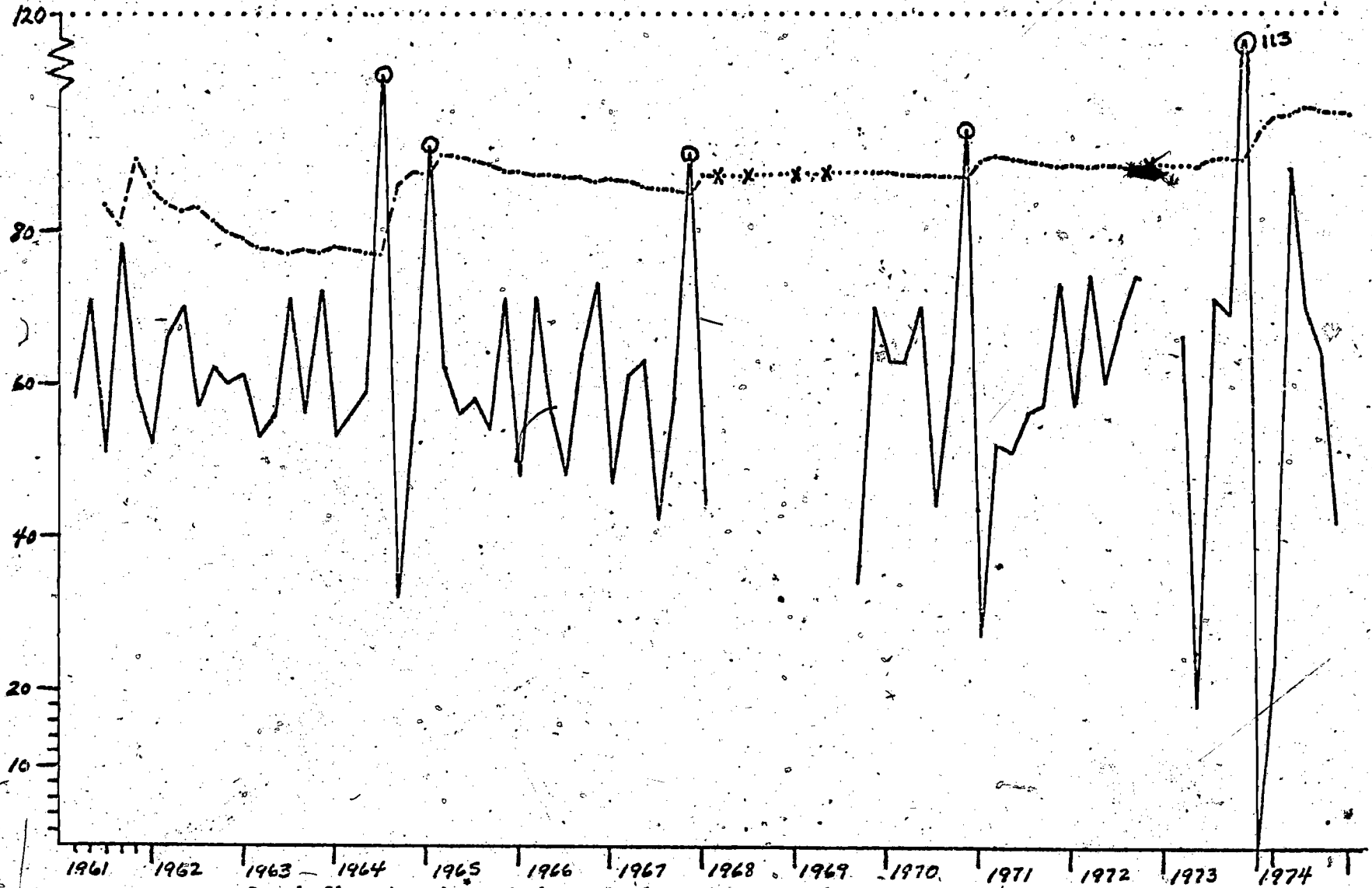
Graph 1a. Archives D'Histoire Doctrinale Et Litteraire Du Moyen Age--annual



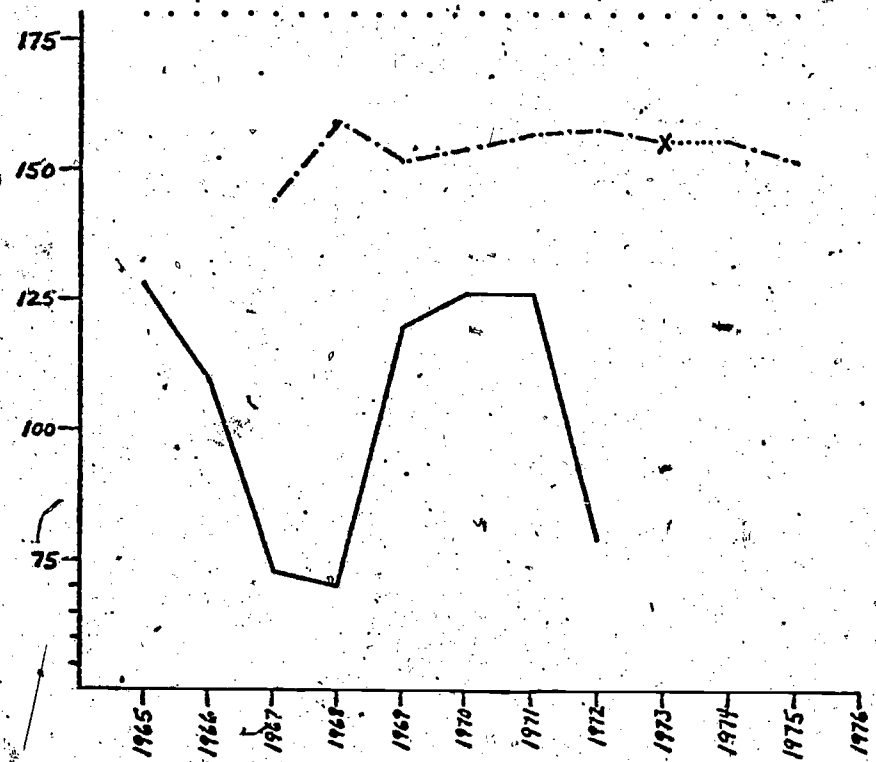
Graph 1b. Archives D'Histoire Doctrinale Et Litteraire Du Moyen Age--annual



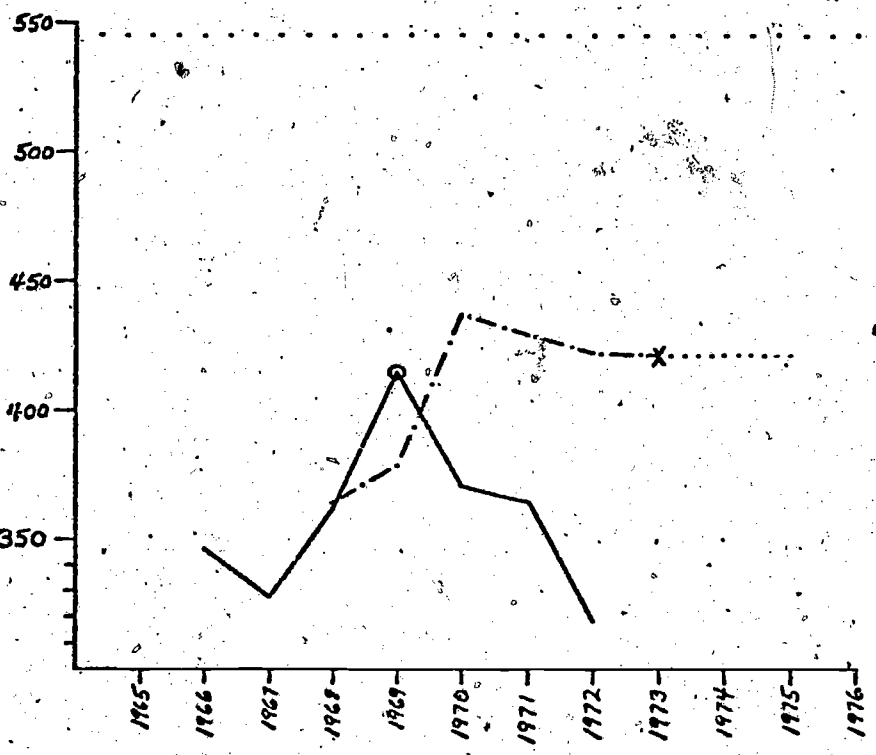
Graph 2a. American Mathematical Society. Bulletin--semi-monthly



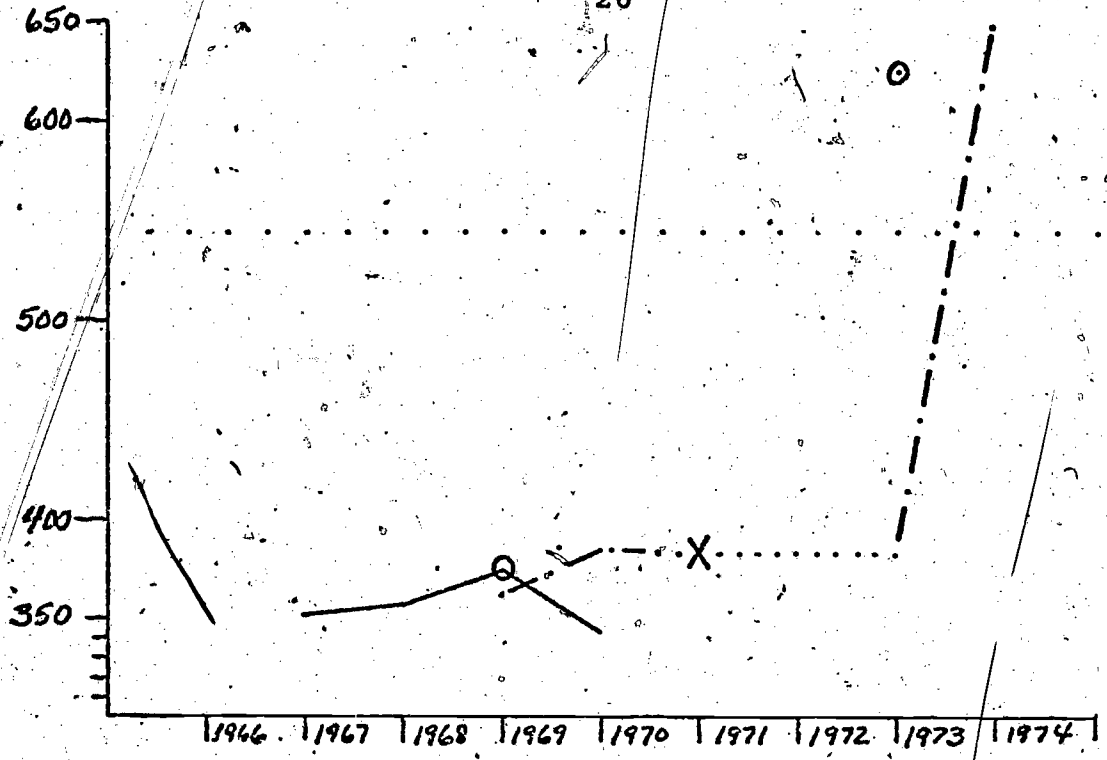
Graph 2b. American Mathematical Society. Bulletin--semi-monthly



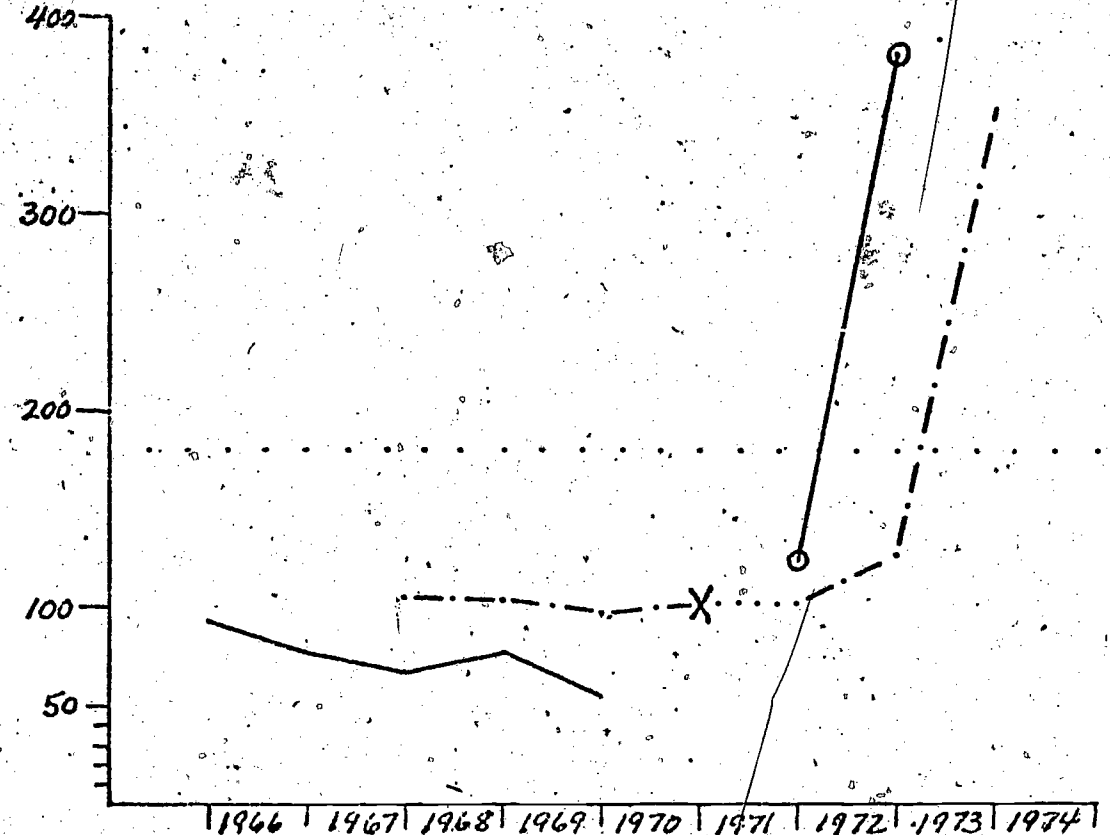
Graph 3a. Britain: An Official Handbook--annual



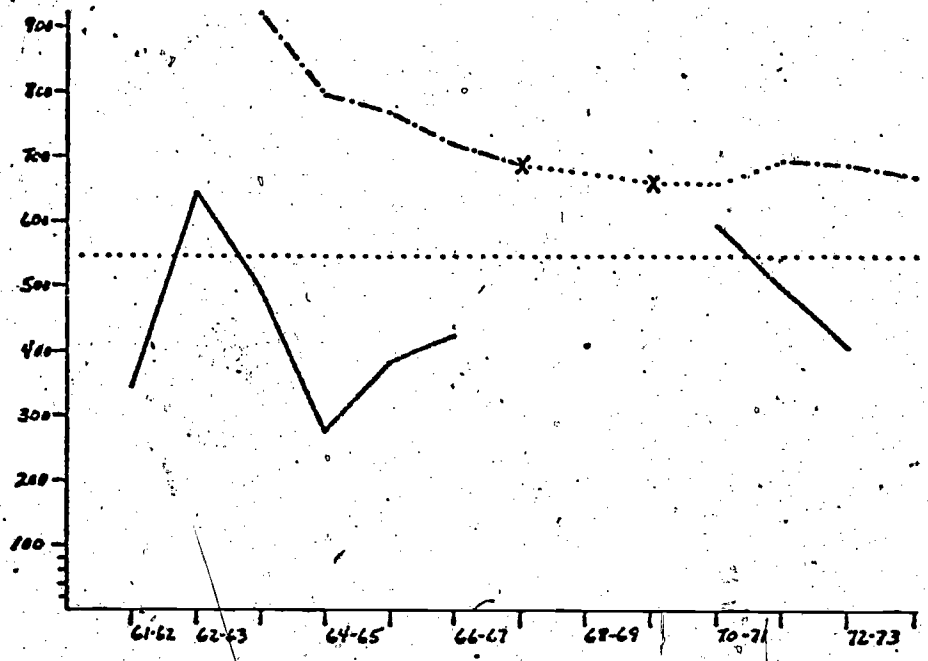
Graph 3b. Britain: An Official Handbook--annual



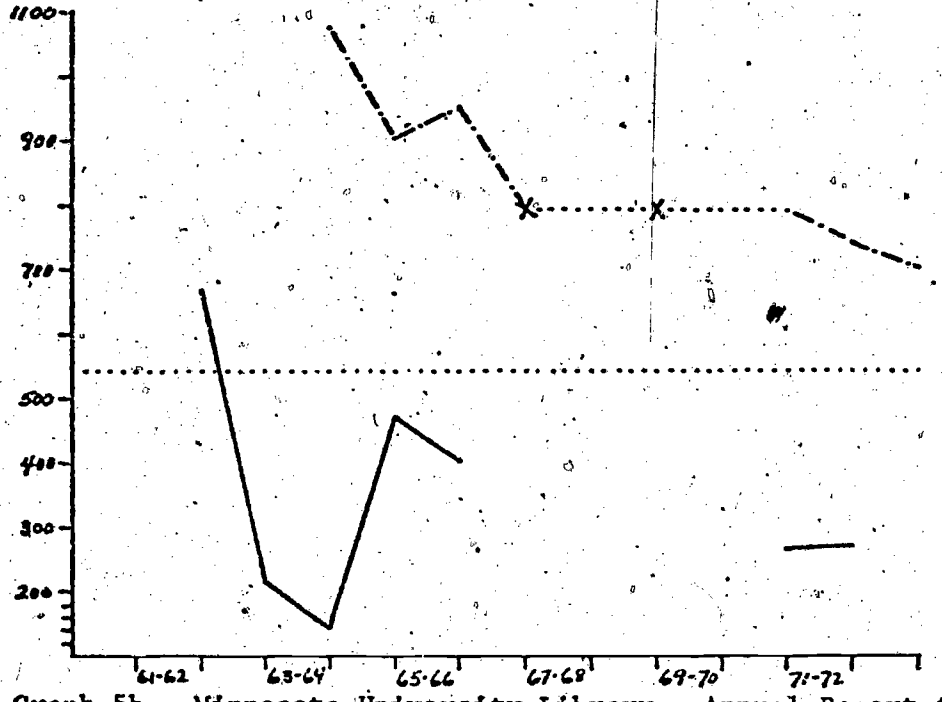
Graph 4b. Denver. Museum of Natural History. Report--annual



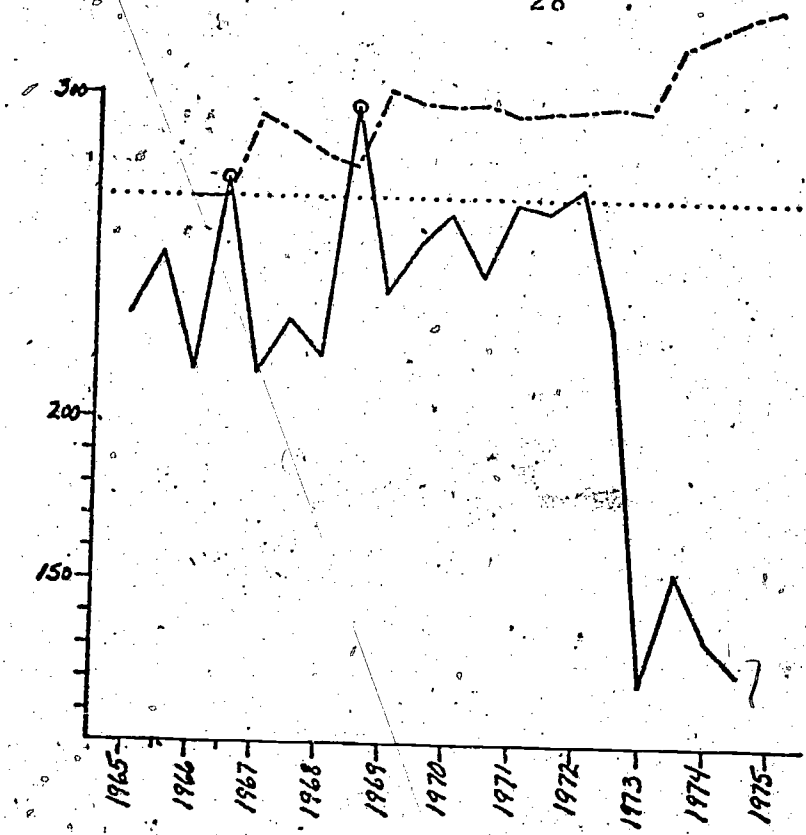
Graph 4a. Denver. Museum of Natural History. Report--annual



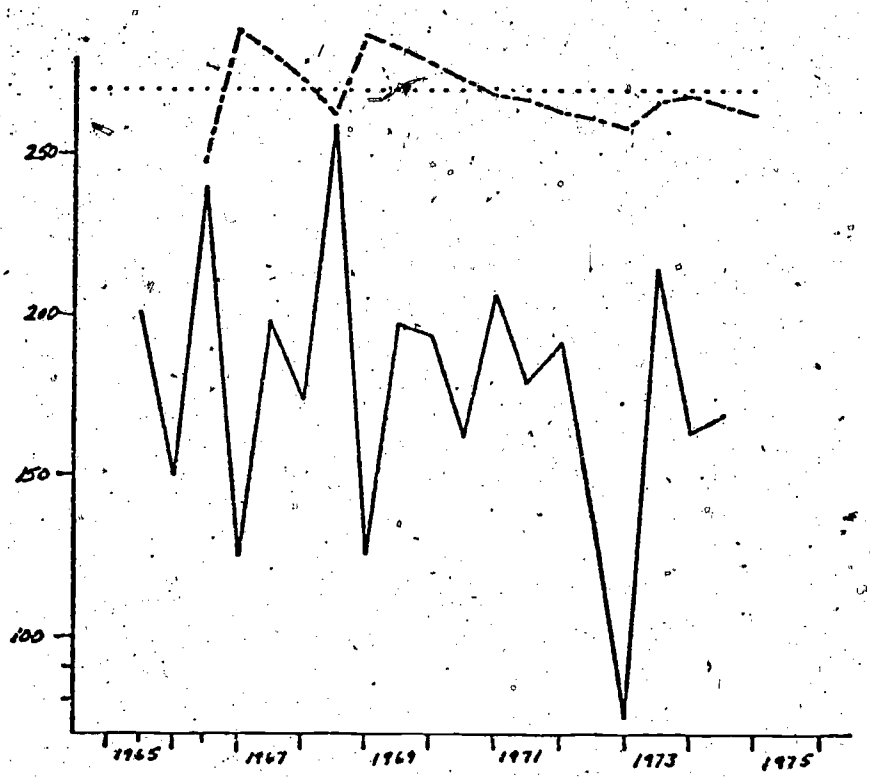
Graph 5a. Minnesota University Library. Annual Report to the President--annual



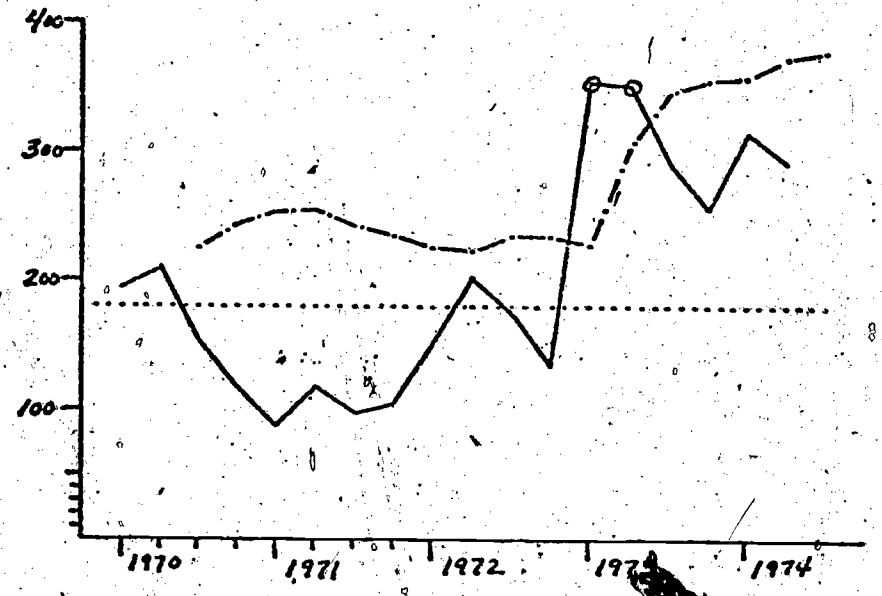
Graph 5b. Minnesota University Library. Annual Report to the President--annual



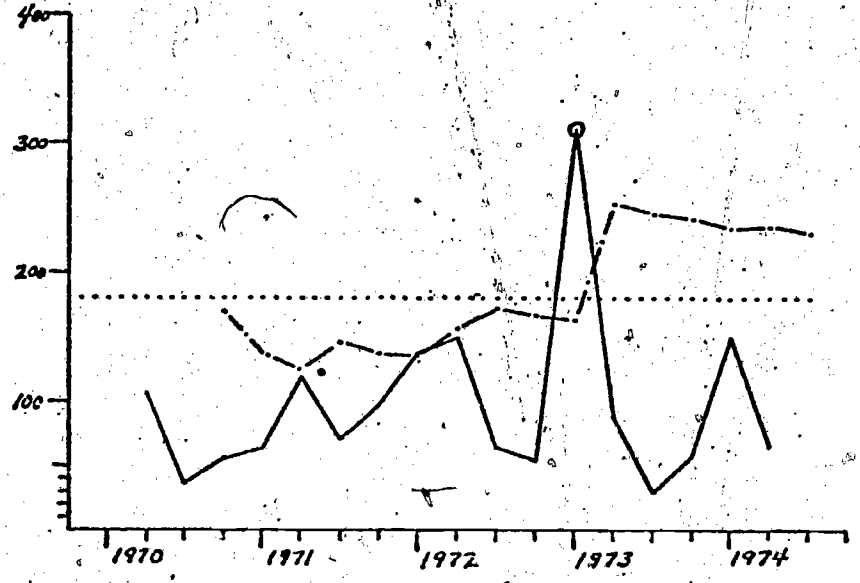
Graph 6a. Byzantinische Zeitschrift--bi-annual



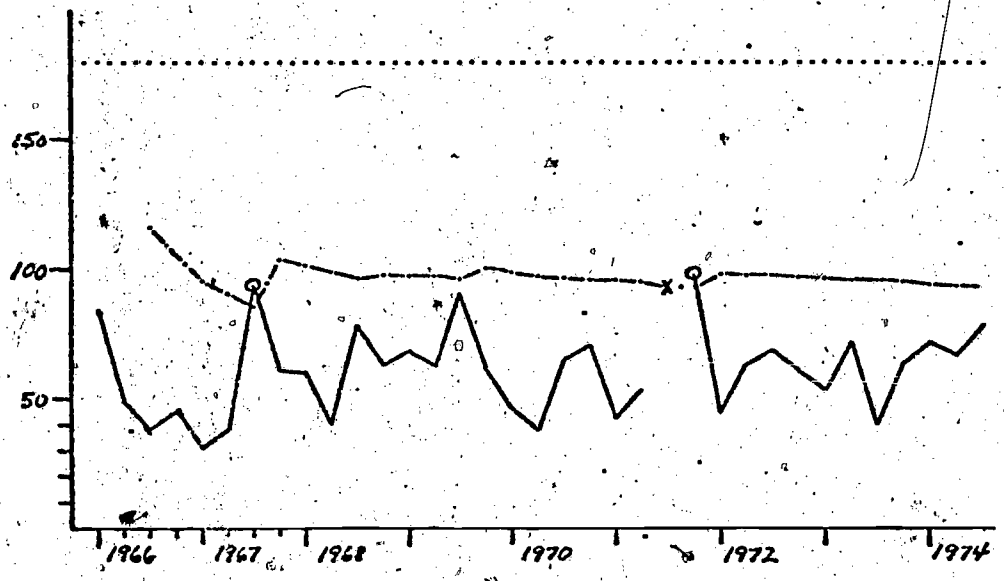
Graph 6b. Byzantinische Zeitschrift--bi-annual



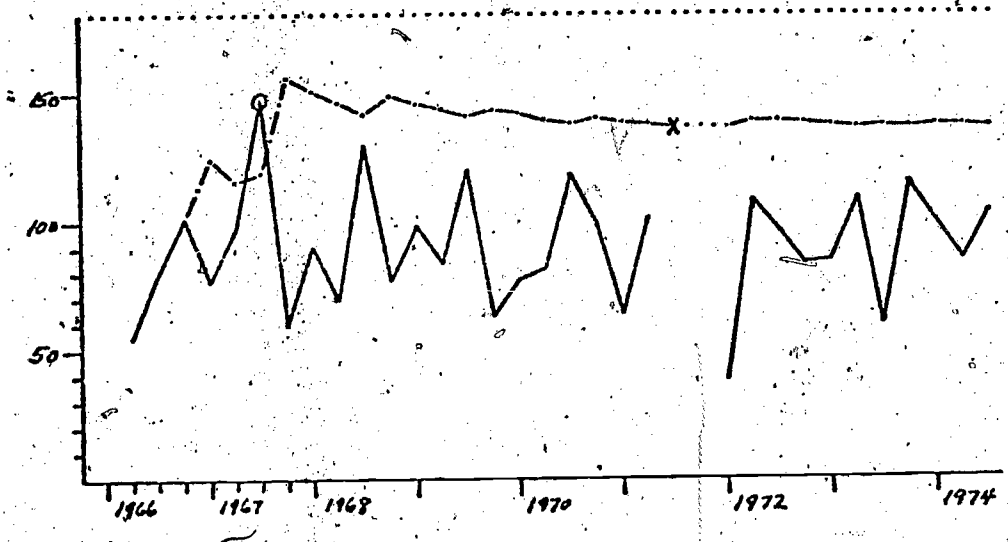
Graph 7a. Bulletin des Sciences Matematiques--quarterly



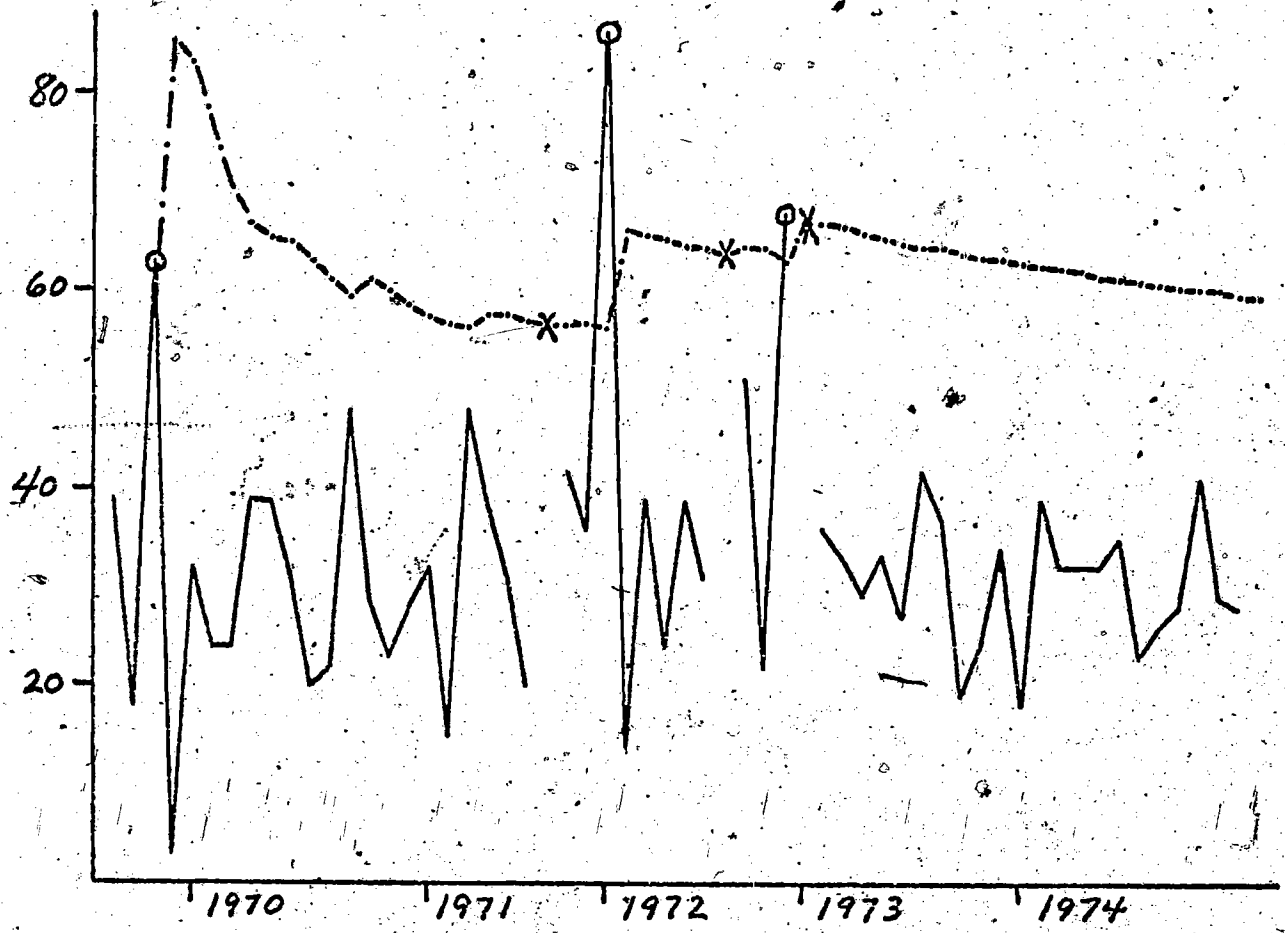
Graph 7b. Bulletin des Sciences Matematiques--quarterly



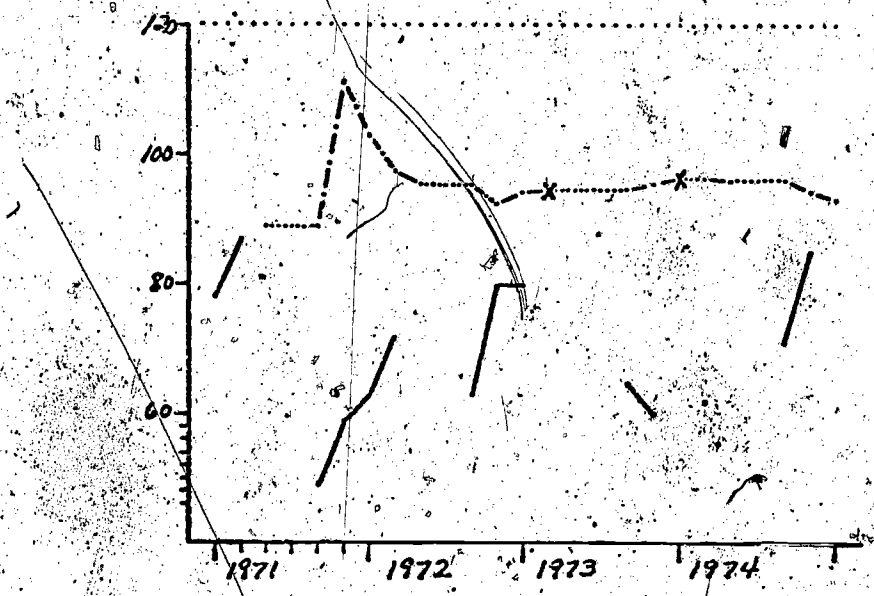
Graph 8a. Canadian Psychologist--quarterly



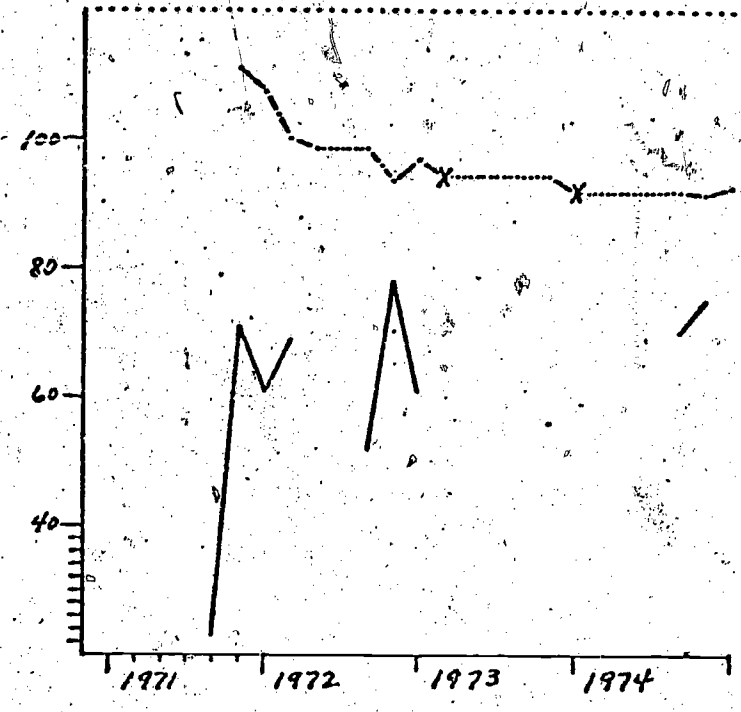
Graph 8b. Canadian Psychologist--quarterly



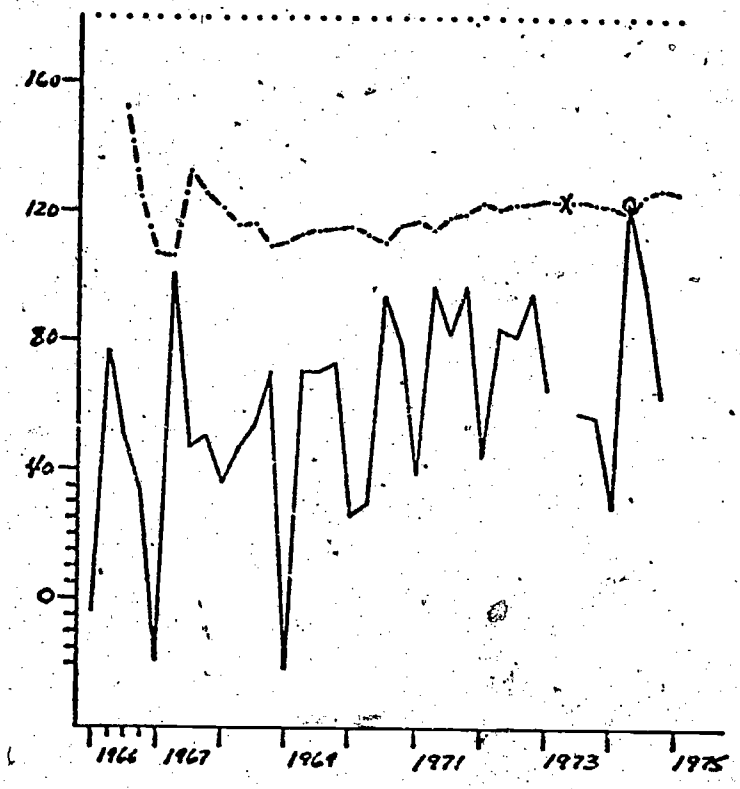
Graph 9. Chamber of Commerce Newsletter--irregular



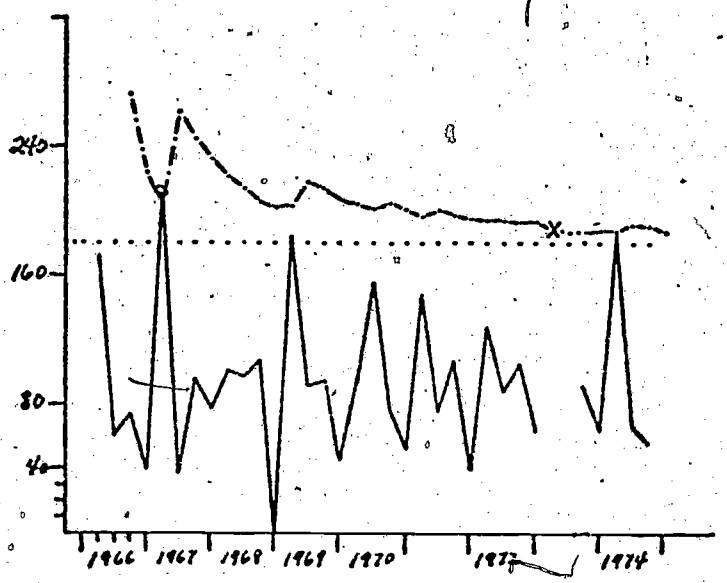
Graph 10a. College English Association Forum--semi-monthly



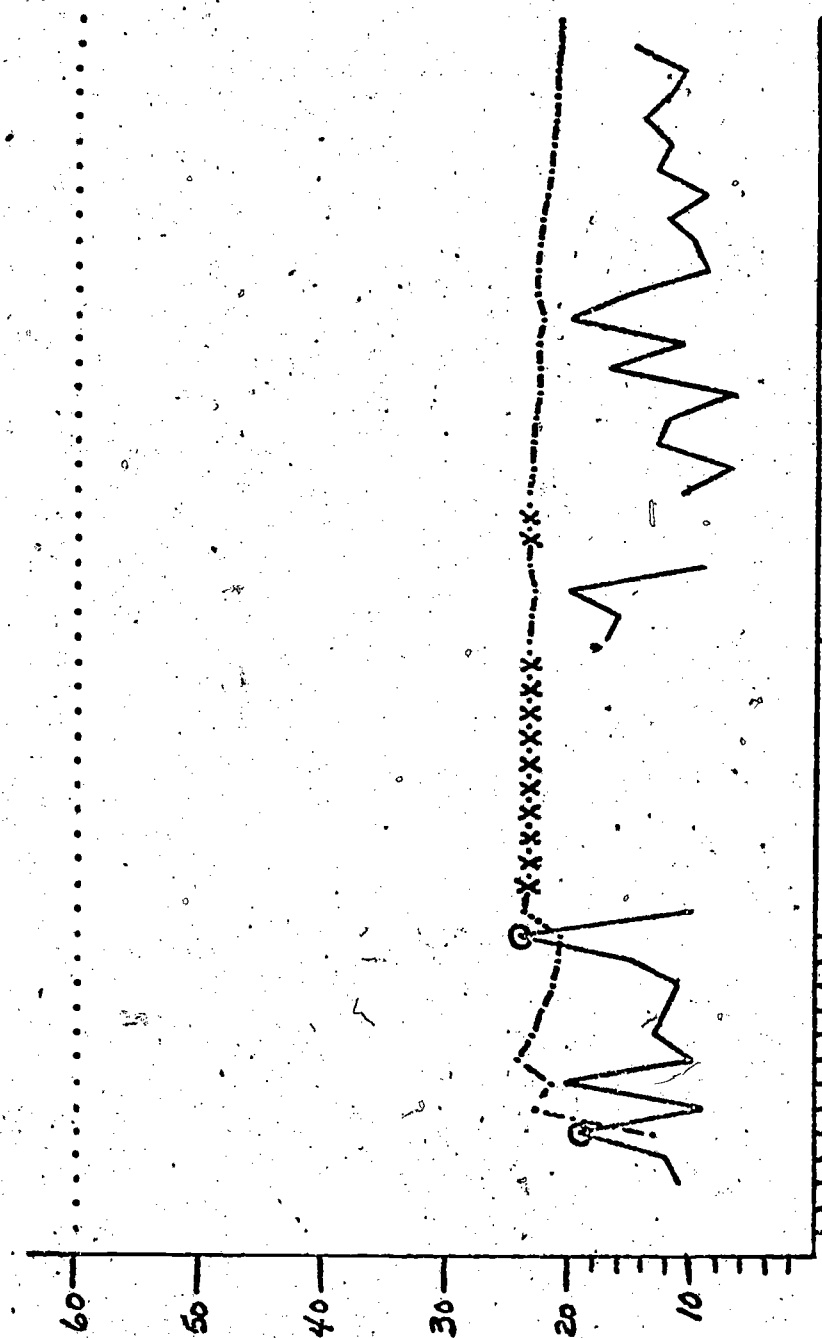
Graph 10b. College English Association Forum--semi-monthly



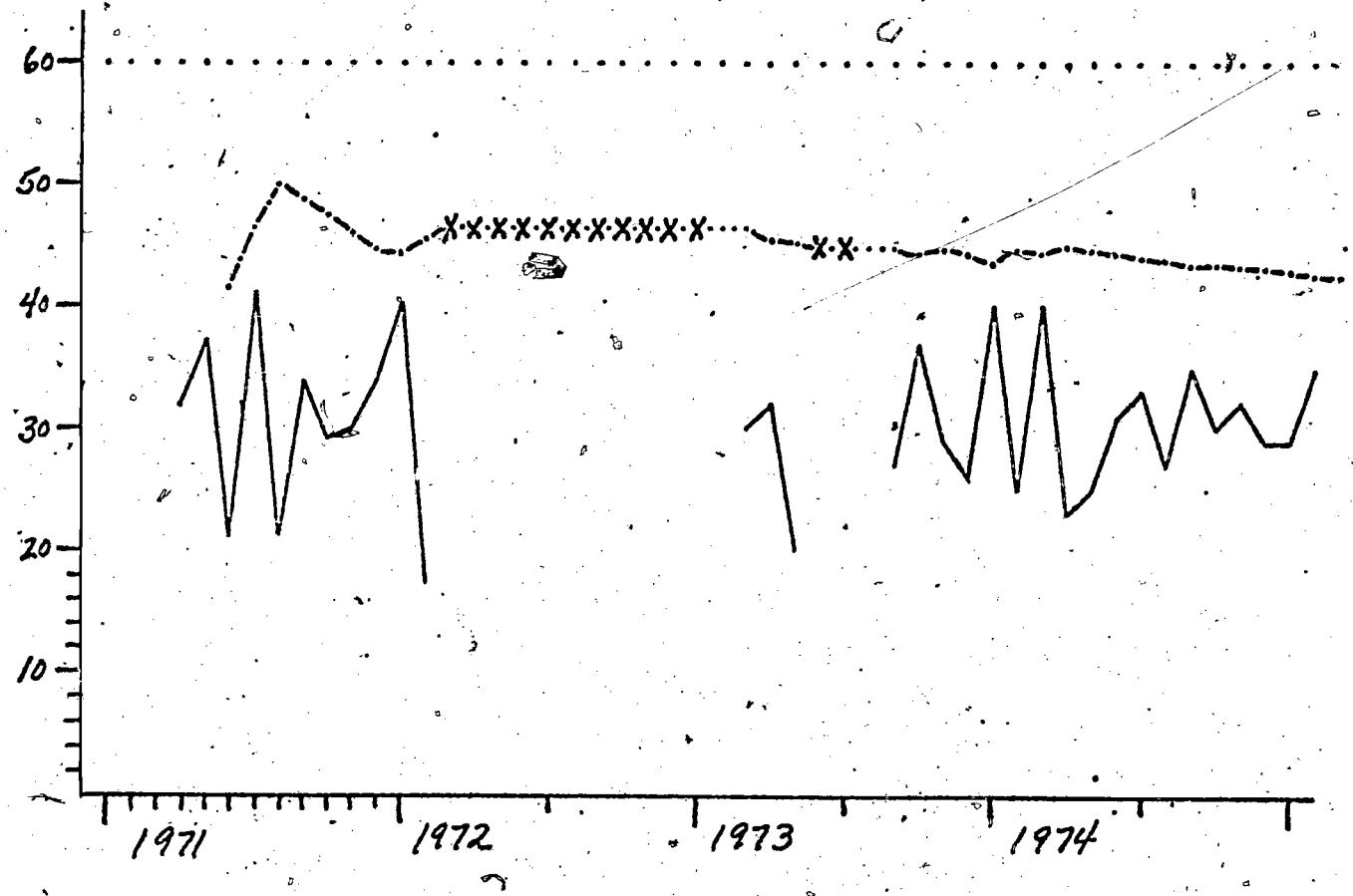
Graph 11a. Cornhill Magazine--quarterly



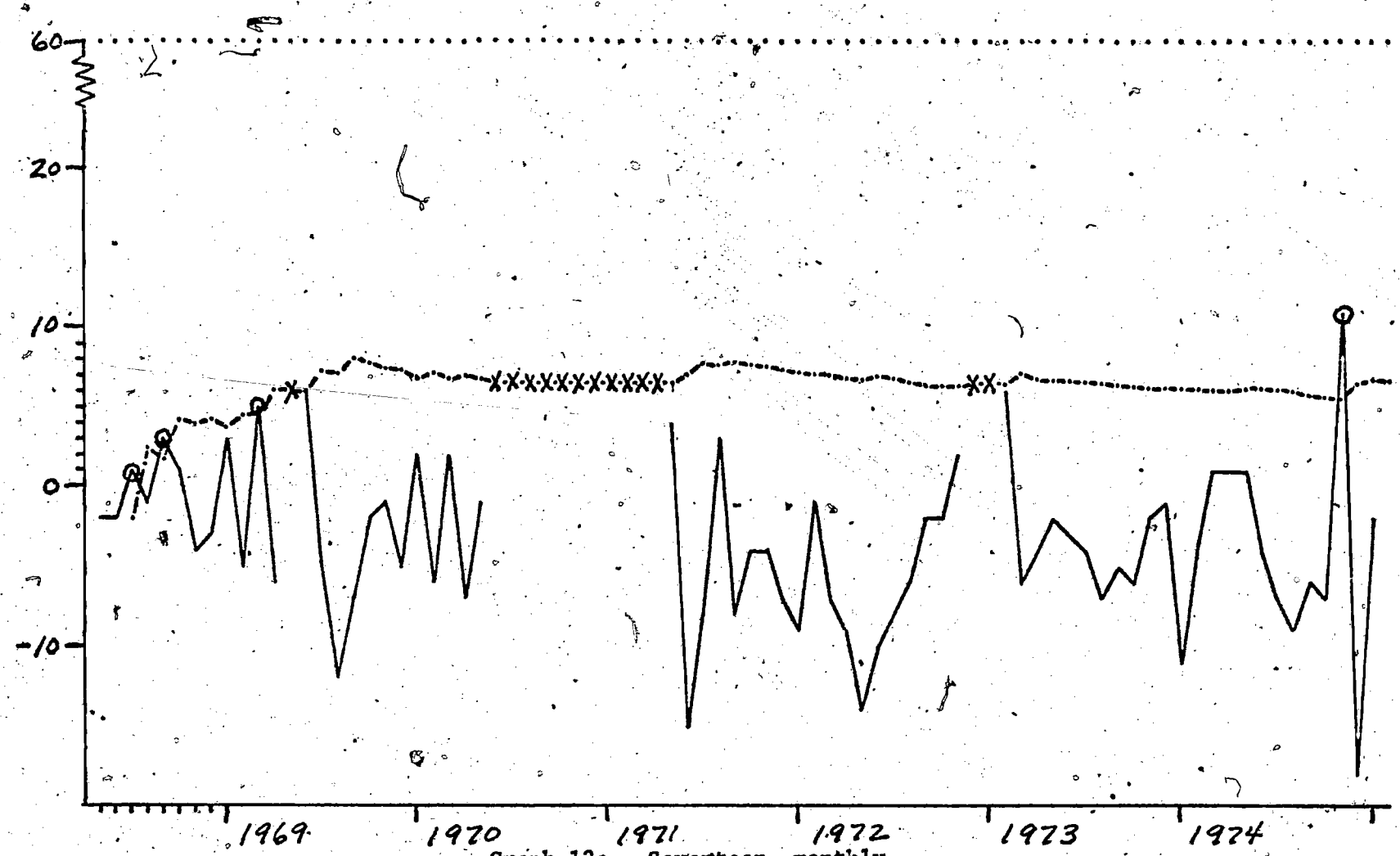
Graph 11b. Cornhill Magazine--quarterly



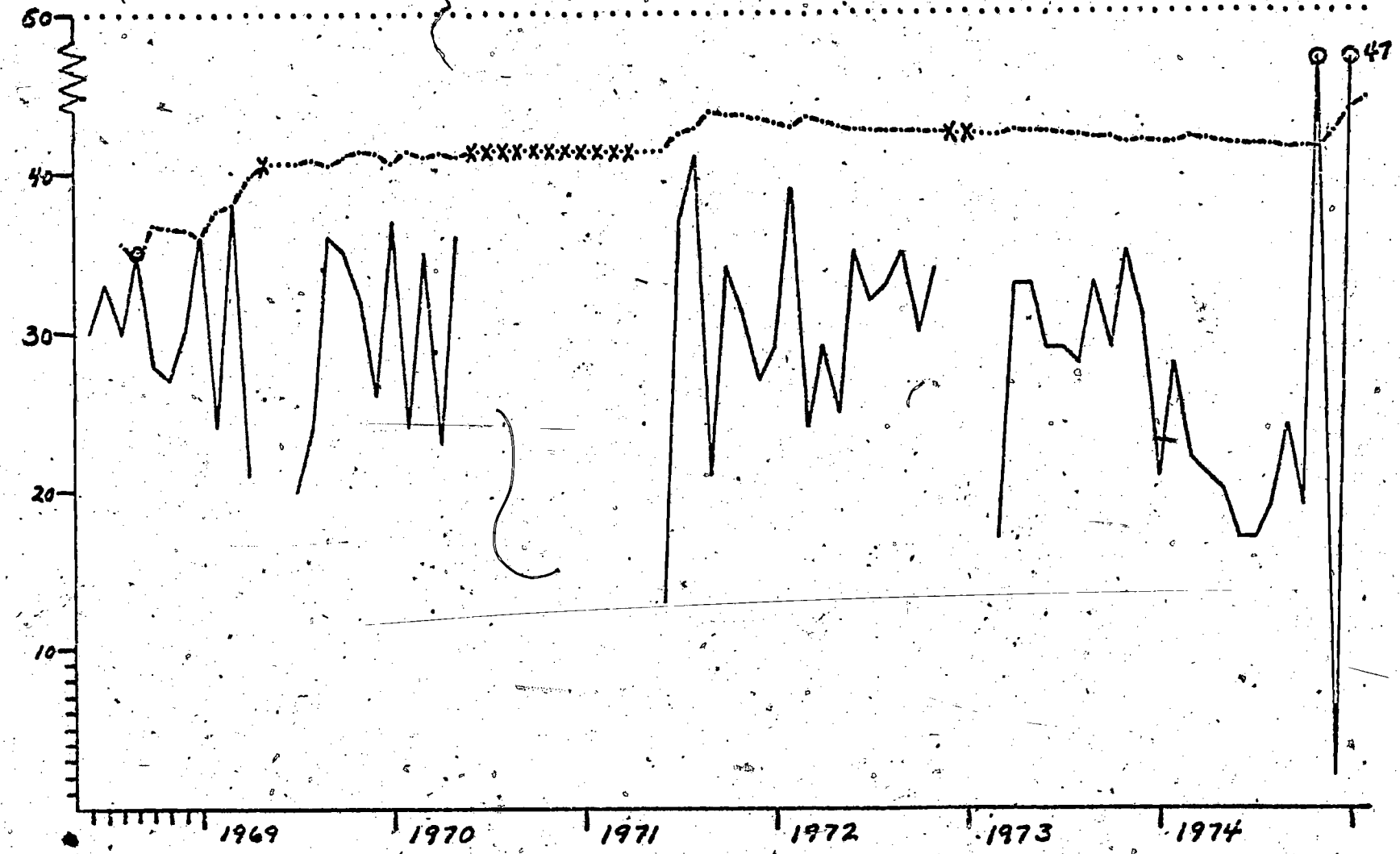
Graph 12a. Outdoor Arizona--monthly



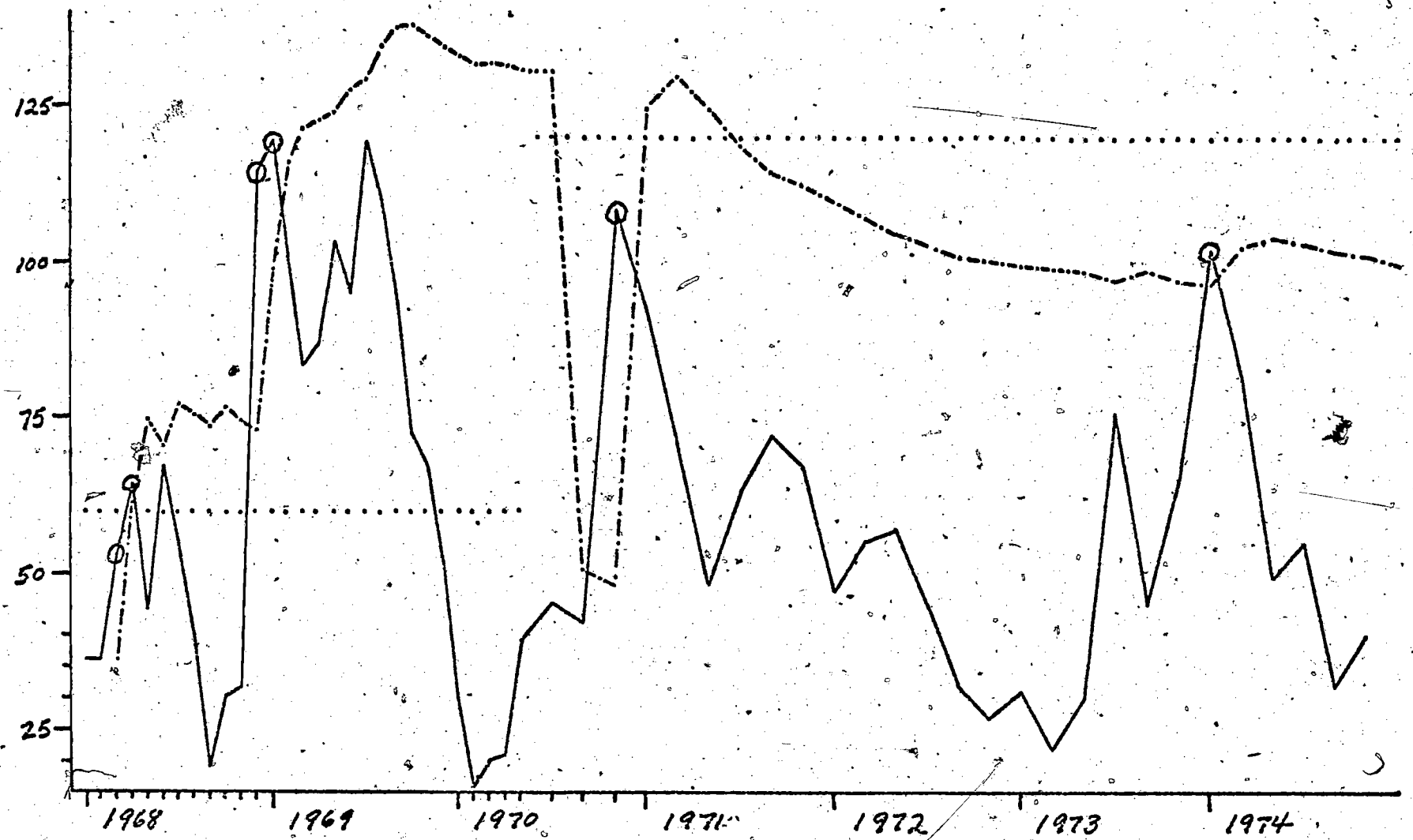
Graph 12b. Outdoor Arizona--monthly



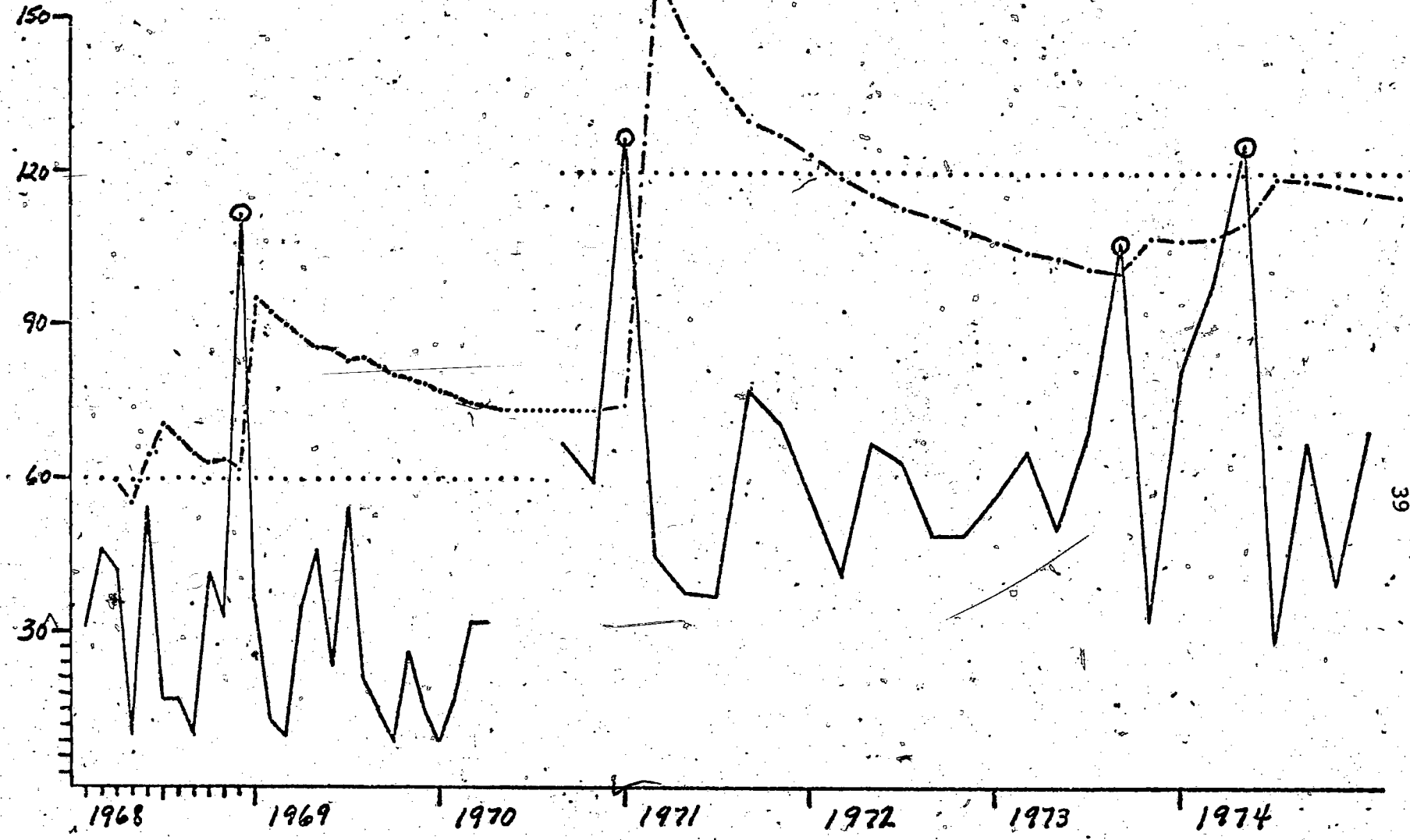
Graph 13a. Seventeen--monthly



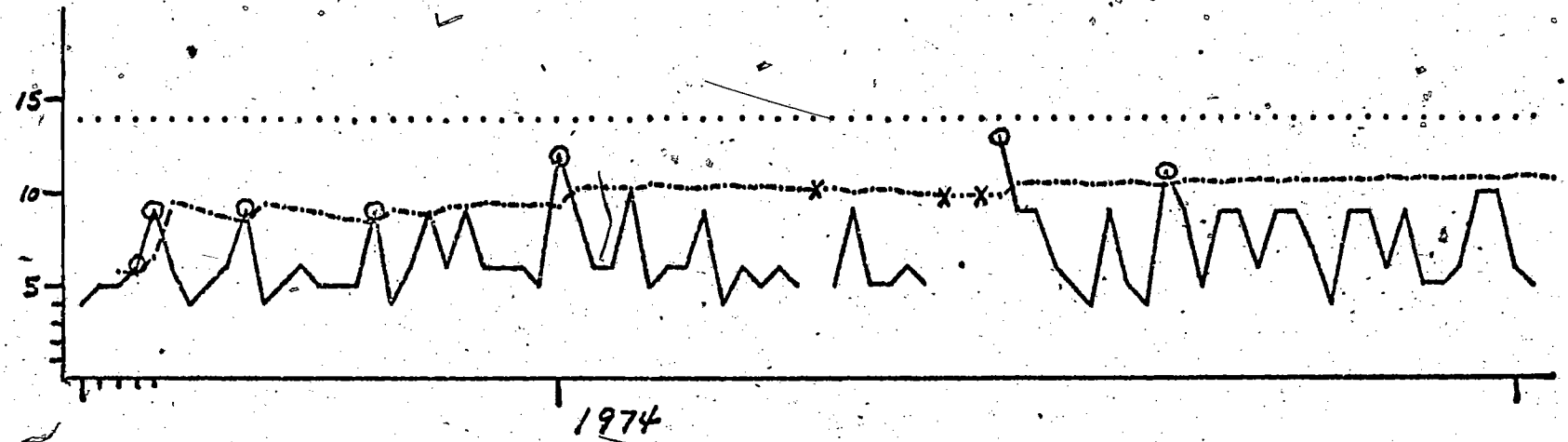
Graph 13b. Seventeen--monthly



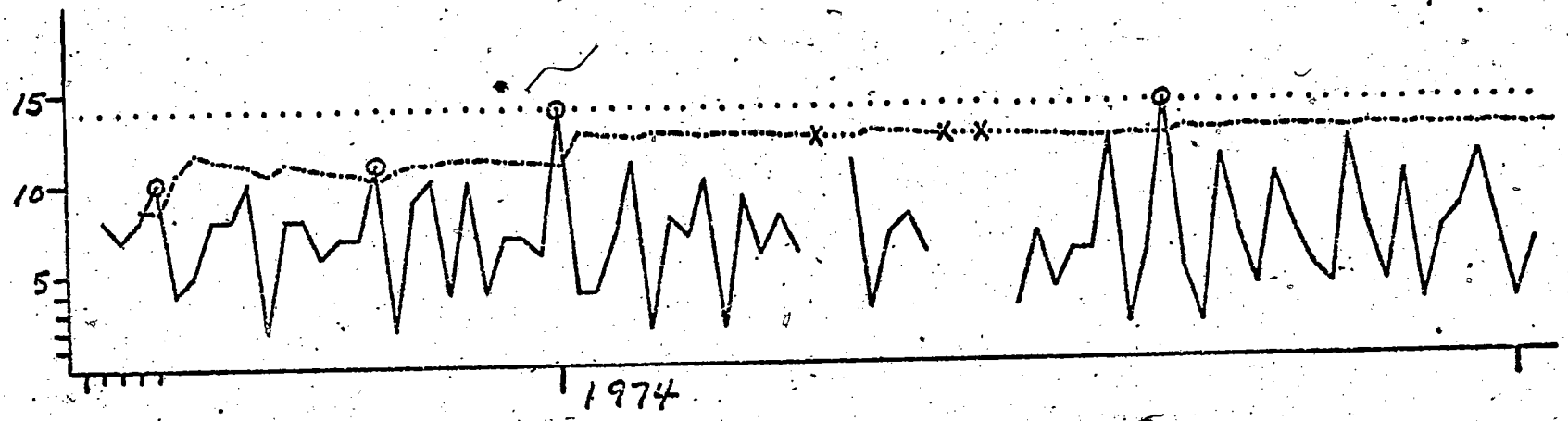
Graph 14a. I E E E Transactions. Power Apparatus and Systems--monthly, semi-monthly



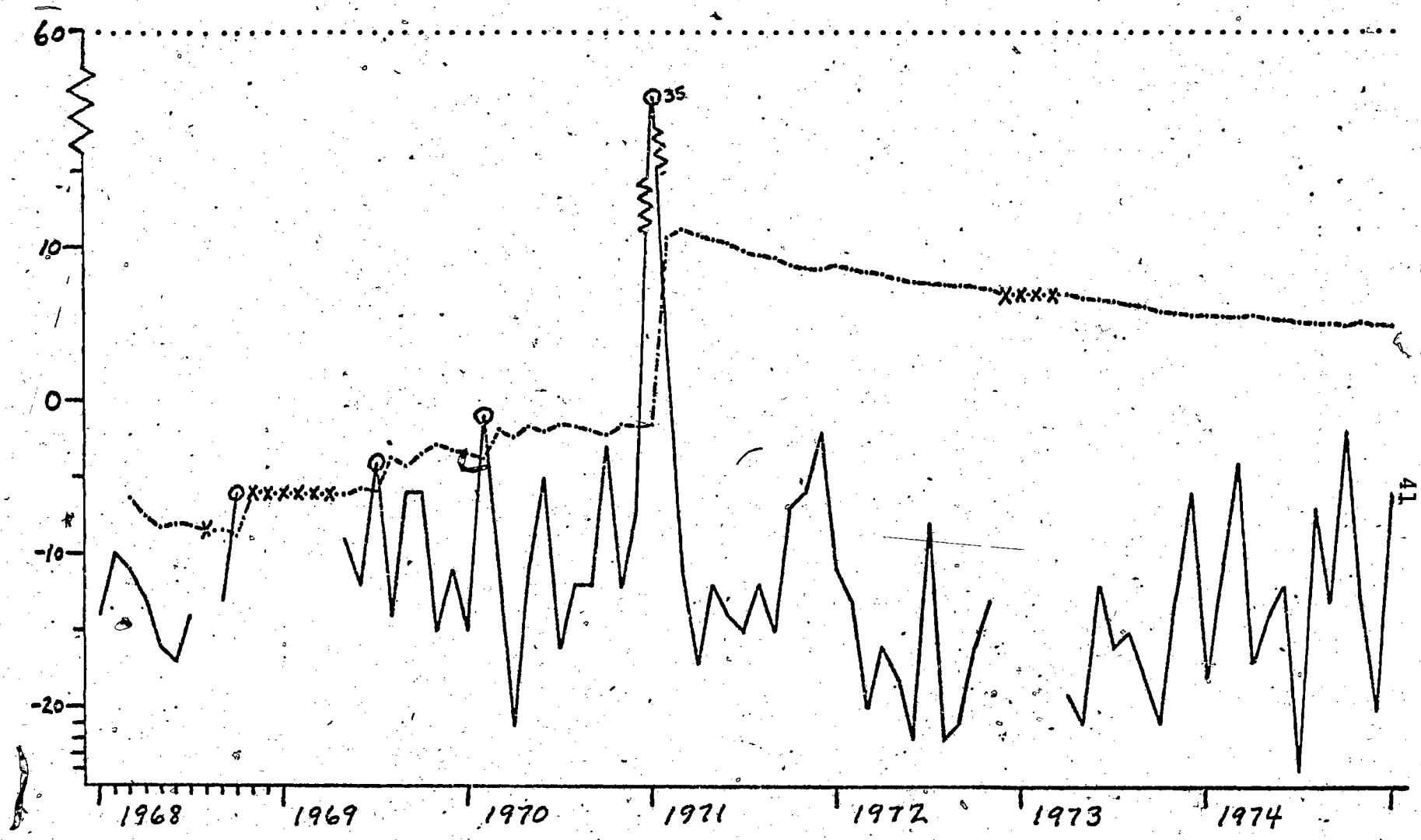
Graph 14b. I E E E Transactions. Power Apparatus and Systems--monthly, semi-monthly



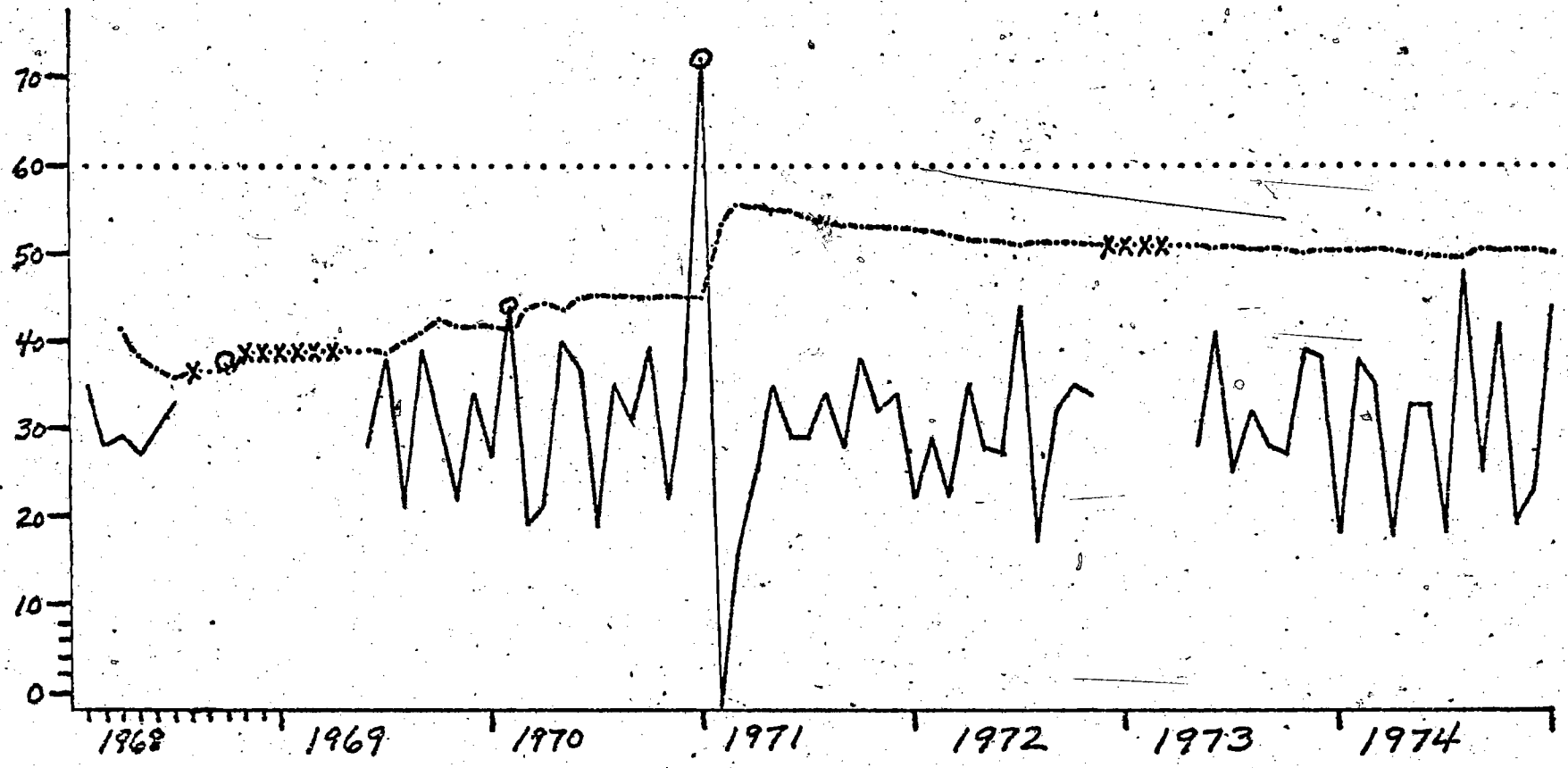
Graph 15a. Times Literary Supplement--weekly



Graph 15b. Times Literary Supplement--weekly



Graph 16a. Stereo Review--monthly



Graph 16b. Stereo Review--monthly

frequencies in the University of Utah Kardex file. The number and percentage of items from the random sample are indicated with each of the classes:

1. Semi-annual (one issue every two years)

   12 (1.6%).

2. Annual (one issue per year)

   112 (14.8%)

3. Bi-annual (two issues per year)

   45 (5.9%)

4. Tri-annual (three issues per year)

   24 (3.2%)

5. Quarterly (four per year and usually every three months)

   184 (24.3%)

6. Semi-monthly (one every two months and usually six per year)

   55 (7.4%)

7. Monthly (one every month and usually twelve per year--herein were included some which did not publish during certain seasons or months of the year but were monthly during the others)

   147 (19.5%)

8. Bi-monthly and Semi-weekly (twice per month or every other week)

   14 (1.9%)

9. Weekly (one per week with seasonal exceptions as in the case of the monthly)

23 (3.0%)

10. Irregular (no determined pattern)

133 (17.5%)

11. Odd (from three per month to five year cumulatives)

9 (1.2%)

Total = 758

Seventy-five of the 758 changed their arrival frequency at least once during the periods covered by the kardex cards in the sample. From the graphs the following data were compiled:

1. In the date-based analysis ('a' graphs) the Memmott algorithm 'claimed' forty-two issues which actually arrived at the University of Utah, without apparent claim, versus thirty-two issues claimed by the Jordan-Larsen algorithm.

2. The Memmott algorithm (on the 'a' graphs) made fifty-one 'claims' of issues, eventually claimed or missed by the University of Utah, earlier than the Jordan-Larsen algorithm versus two which the Jordan-Larsen algorithm made earlier than the Memmott algorithm.

3. In the interval-based analysis ('b' graphs) the Memmott algorithm 'claimed' twenty-five issues which actually arrived at the University of Utah without being claimed, versus ten issues claimed by the O.C.L.C. algorithm.

4. The Memmott algorithm (on the 'b' graphs) made fifty 'claims' of issues, eventually claimed or missed by the University of Utah, earlier than the O.C.L.C. algorithm versus three which the O.C.L.C. algorithm made earlier than the Memmott algorithm.

A breakdown of the results of the graphs follows in figure 6.

Explanation of Figure 6

Column 3. False claims represents the number of issues claimed by the Memmott System which were actually received at the University of Utah apparently without a claim being filed.

Column 4. Correct claims--before J. - L. represents the number of claims or holes found on the kardex card of the serial issue which were claimed by the Memmott System sooner than by the Jordan-Larsen System.

Column 5. False claims represents the number of issues claimed by the Jordan-Larsen System which were actually received at the University of Utah apparently without being claimed.

Column 6. Correct claims--before Memmott represents the number of claims or holes found on the kardex card of the serial issue which were claimed by the Jordan-Larsen System sooner than by the Memmott System.

		'a' graphs				'b' graphs			
		Memmott		Jordan-Larsen		Memmott		O.C.L.C.	
Frequency	Total number of issues	False claims	Correct claims--before J. - L.	False claims	Correct claims--before Memmott	False claims	Correct claims--before O.C.L.C.	False claims	Correct claims before Memmott
Semi-annual									
Annual	39	2	2	4	2	3	2	2	2
Bi-annual	20	2	none missing	3	none missing	0	none missing	0	none missing
Tri-annual									
Quarterly	89	5	2	9	0	3	1	4	1
Semi-monthly	124	12	7	2	0	8	7	2	0
Monthly	233	14	37	14	0	7	37	2	0
Bi-monthly									
Weekly	74	7	3	0	0	4	3	0	0
Irregular	55	-------	----not applicable		----------	3	3	--not applicable--	
Odd									
Total of the 31 graphs	634	42	51	32	2	28	53	10	3

Fig. 6. Results of graphs.

Column 7. False claims represents the number of issues claimed by the Memmott System which were actually received at the University of Utah apparently without a claim being filed.

Column 8. Correct claims--before O.C.L.C. represents the number of claims or holes found on the kardex card of the serial issue which were claimed by the Memmott System sooner than by the O.C.L.C. System.

Column 9. False claims represents the number of issues claimed by the O.C.L.C. System which were actually received at the University of Utah apparently without being claimed.

Column 10. Correct claims--before Memmott represents the number of claims or holes found on the kardex card of the serial issue which were claimed by the O.C.L.C. System sooner than by the Memmott System.

CHAPTER V

SUMMARY

Restatement of the Problem and Procedures

Methods of claiming missing issues of serials publications have been ad hoc. This study described four methods of claiming and analyzed the performance of three of the methods on sixteen arrival patterns. The arrival patterns, selected as workable examples from a random sample of the Kardex files at the University of Utah, were graphed against a statistically-based claiming algorithm (Memmott System) and two algorithms based on the arrival frequency plus a lag factor (Jordan-Larsen and O.C.L.C. Systems).

Conclusions and Recommendations

The Memmott System was shown to be a 'tighter' method of claiming among the sixteen serial issues tested. Of the fifty-three issues missed by the University of Utah library, fifty would have been claimed earlier by the Memmott system than by either the Jordan-Larsen or the O.C.L.C. systems. However, using the date method ('a' graphs) of determining arrival patterns, the Memmott algorithm would have claimed forty-two issues which eventually came without being claimed as compared with thirty-two for Jordan-Larsen. Using

the interval method ('b' graphs) of determining the arrival patterns, the difference was twenty-five early claims for the Memmott algorithm versus ten for the O.C.L.C. algorithm.

The O.C.L.C. and Jordan-Larsen algorithms allowed for more variation in the patterns of the serials and often claimed much later than the Memmott algorithm. It is possible that some items would go out-of-print before they were claimed by the O.C.L.C. or Jordan-Larsen algorithms but fewer items would be claimed too soon. Applied to annuals the Jordan-Larsen and O.C.L.C. algorithms often claimed sooner than the Memmott algorithm. The Jordan-Larsen and O.C.L.C. algorithms definitely performed better on the weekly pattern examined. The Memmott algorithm appears to have performed best on the quarterlies and monthlies.

The interval method of determining arrival patterns produced better results than the date method when applied to the sixteen examples. Fewer 'false' claims were made, and the interval method was the only one of the two which could be applied to irregulars. This might imply that publishers are more inclined to publish by intervals rather than meet a particular deadline, but nothing substantial can be concluded from such a small sample. Further research on that hypothesis would have to consider the various frequencies of publication and types of publishers (commercial, government, private, nonprofit, etc.).

The percentages of serials found in each frequency classification imply that those general percentages exist

in the entrie Kardex files (with the exception of newspapers). The sample size of 758 gives a reliability of ±4 percent with a 95 percent confidence level[1] or ±5 percent with a 99 percent confidence level.[2]

Further study of the data is necessary. More titles should be graphed and additional approaches to claiming must be investigated, but the author makes the following observations and recommendations based on this study:

1. The Memmott algorithm appeared to be feasible, especially if one more standard deviation were added to the claiming date. If C_i had been computed at the 99.9 percent confidence level (three standard deviations), many of the false claims (issues actually received late) would not have been made. If the Memmott algorithm were to be used, claiming would not begin until the arrival dates of several issues (possibly four to ten) were collected to let the pattern of arrival establish itself.

2. The O.C.L.C. and the Jordan-Larsen algorithms appeared to perform best for the longer and shorter frequency serials (annuals, weeklies, etc.).

3. No matter what system is in operation, the human interface should not be forgotten. Because of the nature of this study, no human interface was possible. The computer can only recommend that a claim be issued based on

---

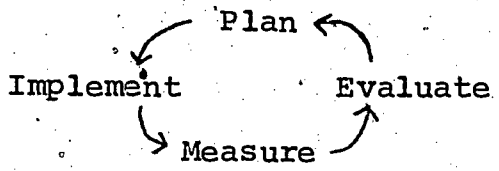
[1]Herbert Arkin and Raymond R. Colton, Tables for Statisticians (New York: Barnes & Noble, Inc., 1963), p. 145.

[2]Ibid., p. 146.

some algorithm it is using. The final judgment should be that of the person best acquainted with the serial title in question. The actual working situation is very different from the simulated situation of this study. The claiming system must have built into it enough flexibility to allow for alterations such as known mail delays, publication breakdown, etc.

4. Some irregulars can be claimed as shown on graph 9.

5. Arrival dates were essential to the analysis in this project. The project could not have been done without the help of the University of Utah Serials Section. The author is reminded of a basic principle from organizational behavior. That is, planning follows a circular process.



Proper planning cannot be done without a means of measuring what is happening. Evaluation cannot be competently done without data to evaluate. When the on-line system is in operation, the arrival dates for the issues could be stored for future use on magnetic tape rather than disc-pack. If the arrival dates were properly indexed for easy analysis, this would avoid the high cost of disc-pack storage space. In either case the arrival dates must be recorded and indexed for any future planning which is intended to be more than just speculation.

Future Research

Many procedures in serials departments are ad hoc. With the rise of computers in libraries many innovations are possible. This project has just touched on an area where little has been done.

The basic premise of the Memmott algorithm was that the arrival of serials in a library acts like a normal distribution. The author suggests a chi-squared test of goodness-of-fit as a possible check on the arrival patterns to determine whether they are normally distributed (see explanations on chi-squared test of goodness-of-fit in Lawrence D. Phillips and Robert Parsons in the Sources Consulted). No studies were found to verify or disprove this premise. The application of statistical methods in operating serials control systems is nonexistent.

Many factors may affect the arrival of a serial. Can the effects of some of these factors (weather, day of the week, strikes, distance from place of publication, country of origin, etc.) be isolated? None of the systems described in this paper attempts discrimination of any factors but simply groups them all together.

Costs are certainly important. What are the costs of maintaining each of the algorithms versus not having them? Would it be justifiable to cancel some subscriptions to have enough money to maintain a good serials claiming system? Computer applications of the claiming algorithms to systems are varied. As computer costs go down and serials

control becomes more imperative, the day might come when graphs such as have been plotted in this project could be done by a computer to give the serials librarian greater decision making capability in claiming.

The data collected in this study provide more opportunities for research. Many more graphs should be plotted; other algorithms should be devised and tested; and other aspects of serials should be investigated with the data in the sample. Samples at other libraries should be used to compare results. The author hopes this project will act as a catalyst to others.

APPENDIX A

IRREGULAR PERIODIC FREQUENCIES AND CLAIM DELAY CODE

APPENDIX A

IRREGULAR PERIODIC FREQUENCIES AND CLAIM DELAY CODE

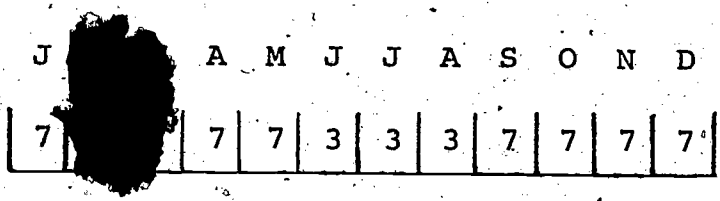


Fig. 1. The example above describes a periodical published biweekly except in the 3 months of summer when it is published monthly.[1]

Ø--Nothing Published
1--1st week
2--2nd week
3--3rd week
4--4th week
5--5th week
6--Weekly
7--Biweekly
8--1st & 2nd weeks
9--1st & 3rd weeks
A--1st & 4th weeks
B--1st & 5th weeks
C--2nd & 3rd weeks
D--2nd & 4th weeks
E--2nd & 5th weeks
F--3rd & 4th weeks
G--3rd & 5th weeks
H--4th & 5th weeks
J--1st, 2nd & 3rd weeks
K--1st, 2nd & 4th weeks
L--1st, 2nd & 5th weeks
M--1st, 3rd & 4th weeks
N--1st, 3rd & 5th weeks
O--1st, 4th & 5th weeks
P--2nd, 3rd & 4th weeks
R--2nd, 3rd & 5th weeks
S--3rd, 4th & 5th weeks
T--1st, 2nd, 3rd & 4th weeks
U--1st, 2nd, 3rd & 5th weeks
V--1st, 2nd, 4th & 5th weeks
W--1st, 3rd, 4th & 5th weeks
X--2nd, 3rd, 4th & 5th weeks

---

[1]Martin D. Fried and Ruth Dunham, California State Library: Processing Center Design and Specifications, Vol. 4: Serials Control System (Berkeley, California: Institute of Library Research, University of California, 1969), pp. 112-19.

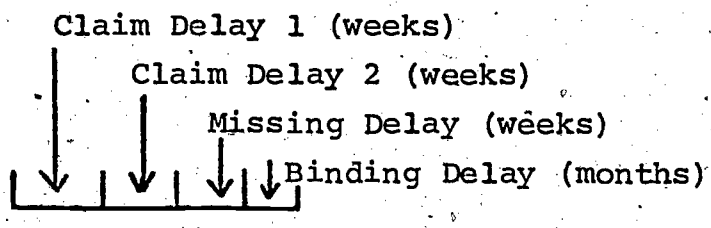


Fig. 2. Claim Delay Code.[1]

Claim Delay 1 indicates the number of weeks after publication to issue the first claim for an unreceived issue.

Claim Delay 2 indicates the number of weeks after publication to issue the second claim for an unreceived issue.

Missing Delay indicates the number of weeks after publication to declare an issue missing and discontinue producing notices.

Binding Delay indicates the number of months after prediction of the last issue of the binding unit before issuing a binding slip.

---

[1]The claim delay code is used for exceptions in shipping and publication pattern. A specified number of days, weeks, or even years is added to the computed date. This, of course, includes refinements according to the season and for special holidays. (See also appendix C, figure 5.)

APPENDIX B

MEAN AND STANDARD DEVIATION PROGRAM

APPENDIX B

MEAN AND STANDARD DEVIATION PROGRAM

```
 1 BEGIN
 2 INTEGER N; NEWLINE(1);
 3 LOOP: WRITE("TYPE THE NUMBER OF DATA POINTS TO BE AVERAGED"); NE
WLINE(1);
 4 READ(N);
 5         BEGIN
 6         INTEGER I,J,K;
 7         REAL SUM,AVE,VAR,STD,FIF;
 8         REAL ARRAY NUM[1:N];
 9         WRITE("INPUT NOS"); NEWLINE(1);
10         K := 2;
11         FOR J := 1 UNTIL N DO READ (NUM[J]);
12         LOOP2:  SUM := 0;
13         FOR I := 1 UNTIL K DO SUM := SUM + NUM[I];
14         AVE := SUM/K;
15         SUM := 0.;
16         FOR I := 1 UNTIL K DO SUM := SUM + (NUM[I] - AVE)↑2;
17         VAR := SUM/(K-1);  STD := SQRT(VAR);
18         FIF := AVE + STD / STD;
19         WRITE("MEAN ="); PRINT(AVE); NEWLINE(1);
20         WRITE("S D ="); PRINT(STD); NEWLINE(1);
21         WRITE("CONFIDENCE ="); PRINT(FIF); NEWLINE(2);
22         K := K + 1;
23         IF K > N THEN GO TO EXIT;
24         GO TO LOOP2;
25         END;
26 EXIT: GO TO LOOP;
27 END
```

APPENDIX C

FLOWCHARTS OF CLAIM SYSTEMS

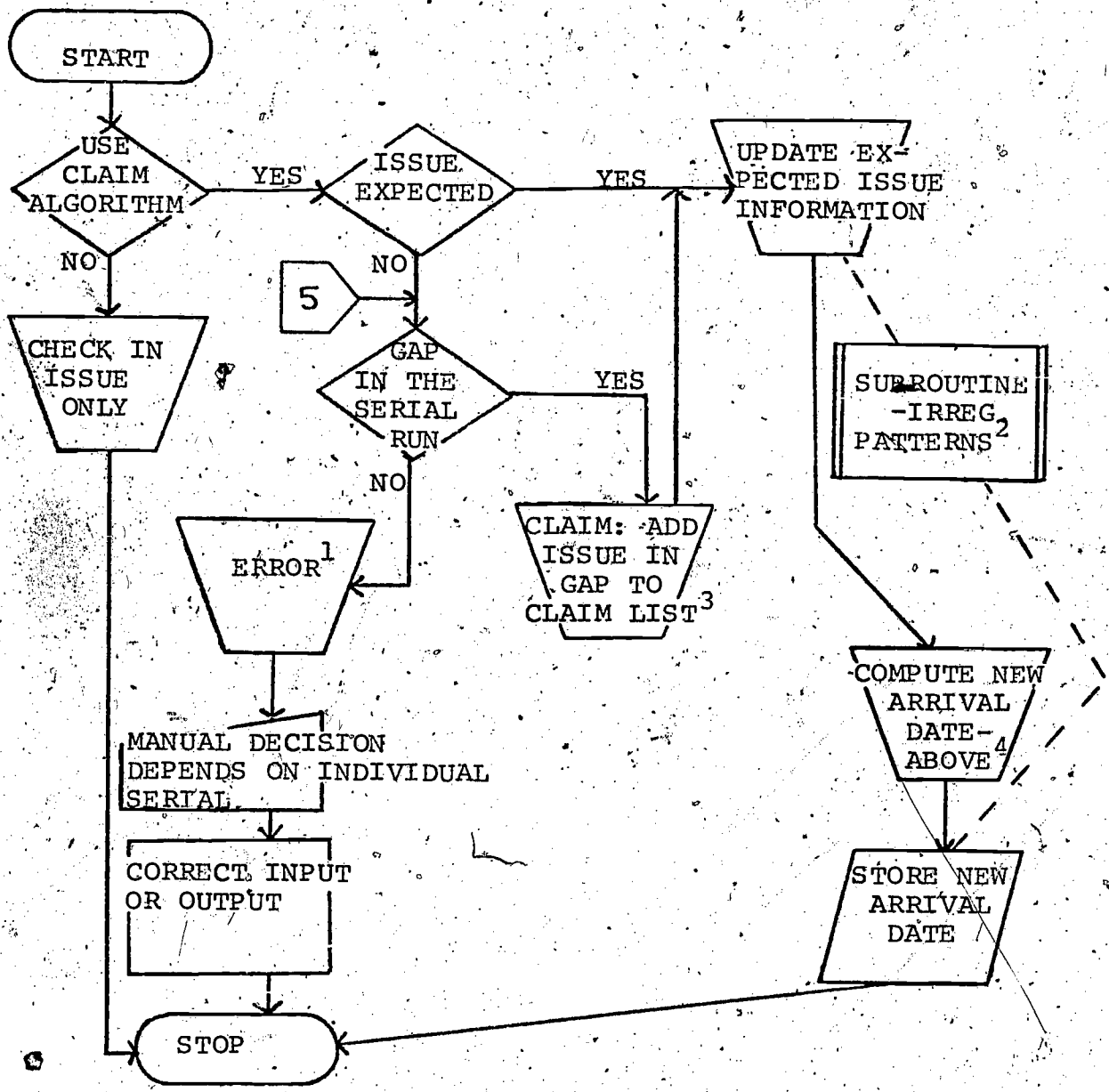


Fig. 1. Jordan-Larsen System

---

[1]An error may be caused by such occurrences as the arrival of a duplicate copy of an issue or the arrival of the first issue of a new subscription.

[2]See Figure 2.

[3]When claiming those issues found to be missing due to gaps they will be added to those claimed due to publisher defaults of expected arrival dates.

[4]See steps nos. 1 through 3 on pages 8-9.

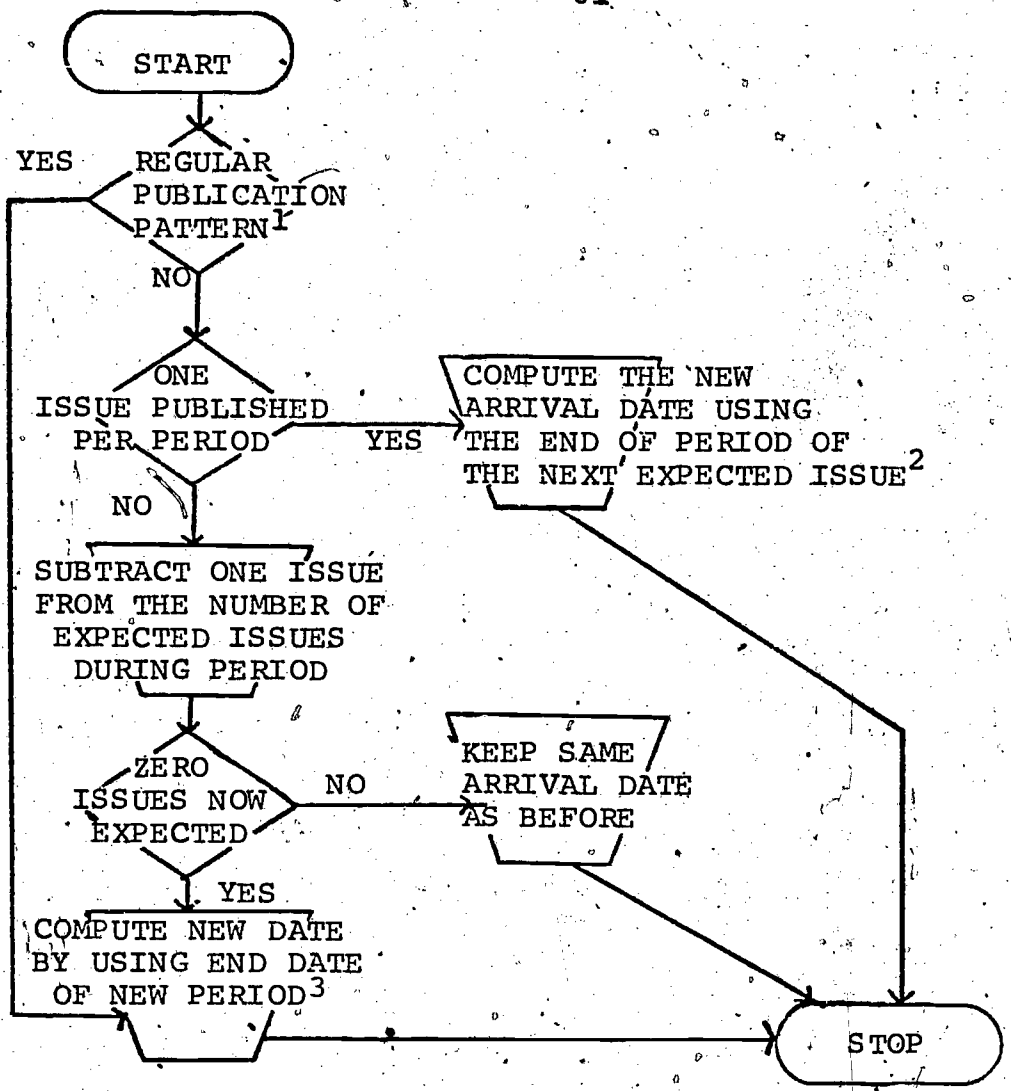


Fig. 2. Jordan-Larsen System: Irreg. Patterns of Publication

---

[1]Regular publication pattern means the serial is published at regular intervals (i.e., equal intervals of time between each issue).

[2]A publication pattern may be constructed to allow for irregular arrivals. The pattern being used above would divide time into intervals such as the months of the year. For each irregular serial using such a pattern the number of issues expected in each period would be noted and the issues would be expected to arrive in order during their respective arrival periods. This would cover the almost regulars (e.g., skip one month).

[3]When all the issues have arrived for a particular period, then the next period will become the new expected arrival period.

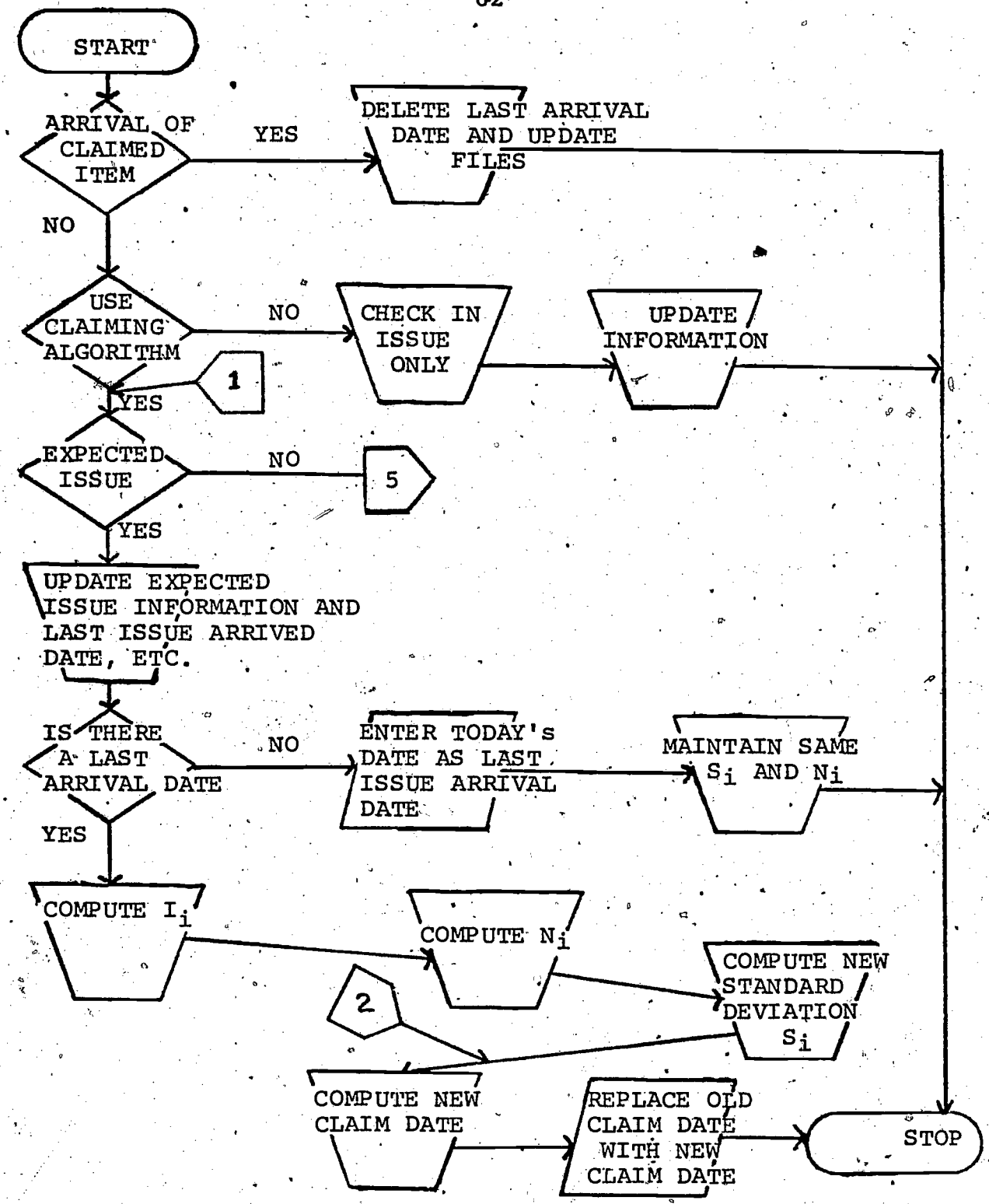


Fig. 3. Memmott System

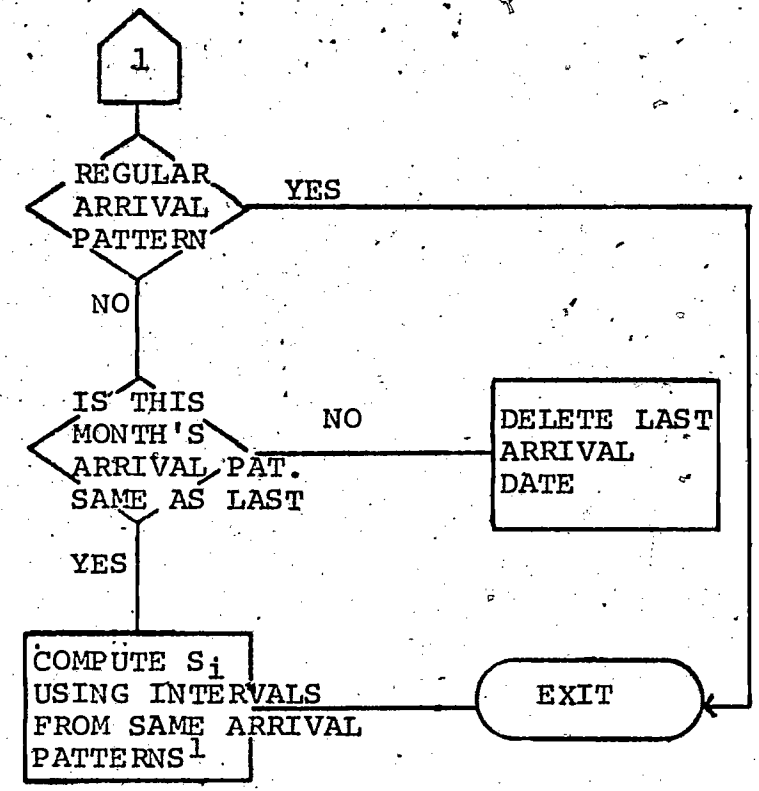


Fig. 4. Irregular Periodic Arrival Subroutine

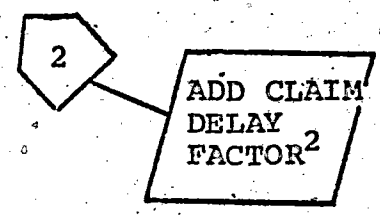


Fig. 5. Claim Delay Subroutine

[1]This is the date on which the next expected issue will be claimed.

[2]See Appendix A, Figure 2.

SOURCES CONSULTED

SOURCES CONSULTED

MONOGRAPHS

Arkin, Herbert, and Colton, Raymond R. Tables for Statisticians. New York: Barnes & Noble, Inc., 1963.

Bosseau, Don L. "The Computer in Serials Processing and Control." In Advances in Librarianship. Vol. 2. Edited by Melvin J. Voigt. New York: Seminar Press, 1971, pp. 104-61.

Bosseau, Don L. "Case Study of the Computer Assisted Serials System at the University of California, San Diego." In Proceedings of the LARC Institute on Automated Serials Systems. Edited by H. William Axford. Tempe, Arizona: The LARC Association, 1973, pp. 77-118.

Brown, Clara D. Serials: Acquisition & Maintenance. Birmingham, Alabama: EBSCO Industries, Inc., 1972.

Fayollat, James. "One-Line Serials Control at UCLA." In Proceedings of the 1972 Clinic on Library Applications of Data Processing. Edited by F. Wilfrid Bancaster. Urbana, Illinois: Graduate School of Library Science, University of Illinois, 1972, pp. 69-81.

Fried, Martin D., and Dunham, Ruth. California State Library: Processing Center Design and Specifications. Vol. 4: Serials Control System. Berkeley, California: Institute of Library Research, University of California, 1969.

Hammer, Donald P. "Serial Publications in Large Libraries: Machine Applications." In Serial Publications in Large Libraries. Edited by Walter C. Allen. Allerton Park Institute, no. 16. Urbana, Illinois: Graduate School of Library Science, University of Illinois, 1970, pp. 120-45.

Hayes, Robert M., and Becker, Joseph. Handbook of Data Processing for Libraries. New York: John Wiley & Sons, Inc., 1970.

International Business Machines Corporation. Library Automation--Computerized Serials Control. Library Automation Series. White Plains, N.Y.: International Business Machines Corporation, 1971.

Lazerow, Samuel. "Serial Records: A Mechanism for Control." In Serial Publications in Large Libraries. Edited by Walter C. Allen. Allerton Park Institute, no. 16. Urbana, Illinois: Graduate School of Library Science, University of Illinois, 1970.

Osborn, Andrew D. Serial Publications. 2d ed. Chicago: American Library Association, 1973.

Parsons, Robert. Statistical Analysis: A Decision-Making Approach. New York: Harper & Row, 1974.

Phillips, Lawrence D. Bayesian Statistics for Social Scientists. London: Thomas Nelson & Sons, Ltd., 1973.

Swihart, Stanley J., and Hefly, Beryl F. Computer Systems in the Library: A Handbook for Managers and Designers. A Wiley Becker & Hayes Series Book. Los Angeles: Melville Publishing Company, 1973.

SERIALS

Balch, Earl, ed. Serials Processing System Reference Manual. The LARC Reports. Vol. 5, no. 3. Tempe, Arizona: The LARC Association, Inc., 1972.

Bishop, David; Milner, Arnold L.; and Roper, Fred W. "Publication Patterns of Scientific Serials." American Documentation 16 (April 1965): 113-21.

Bosseau, Don L. "The University of California at San Diego Serials System--Revisited." Program 4 (January 1970): 1-29.

Fayollat, James. "On-Line Serials Control System in a Large Biomedical Library. Part III: Comparison of On-Line and Batch Operations and Cost Analysis." Journal of the American Society for Information Science 24 (March-April 1973): 80-86.

Fayollat, James, and Luck, Don. "Computer Based Serials Control System, Biomedical Library, UCLA." American Documentation 20 (October 1969): 385.

Hess, Susan M., and Roselle, William C., eds. Library Automation at the University of Wisconsin--Milwaukee. The LARC Reports. Vol. 6, no. 3. Peoria, Illinois: LARC Press, LTD., 1973.

Koenig, Michael E. D.; Finlay, Alexander C.; Cushman, Joanne G.; and Detmer, James M. "SCOPE: A Cost Analysis of an Automated Serials Record System." Journal of Library Automation 4 (September 1971): 129-40.

Markuson, Barbara Evans, ed. Application of Automation in American Libraries; An Analysis of the LARC Survey Returns. The LARC Reports. Vol. 3, no. 1. Tempe, Arizona: The LARC Association, Inc., 1970.

Roper, Fred W. "A Computer-Based Serials Control System for a Large Biomedical Library." American Documentation 19 (April 1968): 151-57.

Surace, Cecily J. "PEARL An Automated Periodicals Control System." Special Libraries 63 (September 1972): 385-93.

Wall, R. A. "A Proposed Experiment in Automated Serials Accessioning." Program 5 (July 1971): 141-51.

Wilkinson, W. A., and Stock, Loretta A. "Machine-Assisted Serials Control." Special Libraries 62 (December 1971): 529-34.

MICROFILM REPORTS

Burns, Robert W., Jr. The Design and Testing of a Computerized Method of Handling Library Periodicals (Title III). Bethesda, Md.: ERIC Document Reproduction Service, ED 050 753, 1970.

Five Associated University Libraries. Joint Serials Control System Project for the Libraries of Cornell University, University of Rochester and the State University of New York at Buffalo. Phase I Feasibility Study Final Report. Bethesda, Md.: ERIC Document Reproduction Service, ED 051 827, 1971.

Overmyer, LaVahn. Library Automation: A Critical Review. Bethesda, Md.: ERIC Document Reproduction Service, ED 034 107, 1969.

Pan, Elizabeth. Library Serials Control Systems. A Literature Review and Bibliography. Bethesda, Md.: ERIC Document Reproduction Service, ED 044 538, 1970.

UNPUBLISHED REPORTS

Taylor, John R. "Report on Visit to the Serial Section of the U.C.L.A. Biomedical Library." Provo, Utah, 26 August 1971. (Typewritten.)