GENERATING SYNTHETIC STREAM-FLOW RECORDS FOR NEW ZEALAND RIVERS

W. C. Boughton* and A. McKerchar*

ABSTRACT

The Thomas-Fiering model for generation of synthetic stream-flow records has been used with data from some New Zealand catchments. Available records are sparse and only four gauging stations have records of more than 20 years. Possible errors which can occur when short records are used to estimate monthly means, serial correlations, etc., are demonstrated. A simple simulation model for testing the difference between historic and synthetic records in determining reservoir size is described.

INTRODUCTION

The sequence of monthly flows which form the stream-flow record at a gauging site on a river is the prinicipal factor determining the size of reservoir needed to meet a given demand, or in determining the amount of water that can be utilized for irrigation, power production, or other purposes. Rarely is the record long enough to give confidence in the results which are obtained.

The factor which most influences the utilization of the flow is the worst sequence of low flows occurring together in the record. One critical sequence of low flows is usually the factor which determines the size of a reservoir for regulating the flow, or determines the amount of water which can be continuously withdrawn from the river.

The short samples of stream-flow record which are normally available may not contain a representative sample of the low-flow sequences which can occur. If a number of different samples of stream-flow record from the same site could be analysed, different results of reservoir size or water availability would be obtained because of the different patterns of flow in the samples.

Beard (1967) demonstrated this effect with 12 stream-flow records, each about 50 years in length, from rivers in the United States. He divided each record into two samples of 25 years, and determined the reservoir size required to meet a stated demand

* Agricultural Engineering Department, Lincoln College.

using first one 25-year sample and then the other. The storage sizes determined from the two halves of a record differed by more than a factor of 2 in two-thirds of the cases.

If a large number of samples of stream-flow record were available at a gauging site, it would be possible to examine all of them and determine the probability of a reservoir of given size meeting a particular demand at all times. In the absence of very long records of flow, from which many samples of, say, 50 years could be drawn, use is made of synthetic-hydrology techniques to generate long periods of artificial records which are statistically the same as the historic record.

The concept of generating artificial records is not new. As early as 1914, Hazen (1914) synthesized a flow sequence of 300 years by combining the records from 14 streams into a single record. Later, Sudler (1927) introduced an element of randomness by printing the 50 annual flows from a 50-year record onto cards with one annual flow per card. By repeated shuffling and dealing 20 times through the deck, a sequence of 1,000 years of flow was generated.

Barnes (1954) used a method similar to that of Sudler, except that the synthetic flows were drawn from a normal distribution with the same mean and variance as flows in the historic record. Barnes introduced the use of random numbers for determining the values of the flows.

The early attempts to synthesize records had a number of weaknesses. Serial correlation between flows was generally ignored, and this prevents use of the methods with monthly flow sequences. Also, the restricted range of flows in the synthetic records and the types of distribution used to describe the flows are other limitations of significance.

Thomas and Fiering (1962) introduced a statistical model which overcame the weaknesses of those earlier models and which has now found wide acceptance. The method is flexible in that it can be used for weekly, monthly, or seasonal flows as well as annual flows. It does not require that the flow data be normally distributed and may be used with skewed distributions. It also incorporates serial correlation between successive flows so as to accord with observed stream flows.

THE THOMAS-FIERING MODEL

Thomas and Fiering use a linear-regression model between flows in successive time periods. For example, using monthly flows, flow in February is related to flow in January by a linear regression as shown in Fig. 1. Flow in March is related to flow in February by another linear regression. Twelve linear regressions are used to

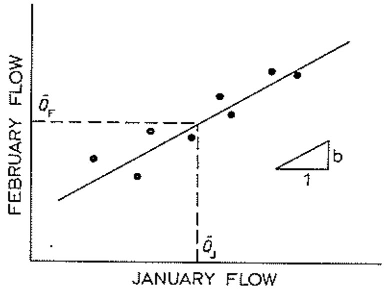

FIG. 1 — Linear regression relating February flows to January flows.

describe the serial correlation of flows through the year, the final regression relating flow in January to flow in the preceding December.

The regression line relating flow in February to flow in January is given by:

$$(Q_F - \bar{Q}_F) = b \, . \, (Q_J - \bar{Q}_J) \quad \ldots\ldots \quad \ldots\ldots \quad \ldots\ldots \quad (1)$$

where Q_F = flow in February,
$\bar{Q}_F$ = mean of Febuary flows,
Q_J = flow in January,
$\bar{Q}_J$ = mean of January flow,
b = slope of the regression line.

The linear regression line passes through the point given by the two mean values, $\bar{Q}_J$ and $\bar{Q}_F$, and has a slope of b.

The regression line is inadequate by itself to describe the serial correlation between monthly flows. For example, if the starting flow in January was equal to the mean January flow $\bar{Q}_J$, this would generate the mean February flow $\bar{Q}_F$. This in turn would generate the mean March flow, and so on, with the mean December flow generating the mean January flow and repeating the same sequence each year.

Some random component is required in the generating procedure to allow for the natural variations in flow which are not accounted for by the regression line. It should be noted that the February flows form a set of data with mean value $\bar{Q}_F$ and variance s_F^2. Part of the variance is accounted for by the regression line. If the correlation coefficient between February flows and January flows is r, then the proportion of the total variance of February flows that can be attributed to variation in January flows is equal to r^2.

$$\underset{\text{total variance}}{s_F^2} = \underset{\text{variance explained by regression}}{s_F^2 . r^2} + \underset{\text{variance unaccounted for by regression}}{s_F^2 . (1 - r^2)} \quad \ldots\ldots \quad \ldots\ldots \quad \ldots\ldots \quad (2)$$

The standard error of estimate of the February flows from the January flows is $s_F \cdot (1-r^2)^{\frac{1}{2}}$, and this is the measure of the random or unexplained variation in February flows. Thomas and Fiering incorporated this additional variance into their generating model as follows:

$$(Q_F - \overline{Q}_F) = b_1 \cdot (Q_J - \overline{Q}_J) + t.s_F \cdot (1 - r_1^2)^{\frac{1}{2}} \quad \ldots\ldots \quad (3)$$

Where t is a random number drawn from a standardized normal distribution with mean = zero, and variance = 1.

Forty-eight parameters are required to describe the 12 regression equations which form the model:

12 values of mean monthly flow ($\overline{Q}_J$, $\overline{Q}_F$, etc.),
12 slopes of the regression lines (b_1, b_2, etc.),
12 correlation coefficients (r_1, r_2, etc.),
12 standard deviations (s_J, s_F, etc.).

A starting value is assumed for flow in the first month. A random number, t, is selected and the value of flow in the second month is calculated from equation (3). A second random number is then used with flow in the second month to generate flow in the third month, and so on. Once started, the generating procedure can be used to produce a synthetic record as long as is required.

The Fortran programme for operation of the model is set out in Appendix 1.

RESULTS

The present study of the Thomas-Fiering model has so far dealt with variations in values of the parameters from different samples of record, and with application of the model to a simple reservoir-simulation model.

TABLE 1 — Monthly flow records.

River	*Gauge*	*Record (years)*
Kaituna	L. Rotoiti outlet	59
Waikato	Taupo	58
Waiau	L. Manapouri outlet	33
Waiau	L. Te Anau outlet	33
Pukaki	Lake spillway	18
Tarawera	Lake outlet	16
Ahuriri	Benmore	15
Tekapo	Tekapo dam	14
Rangitaiki	Te Teko	11
Kaituna	Te Matai	10
Whakatane	Whakatane	8
Motu	Houpoto	8
Tongariro	Turangi	8
Mohaka	Raupunga	7

Most of the long records of stream flow in New Zealand are from stations at the outflow from lakes or storages, and are records of regulated flow instead of natural flow. Table 1 lists the records of monthly flow, taken from the 1963 and 1964 *Hydrology Annual*, which have been used in the study.

Two of these records, Pukaki and Tekapo, are recorded at hydro-electric power installations and are therefore of little value for simulation use. Only four records are of more than 20 years length and these are all recorded at lake outlets. A certain degree of natural regulation is imposed on each. The two long-term records, from the Waikato and Kaituna Rivers, will each extend for 62 years if updated to the end of 1967.

For some years to come it will be necessary to use short lengths of record to estimate the monthly means, variances, and correlations for use in the model. Current opinion is that 20 to 30 years of stream-flow record are needed in order to get reliable estimates of these parameters. Studies by the authors to date indicate that this opinion is probably correct for New Zealand rivers.

The 48 parameters for the Waiau River at Lake Te Anau outlet based on a record of 33 years are listed in Table 2.

TABLE 2 — Data for Waiau River at Lake Te Anau outlet.

Month	*Monthly mean flow (cusecs)*	*Standard deviation*	*Slope (b)*	*Correlation coefficient (r)*
Jan	10,172	2,993	0.43	0.50
Feb	10,097	4,042	0.80	0.59
Mar	9,441	3,637	0.66	0.74
Apr	9,964	3,733	0.80	0.77
May	10,100	3,975	0.74	0.70
Jun	9,056	2,768	0.43	0.62
Jul	8,196	2,210	0.25	0.31
Aug	7,662	2,017	0.15	0.16
Sep	8,776	2,387	0.68	0.59
Oct	11,347	3,454	0.71	0.47
Nov	12,082	3,495	0.46	0.46
Dec	11,276	3,542	0.68	0.67

There are 18 different samples, each 16 years in length, which can be drawn from this historic record. The first sample starts in year 1 and finishes in year 16, then year 2 to year 17, and so on, finishing with the sample from year 18 to year 33.

A study has been made to see how the parameters of this record differ among the different 16-year samples. Fig. 2 illustrates how the correlation coefficient between February flows and January flows varies among the samples. If a 15-year record from 1933 to 1947 had been used to estimate the coefficient, a value of 0.02 would have

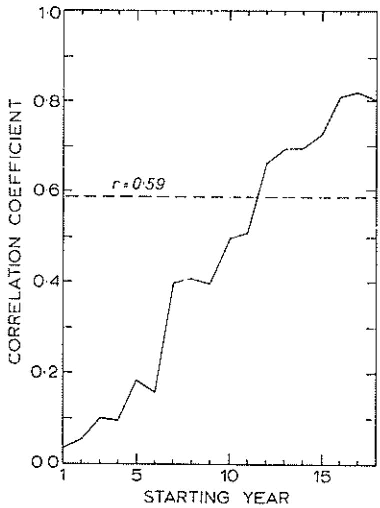

FIG. 2 — Variation in February–January correlation coefficient using different samples of flow record.

been obtained. If the sample from 1948 to 1963 had been used, a value of 0.82 would have been obtained. Using the whole record, a value of 0.59 was obtained.

This variation is the worst encountered in this particular record, but it does indicate the variation which can be encountered and which should be taken into account. No finality has yet been reached about the effect such variation is likely to have on practical design such as in determining the capacity of a proposed reservoir.

As a preliminary step before using synthetic flows for study of an actual river basin, a simple simulation model has been used to examine the difference between the results obtained from use of the historic record and use of synthetic flows. The model comprised a single reservoir with a steady uniform demand and without consideration of any seepage, evaporation, or other losses. A number of combinations of storage capacity and demand level were considered in turn. For each combination the behaviour of the reservoir was studied using first the historic record and then four synthetic records.

Failure of the dam was recorded whenever there was insufficient water remaining in storage to meet the monthly demand. The percentages of time the demand was met without failure are listed in Fig. 3, which shows the results from use of the historic record, and the average of results from the four synthetic records in brackets below. Both capacity and demand are expressed in terms of the mean monthly flow (i.e. one-twelfth of the mean annual flow).

A similar pattern to that in Fig. 3 has been found with stream-flow records from other gauging stations.

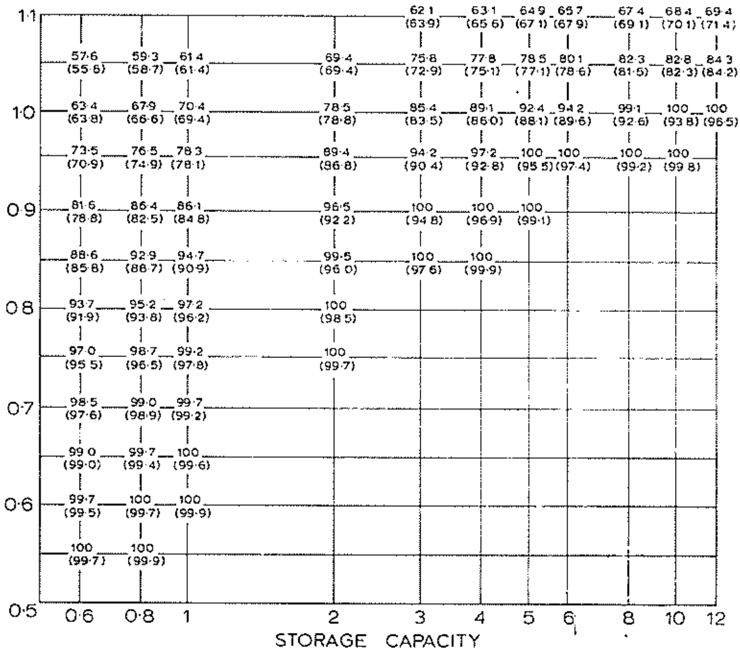

FIG. 3 — Percentage of time supply from dam meets demand — synthetic record in brackets. Horizontal and vertical scales are in terms of the mean monthly flow.

COMMENTS

This paper reports work in progress rather than a completed study, and it seems pertinent to give a description here of some related aspects which form part of the project.

Other Models

An alternative model to the Thomas-Fiering model has been put forward by Quimpo (1968) and by Yevjevich and Roesner (1966). This model is based on three components of a stream-flow record — a periodic component, a serial correlation component, and a random component.

A detailed review of this and other models for the generation of daily, monthly, and annual stream flows has recently been made (McKerchar, in press). When published, this review will give a description of additional aspects of daily and annual flows which are not covered here.

Multi-stream Models

A number of generating procedures have been put forward for the simultaneous generation of flows at a number of related stations. When flows are to be generated at two or more stations in close proximity to each other, it is necessary to consider cross-correlations between the stations in addition to serial correlation at each station.

Beard (1965) proposed a multiple linear-regression approach which has certain deficiencies. Fiering (1964) used a principal-component analysis which improved on this. Young and Pisano (1968) have demonstrated a straightforword approach which seems suitable for most analyses of river basins. This latter method is being programmed for the computer at Lincoln College and will be used for future studies of river basins in New Zealand.

Other Approaches

At present there is reason to believe that, in some cases, better results in the generation of monthly flows may be obtained by generation of daily rainfalls for input to a catchment model which in turn determines the volume of run-off. The cases in which this procedure seems desirable are where flow records are very short but rainfall records long and reliable, and also on small catchments where flow is ephemeral and many months have zero flow.

The generation of daily rainfall sequences is being based on the Bernoulli Urn models described by Wiser (1966) and is intended for use with the mathematical catchment model described by Boughton (1968).

REFERENCES

Barnes, F. B. 1954: Storage required for a city water supply. *J. Inst. Engs Aust. 26*: 198.

Beard, L. R. 1965: Use of interrelated records to simulate streamflow. *J. Hyd. Div. ASCE 91*: 13–22.

Beard, L. R. 1967: Hydrologic simulation in water yield analyses. *J. Irr. and Drain. Div. ASCE 93* (IR1): 33–42.

Boughton, W. C. 1968: A mathematical catchment model for estimating run-off. *J. Hydrol. (N.Z.) 7* (2).

Fiering, M. B. 1964: A multivariate technique for synthetic hydrology. *J. Hyd. Div. ASCE 89*: 43.

Fiering, M. B. 1967: *Streamflow synthesis*. Cambridge, Harvard University Press.

Hazen, A. 1914: Storage to be provided in impounding reservoirs for municipal water supply. *Trans. ASCE 77*: 1539.
Hufschmidt, M. M.; Fiering, M. B. 1966: *Simulation techniques for the design of water resource systems.* Cambridge, Harvard University Press.
Quimpo, R. G. 1968: Stochastic analysis of daily river flows. *J. Hyd. Div. ASCE 94* (HY1): 43–57.
Sudler, C. E. 1927: Storage required for the regulation of streamflow. *Trans. ASCE 91*: 622.
Thomas, H. A.; Fiering, M. B. 1962: Mathematical synthesis of streamflow sequences in the analysis of river basins. In: A. Maass *et al. Design of water resource systems.* Cambridge, Harvard University Press. Ch. 12.
Wiser, E. H. 1966: Monte Carlo methods applied to precipitation-frequency analysis. *Trans ASCE 19* (4): 538–542.
Yevjevich, V. M.; Roesner, L. A. 1966: *Mathematical models for time series of monthly precipitation and monthly run-off.* (Hydrology Paper 15, Colorado State University, Fort Collins, Colorado.)
Young, G. K.; Pisano, W. C. 1968: Operational hydrology using residuals. *J. Hyd. Div. ASCE 94* (HY4): 909–923.

APPENDIX 1

Computer Programme for Generating Synthetic Stream-flow Records.

```
      SUBROUTINE ECORE (X,Y,N,S,LSP,M,EN,KKK,RC,CC)
C
C            A.I.MCKERCHAR, LINCOLN COLLEGE, 28 FEB. 1968
C            ECORE FINDS MEANS, STANDARD DEVIATION,
C            CORRELATION AND REGRESSION COEFFICENTS ON
C            A MONTHLY BASIS. AND ALSO THE COEFFICIENTS
C            OF SKEWNESS.
C  IF  KKK=2 OR 3, NO OUTPUT IS PRODUCED AND S/R IS
C            USED BY EMILY
      DIMENSION X (12,60),Y(12),S(12)
      DIMENSION RC(12),CC(12)
      MD=M+LSP-1
   78 FORMAT(//' MONTH MEAN ST. DEV. COEF. SKEWNESS
             REGRESS.COEF.
     1CORREL.COEF. NO.MONTHS. '//)
      GO TO(30,31,31),KKK
   30 WRITE(3,78)
   31 DO 73 I=1,12
      T1=0.
      T2=0.
      T3=0.
      SUMX=0.0
      SUMY=0.0
      SUMSQ=0.0
      SUMCU=0.0
      ENN=EN
      K=I-1
      IF(K)70,70,71
   70 K=12
   71 DO 72 J=M,MD
      IF(12-K)81,81,82
   81 L=J-1
      IF(L)75,75,83
```

```
82  L=J
83  IF(X(I,J))75,76,76
83  IF(X(K,L))75,77,77
77  T1=T1+X(I,J)*X(K,L)
    T2=T2+X(I,J)**2
    T3=T3+X(K,L)**2
    SUMX=SUMX+X(I,J)
    SUMY=SUMY+X(K,L)
    GO TO 72
75  ENN=ENN-1
72  CONTINUE
    SUMX=SUMX/ENN
    SUMY=SUMY/ENN
    Y(I)=SUMX
    DO 26 J=M,MD
    IF(X(I,J))26,28,28
28  SUMSQ=SUMSQ+(X(I,J) -SUMX)**2
    SUMCU=SUMCU+(X(I,J) -SUMX)**3
26  CONTINUE
    S(I)=SQRT(SUMSQ/(ENN-1.0))
    CSK=SUMCU/S(I)**3*(ENN/((ENN-1)*(ENN-2)))
    CCOEF=(T1-ENN*SUMX*SUMY)/SQRT((T2-ENN*
   1SUMX**2)*(T3-ENN*SUMY**2))
    RCOEF=(T1-ENN*SUMX*SUMY)/(T3-ENN*SUMY**2)
    GO TO(40,42,42),KKK
42  CC(I)=CCOEF
    RC(I)=RCOEF
    GO TO 73
40  CONTINUE
    SO=S(I)+0.005
    CCOEF=CCOEF+0.00005
    RCOEF=RCOEF+0.00005
    SUMXO=SUMX+0.005
    WRITE(3,74) I,SUMXO,SO,CSK,RCOEF,CCOEF,ENN
74  FORMAT(I5,2F9.2,F14.3,2F14.4,F11.0/)
73  CONTINUE
    RETURN
    END
    SUBROUTINE EGENR(Q,XB,S,RC,CC,N,KANN,KTYPE,
        NYEAR,NCYCL,KPRNT,IX)
    DIMENSION Q(12,60),XB(12),S(12),RC(12),CC(12)
    COMMON J12,J16,J01,J02,J03
    COMMON RB,RE,RS,SB,SE,SS,NSTOR,NREG,BUGQ
C       INITIALIZE GENERATION BY READING 2 RANDOM
C       DIGITS FIRST JANUARY IS REGRESSED ON ANY
C       DECEMBER PICKED FROM RECORD.
20  FORMAT(' THINK OF A 5 DIGIT ODD NUMBER, LESS
   1THAN 30000. AND TYPE IT ON THE CONSOLE KEYBOARD,
   1OR READ IT FROM A CARD. ')
21  FORMAT(I5)
22  FORMAT(/' THANK YOU, NOW ANOTHER 2 DIGIT
   1NUMBER WITHIN THE RANGE 1-',I2)
23  FORMAT(I2)
36  FORMAT(' TRY A POSITIVE NUMBER. ')
37  FORMAT(' TRY AN ODD NUMBER. ')
29  FORMAT(' NUMBER IS OUTSIDE PERMITTED RANGE,
   1TRY ANOTHER. ')
```

```
   30 FORMAT(' FOR CERTAIN REASONS THAT NUMBER IS
     1UNSUITABLE, TRY ANOTHER. ')
      IV3=0
      IF(NCYCL-1)38,38,40
   38 WRITE(J01,20)
   35 READ(JI6,21)IX
      IF(IX)31,31,32
   32 RX=IX
      IRX=RX/2.0
      IRX=IRX*2
      IF(IX-IRX)33,15,33
   31 WRITE(J01,36)
      GO TO 35
   15 WRITE(J01,37)
      GO TO 35
   33 WRITE(J01,22)N
   96 READ(JI6,23)NRAND
C         CHECK RANGE OF NRAND
      IF(NRAND-N)24,24,25
   24 IF(NRAND)25,25,26
   26 IF(Q(12,NRAND))27.27.28
   25 WRITE(J01,29)
      GO TO 96
   27 WRITE(J01,30)
      GO TO 96
   28 PREVM=Q(12,NRAND)-XB(12)
      WRITE(J03,79)IX,NRAND
      GO TO 42
C         NRAND PERMITS REGRESSING 1ST JAN. ON ANY
C         HISTORIC DEC. FOR STORING ONTO DISK MAKE
C         NYEAR A MULTIPLE OF 10.
   40 CONTINUE
      GO TO (50,50,51),KTYPE
   50 PREVM=BUGQ—XB(12)
      GO TO 42
   51 PREVM=ALOG(BUGQ)-XB(12)
   42 QNEG=0.0
      NEG=0
      DO 66 J=1,NYEAR
      DO 61 I=1,12
      CALL GAUSS(IX,1.0,0.0,V)
      IF(V-3)550,551,551
  551 IV3=IV3+1
      V=3
  550 IF(V+3)552,552,553
  552 V=-3
      IV3=IV3+1
  553 CONTINUE
      Q(I,J)=XB(I)+RC(I)*PREVM+V*S(I)*SQRT(1.0-CC(I)
     1**2)
      PREVM=Q(I,J)-XB(I)
C         TAKE NATURAL VALUES OF SYNTHETIC RECORD,
C         COUNTING NUMBER AND MAGNITUDE OF NEGA-
C         TIVE VALUES, AND THEN SET THEM TO ZERO. IF
C         KANN=2, ADJUST MONTHLY FLOWS FOR COMPATI-
C         BILITY WITH ANNUAL VALUES.
      IF(Q(I,J))62,60,60
```

```
 62   NEG=NEG+1
      QNEG=QNEG+Q(I,J)
      Q(I,J)=0.0
      GO TO 61
 60   CONTINUE
      GO TO (61,61,64),KTYPE
 64   Q(I,J)=EXP(Q(I,J))
 61   CONTINUE
 66   CONTINUE
      BUGQ=Q(12,NYEAR)
555   FORMAT(' NUMBER OF RANDOM NUMBERS REQUIRING
     1TRUNCATION TO WITHIN RANGE -3 TO +3 WAS ',I3)
 65   FORMAT(' NUMBER OF NEGATIVE FLOWS GENERATED
     1AND SUBSEQUENTLY SET TO ZERO WAS ',I3,'. THEIR
     1TOTAL WAS 'F10.1,/)
 69   FORMAT(///' SYNTHETIC SEQUENCE NUMBER ',I3/' ------
     1----------',/)
 70   FORMAT(' THIS SEQUENCE WAS GENERATED USING A
     1NORMAL DISTN.',/)
 71   FORMAT(' THIS SEQUENCE WAS GENERATED USING A
     1LOGARITHMIC TRANSFORM.',/)
 79   FORMAT(' 1ST RANDOM NO. WAS',I7, '. INITIALIZING DEC.
     1VALUE WAS FROM HIST. RECORD, YEAR NO. ',I2,/)
      WRITE (J03,69)NCYCL
      GO TO (75,75,76),KTYPE
 75   WRITE (J03,70)
      GO TO 77
 76   WRITE(J03,71)
 77   WRITE(J03,65)NEG,QNEG
      WRITE(J03,555)IV3
      RETURN
      END

      SUBROUTINE RANDU(IX,IY,YFL)
      IY=IX*899
      IF(IY)5,6,6
  5   IY=IY+32767+1
  6   YFL=IY
      YFL=YFL/32767.
      RETURN
      END

      SUBROUTINE GAUSS(IX,S,AM,V)
      A=0.0
      DO 50 I=1,12
      CALL RANDU(IX,IY,Y)
      IX=IY
 50   A=A+Y
      V=(A-6.0)*S+AM
      RETURN
      END
```

Note: When punching long statements in these subroutines it is necessary to use the full field before continuing on a subsequent card.