Appendix C:

TECHNICAL NOTES ON PERFORMING THE TIME SERIES ANALYSES.

Technical Notes On Performing The Univariate Time Series Analysis.

The Programme:

The programme used to run the univariate time series analysis is called TAØ1UNIV. A full listing of TAØ1UNIV is presented in Appendix

The Control Characters:

The control characters appear in the datafile which is appended to TAØ1UNIV, placed immediately before and after the data file title. For example, the control characters for the datafile TAØ1BAT look like this:

```
 2 30 10
TOTAL BATON ROUNDS FIRED 1969-81.
 12120  39
```

The first line of control characters set the time lag (2); the series length (30); and the interval between subsequent series. The second line of control characters set the number of data points on the first card (12); the total number of data points (120); and the variable number (39).

Changing the Control Characters:

The required parameters of any specified time scan can be set by simply editing the two rows of control characters. However, if a longer or shorter number of time points are to be examined, the control characters on lines, 1;2;3;4; 121; & 167 of the programme TAØ1UNIV must also be changed.

Running the Univariate Time Series Analysis:

On the UMRCC CP76 system, the Univariate Time Series Analysis can be run using the simple run job command: RJ TAØ1---,TAØ1UNIR,CP76(P4000,T4,SP), where TAØ1--- is the job name; TAØ1UNIR is the package which merges TAØ1UNIV - the programme and the data; CP76 is the computer label, P4000 is the priority and T4 the required time to run.

Changing Data Files:

A new data file can be appended to the Univariate programme by simply editing TAØ1UNIR which looks like this:-

```
0. ATTACH(UNIV,TAØ1UNIV,ST=S6A)
1. FTN(I=UNIV,SL=0)
2. ATTACH(IRD,TAØ1BAT,ST=S6A)
3. LDSET(MAP=B/ZZZZMP,PRESET=ZERO)
4. LGO(PL=15000)
5. ## ## ## ##
```

TAØ1BAT on line 2 is simply replaced by editing in the new data file and the job run in the normal manner as explained above.

Drawing Out The Correlograms:

The Univariate programme automatically draws out the correlograms with traces plotted for lag.1.,lag 2., and Mean Level 1 & 2. However, the refined plots presented in the text are drawn using the "Simple Plot" package.

Technical Notes On Performing The Bivariate Time Series Analysis.

The Programme:

The programme used to run the bivariate time series analysis is called TAØ1IRISHP. A full listing of TAØ1IRISHP is presented in Appendix

The Control Characters:

The control characters appear in the datafile which is appended to TAØIRISHP, immediately before the first variable file label. For example, the control characters in the data file IRISTA look like this:-

---

3 4 212 348
ASSASINATIONS OF PROTESTANTS IN NORTHERN IRELAND.

In sequence, the first control character (3) identifies the first variable; the second control character (4), the second variable. The next control character (2) sets the time lag, i.e. the number of time points between series. The following control characters set the series length (12); the increment between series (3); and the last series length to be run (48).

Changing The Contol Characters:

The required parameters for any specified time scan between variable pairs can be set by simply editing this row of control characters. However, if a longer or shorter number of time points are to be examined, the relevant parts of the programme TAØ1IRISHP must be altered accordingly.

Running the Bivariate Time Series Analysis:

On the UMRCC CP76 system, the Bivariate Time Series Analysis can be run using the simple run job command: RJ TAØ1 ---,TAØ1BIV,CP76(P4000,T4,SP), where TAØ1--- is the job name; TAØ1BIV is the package which merges the programme TAØ1IRISHP, with the data file e.g. IRISTA; CP76 is the computer label; P4000 is the priority and T4 the required time to run.

Changing Data Files:

A new data file can be appended to the Bivariate programme by simply editing TAØ1BIV which looks like this:-

```
0. ATTACH(IRISHP,TAØ1IRISHP,ST=S6A)
1. FTN(I=IRISHP,SL=0)
2. ATTACH(IRISHD,IRISTA,ST=S6A)
3. LDSET(MAP=B/ZZZZMP,PRESET=0)
4. LGO(PL=15000)
5. ££££S
6. ****
```

IRISTA on line 2 is replaced by editing in the name of the new data file and then the job can be run in the normal manner as described above.

Drawing Out The Correlograms:

The Bivariate programme automatically draws out the correlograms with traces plotted for lag.1., lag 2., Mean levels 1 & 2. However the more refined plots presented in the text were drawn up using the "Simple Plot" package.

Technical Notes On Performing The Multivariate Time Series Analysis.

The Programme:

The programme used to run the multivariate time series analysis is called TAØ1IRISHP5. A full listing of TAØ1IRISHP5 is presented in Appendix

The Control Characters:

The control characters appear in the datafile which is appended to TAØ1IRISHP5, placed immediately before the title of the first variable in the data file.

For example, the control characters for the datafile TAØ1SCORPIO, look like this:-

---

```
 24  5  3138
14138 12 12  2
VEHICLES SEARCHED (IN 100'S) 1970-81.
```

The first line of control characters set the series length (24); the interval between subsequent series (5); the spread between the time lag components t -2, t - 1, t, (3); and the total number of data points in each variable (138).

The second line of control characters set the total number of variables in the datafile (14); the total number of data points in each of these variables (138); the number of data points on the first card (12); the number of data points on subsequent cards (12); and the number of time lags used in the analysis (2).

Changing The Control Characters:

The required parameters for any selected time scan can be set by simply editing the two rows of control characters preceeding the datafile. However, if a greater number of variables than 38 is to be analysed, further instructions are included in the programme TAØ1IRISHP5.

Running The Multivariate Time Series Analysis:

On the UMRCC CP76 system, the Multivariate Time Series Analysis can be run using the simple run job command: RJ TAØ1---,TAØ1RUN30,CP76(P4000,T4,SP) - where TAØ1RUN30 is the package which merges programme and data, TAØ1---, is the jobname, CP76 is the computer label, P4000 the priority & T4 the required run time in seconds.

Changing Data Files:

A new data file can be appended to the Multivariate programme by simply editing TAØ1RUN30 which contains the name of the data file on line 2 , and replacing the old file name with the new datafile to be run.

Drawing Out The Systemograms:

The Multivariate Time Series Programme prints out a list of variable names following the order in which they are organised in the data field. For example, an analysis of the data in file TAØ1SCORPIO, first prints out:-

Vehicles Searched
Occupied Houses Searched
Total Number Of CS Projectiles Fired
Total Baton Rounds Fired
Assassination Of Catholics
Assassination Of Protestants
Internment Of Republicans
Internment Of Loyalists
Number Of Bomb Explosions
Total Number Of Shooting Incidents
Total Number Of Civilians Injured & Wounded
Military Injured & Wounded (Army + UDR)
Military Deaths (Army + UDR)
Number Of Civilians Killed

This sequence is used to form a matrix which is designed to determine the direction of any inferred causal influences between sets of variables. Given that each variable in a two time lag analysis has three possible states, t - 2 ; t - 1 ; & t; each of these states is provided with a unique identifier number in the matrix. For example, Vehicles Searched, the first variable, has states 1,2 & 3. Houses Searched, the second variable, is identified with 4, 5, & 6, and so on. Since the flow of time is from t - 2, to t - 1, to t, the convention used is in reverse order. The pattern is demonstrated in the matrix formed from TAØ1SCORPIO drawn out overleaf.

To determine the direction of causal influence inferred between variables, simply examine their relative precedence in the matrix. All variables with an identifier number in the same column are interactive , have no causal direction inferred since they rise and fall together. A variable in the t - 2 column linked with another variable in either the t - 1 or the t columns is seen as directing the associated variable. Arrowheads from the active to directed variables are used to denote the direction of inferred influence. If the respective variable identifier numbers are two columns away, a double arrowhead is used; if they are only seperated by one column, a single arrow is the convention; whilst interactive variables in the same column are linked by arrowless lines. If a negative correlation emerges, the convention used to denote this is a broken line. The actual level of correlation between related variable pairs is indicated inside a box attached to the line connecting the relevant variables.

The best way of clarifying this procedure is through actually drawing out one of the clusters produced by running a typical multivariate time series analysis. The example worked through below resulted from an analysis of the data in TAØ1SCORPIO, using a series length of 24. The matrix used to determine the direction of inferred causal influence is thus the same one referred to earlier and drawn out overleaf.

The actual printout for this analysis consists of a sequence of clusters for each timeperiod. However, the method is standard for each time period, so only one need be considered below in full.

MATRIX USED TO DETERMINE THE DIRECTION OF INFERRED CAUSAL INFLUENCE.

VARIABLE	t - 2	t - 1	t
Vehicles Searched	3	2	1
Houses Searched	6	5	4
CS Munitions Fired	9	8	7
Baton Rounds Fired	12	11	10
Assassination of Catholics	15	14	13
Assassination of Protestants	18	17	16
Internment of Republicans	21	20	19
Internment of Loyalists	24	23	22
Bomb Explosions	27	26	25
Shooting Incidents	30	29	28
Civilians Injured & Wounded	33	32	31
Military Injured & Wounded	36	35	34
Military Killed	39	38	37
Civilians Killed	42	41	40

The actual output printed for the example timeperiod 76 looks like this:-

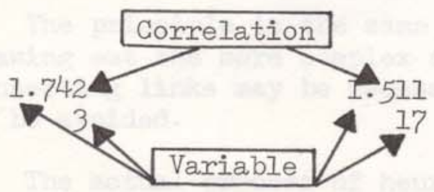
```
OUTPUT FOR TIME SERIES CLUSTER FOR SERIES LENGTH    24

FOR TIME PERIOD        76
MULT   TYPAL START
1      2      1.860        2.860
                 0            0
2      2      1.742        1.511
                 3           17
3      1    - 3.710
                28
4      2     28.702       28.652
                29           29
5      1    -29.674
                25
6      1     25.580
                26
8      2     31.792       40.792
                 0            0
9      7     40.687       31.564   40.601   31.666   40.754   31.701   31.599
                 4            5       27       32       33       41       42
11     2     12.632       34.632
                 0            0
13     2     13.617       37.617
                 0            0
15     2     10.603       18.603
                 0            0
```

Reading from left to right, the MULT and TYPAL columns give the sequence and number of clusters respectively and are not used in the process of drawing out the systemograms. The figures required for creating the systemograms occur as a sequence of blocks, begining under the START heading, and read from left to right. Thus the first line of blocks look like this:-

```
1.860      2.860
   0          0
```

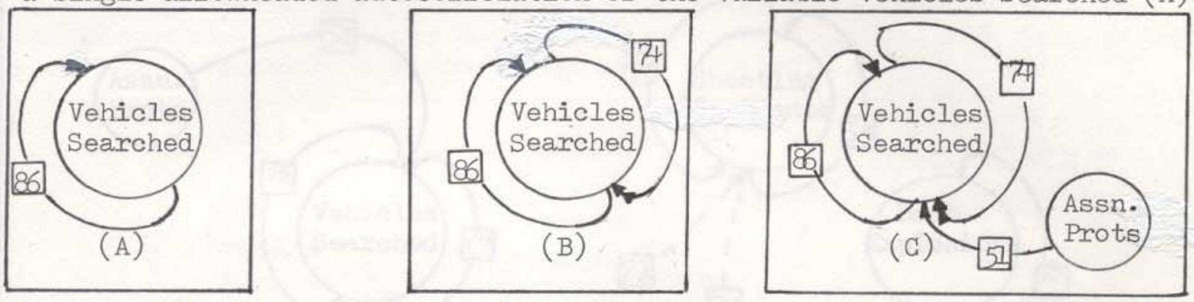
The number before the decimal point (1) refers to the variable - in this case Vehicles Searched. The number after the decimal point - (860) refers to the level of correlation. On the first line of a new cluster sequence, the first two variables, (usually the most highly correlated), emerge with the <u>same</u> level of correlation. These first two blocks of a new sequence of clusters have a zero underneath the correlation figure, whilst subsequent blocks have another variable identifier number. Thus line two looks like this:-



Thus the first block denotes a correlation of .742 between variable identifiers 1 & 3; block two shows a correlation of .511 between variable identifiers 1 & 17.

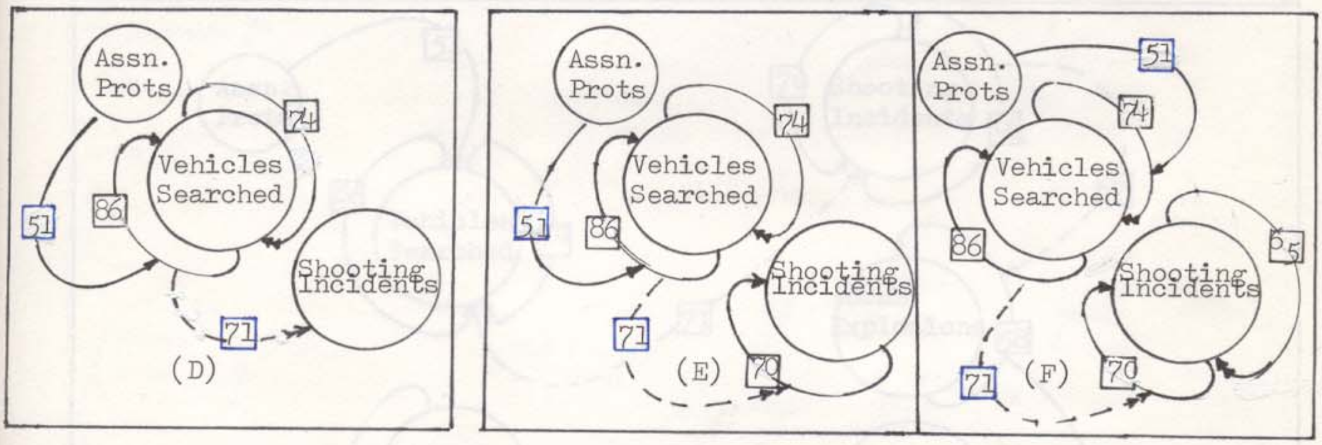
Formally working through the output for timeperiod 76 we have:-

Line 1. Variable identifiers 1 & 2 which have a correlation of .86. Identifiers 1 & 2 both represent Vehicles Searched. The direction of influence is from t - 1 to t, the first block set therefor draws out as a single arrowheaded autocorrelation of the variable Vehicles Searched (A).



Line 2. The next block consists of variable identifiers 1 & 3 which have a correlation of .742. Again, identifiers 1 & 2 represent vehicles searched but this time the flow of influence is from t-2 to t. This result is therefor drawn out as a doublearrowheaded autocorrelation of Vehicles Searched (B). The next block links variable identifiers 1 & 17 by a correlation of .511. Using the matrix above, identifier 17 represents the variable Assassination of Protestants at t - 1; identifier 1 is variable Vehicles Searched at t. The flow of associated influence is therefore from Assassination of Protestants to Vehicles Searched, denoted by a single arrowheaded line. (C)

Line 3. has only 1 block which links variable identifiers 3 & 28 with a negative correlation of .71. Since 3 is in the t - 2 column and 28 lies in the t column, this is drawn as a double arrowheaded broken line pointing from Vehicles Searched to Shooting Incidents. (D)

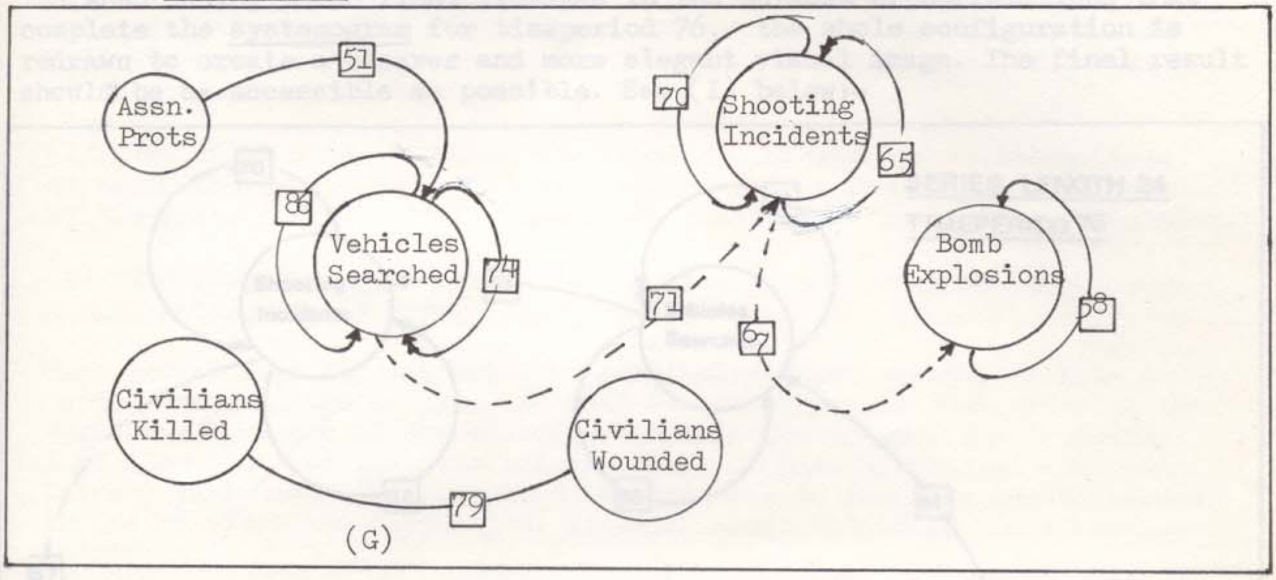


Line 4 adds a single arrowheaded autocorrelation of Shooting Incidents (E), and a double arrowheaded autocorrelation too of levels.702 & .652 respectively.(F)

The principle is the same for all subsequent blocks. However, when drawing out the more complex clusters, considerable remapping of the connecting links may be necessary, if a messy looking descriptive image is to be avoided.

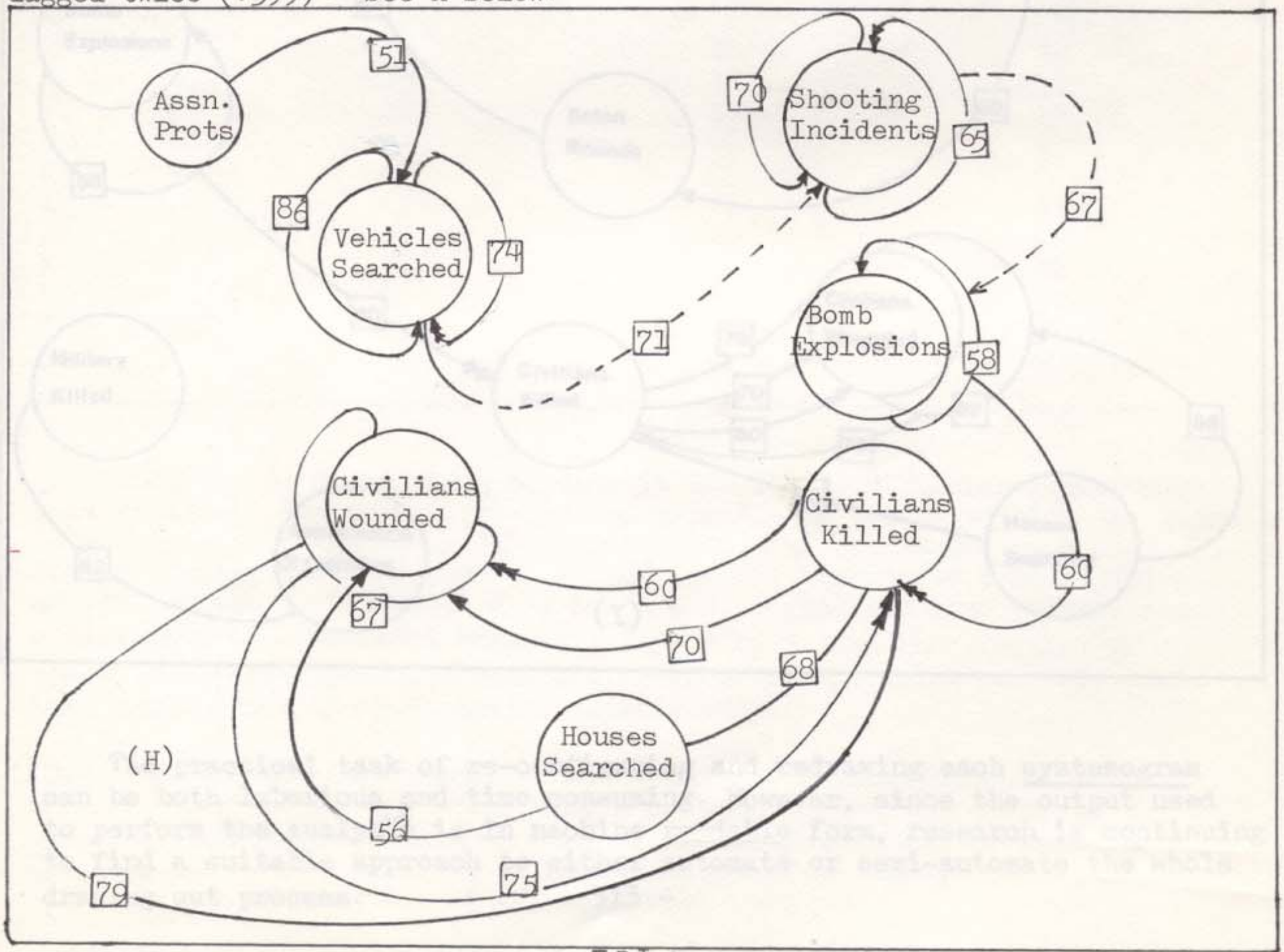
The actual process of heuristically forming the final layout of each systemogram is best illustrated by drawing out the current example to completion.

Line 5 reveals a negatively correlated link from the variable, Shooting Incidents to Bomb Explosions(.674). Line 6 indicates an autocorrelation in the variable Bomb Explosions of (.58). Line 7 describes an interactive correlation between Civilian Wounded & Civilian Killed (792). Drawing out these clusters, the embryonic systemogram structure (G) is obtained - (See below).



(G)

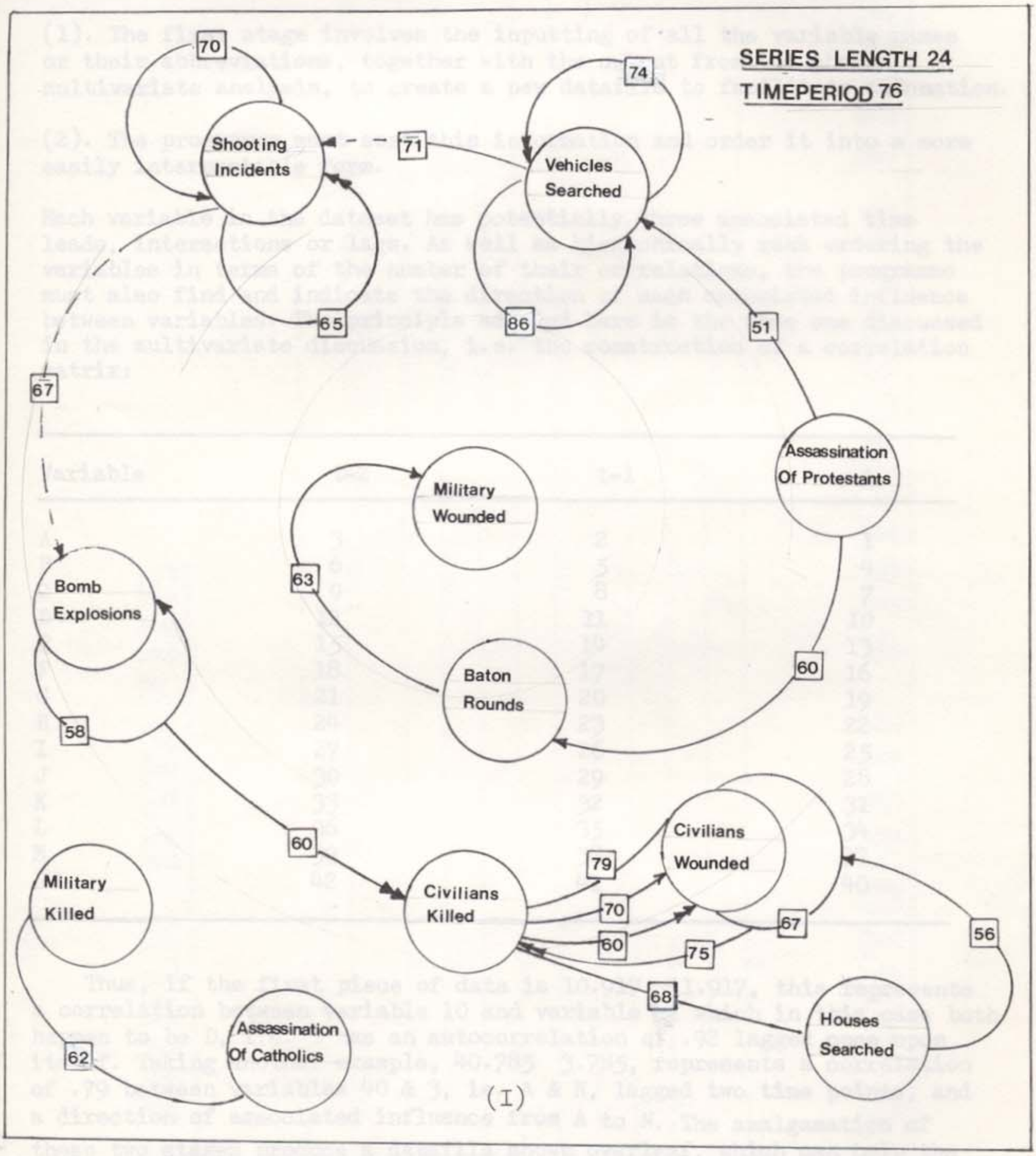
Line 8 expands these last two clusters, with an interactive link between Houses Searched & Civilians Killed (.682); a driving link from Houses Searched to Civilians Wounded (564); a driving link from Bomb Explosions to Civilians Killed(601); an autocorrelation for Civilians Wounded, lagged once (666); a driving link from Civilians Wounded to Civilians Killed (754); and corresponding driving links from Civilians Killed to Civilians Wounded, lagged once (.701) and lagged twice (.599) - See H Below.



(H)

Lines 9,10 & 11 contain blocks which correlate influences from Baton Rounds to Military Wounded(632); an interactive correlation of Assassination Of Catholics and Military Killed(.62); and a correlation from Assassination of Protestants to Baton Rounds Fired(.603).

When adding these final linkages in the network of correlations that complete the systemogram for timeperiod 76, the whole configuration is redrawn to create a clearer and more elegant visual image. The final result should be as accessible as possible. See (I) below:-



(I)

The practical task of re-configuring and redrawing each systemogram can be both laborious and time consuming. However, since the output used to perform the analysis is in machine readable form, research is continuing to find a suitable approach to either automate or semi-automate the whole drawing out process.

AUTOMATING THE SYSTEMOGRAM DRAWING OUT PROCESS.

Although full automation of the systemogram drawing out process is still some time off, during the summer of 1984 work began to produce a programme to semi-automate this task, at the University of Manchester Institute of Science & Technology. Mr. Tim Walker, an undergraduate in UMIST's Department of Computation, has suggested the following approach as part of his final year dissertation. It consists of essentially five stages:-

(1). The first stage involves the inputting of all the variable names or their abbreviations, together with the output from the chosen multivariate analysis, to create a new datafile to facilitate automation.

(2). The programme must sort this information and order it into a more easily interpretable form.

Each variable in the dataset has potentially three associated time leads, interactions or lags. As well as hiearchically rank ordering the variables in terms of the number of their correlations, the programme must also find and indicate the direction of each associated influence between variables. The principle adopted here is the same one discussed in the multivariate discussion, i.e. the construction of a correlation matrix:

Variable	t-2	t-1	t
A	3	2	1
B	6	5	4
C	9	8	7
D	12	11	10
E	15	14	13
F	18	17	16
G	21	20	19
H	24	23	22
I	27	26	25
J	30	29	28
K	33	32	31
L	36	35	34
M	39	38	37
N	42	41	40

Thus, if the first piece of data is 10.917 11.917, this represents a correlation between variable 10 and variable 11 which in this case both happen to be D, i.e. D has an autocorrelation of .92 lagged once upon itself. Taking another example, 40.785 3.785, represents a correlation of .79 between variables 40 & 3, ie. A & N, lagged two time points, and a direction of associated influence from A to N. The amalgamation of these two stages produce a datafile shown overleaf, which can help the hand drawing out of systemograms considerably.

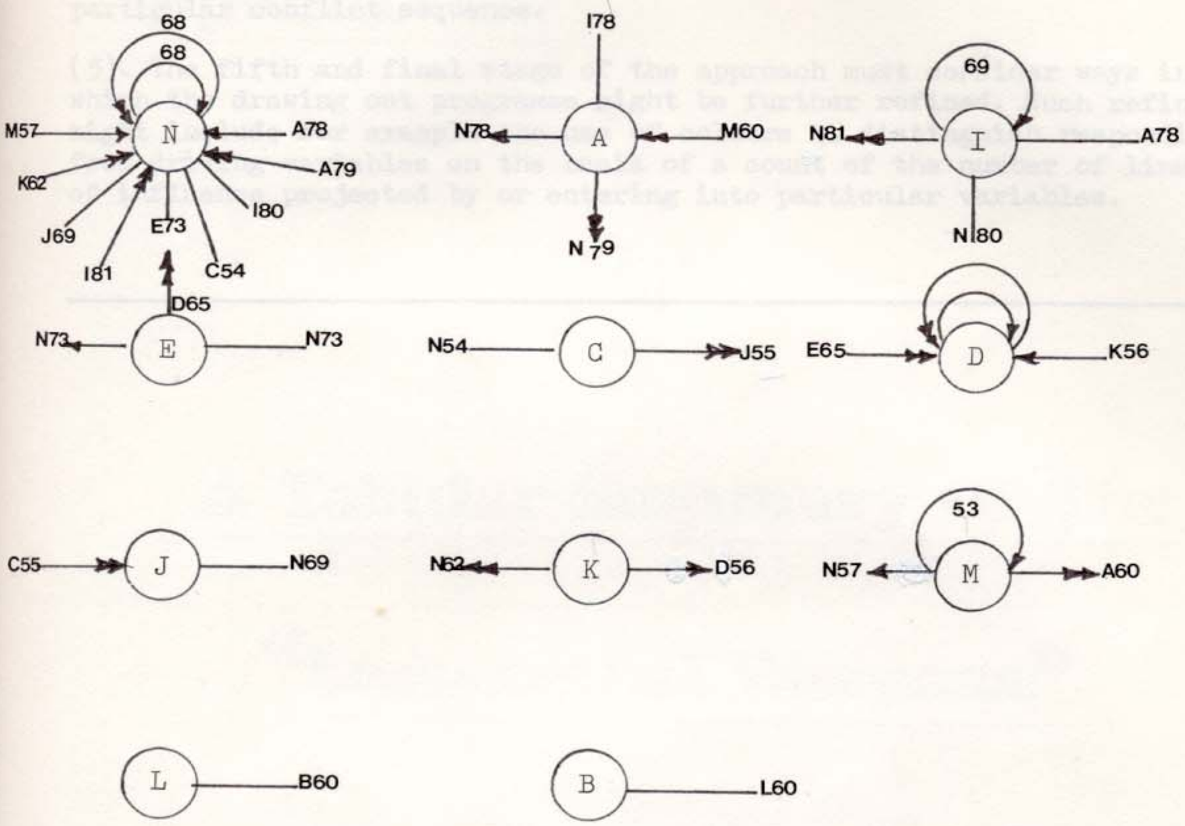
(3). Entries into this datafile have been rank ordered in terms of the numeric order of the variables they correlate to, and according to the number of non-autocorrelations they have. The third stage of automation involves using this datafile to produce a graphic display of all the relationships between all of the variables analysed.

SAMPLE DATAFILE

Name	No. of Correlations	No. of Autocorrelations	No. of non-Autocorrelations	Autocorrelations	Correlations
NN	12	2	10	68NN1 68NN2	78AA1 79AA2 54CC0 73EE0 73EE1 80II0 80II2 69JJ0 62KK2 57MM1
AA	4	0	4		78II0 60MM2 78NN-1 79NN-2
II	4	1	3	69II1	80AA0 80NN0 81NN-2
EE	3	0	3		65DD-2 73NN0 73NN-1
CC	2	0	2		55JJ-2 54NN0
DD	4	2	2	92DD1 83DD2	65EE2 55KK1
JJ	2	0	2		55CC2 69NN0
KK	2	0	2		55DD1 62NN-2
MM	3	1	2	53MM1	60AA-2 57NN-1
LL	1	0	1		60BB0
BB	1	0	1		60LL0

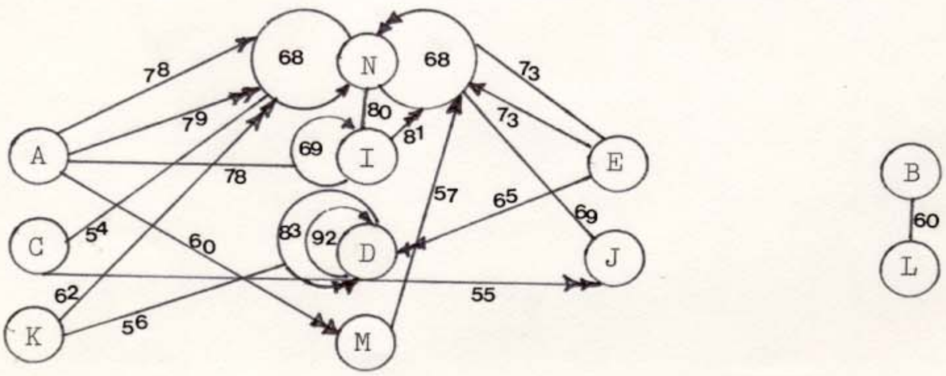
TO AUTOMATE SYSTEMOGRAM PRODUCTION.

The output to be achieved is shown below:-



The arrows shown on the links between variables indicate the direction of associated influence, according to the datafile schema 2,1,0,-1,2. Although there are no negative correlations in this particular systemogram, the same convention of broken line links has also been adopted here.

(4). The fourth and most vital stage is to produce some form of complete systemogram by combining the component parts expressed above. This stage requires a standard set of conventions for deciding the relative positions of the variables and a methodology consisting of a set of standard operating procedures which will, where possible, avoid crossover connecting lines between variables. Using a simple hierachical approach, the systemogram below is produced.



Although this schema clearly shows the relationship between the variables, (it is in fact SL30-Systemogram TP 76), it is visually unacceptable for display purposes. The system does enable a conflict analyst skilled in drawing systemograms, a clear framework from which a more elegant image can be produced. The automatic production of such conflict diagrams will significantly reduce the amount of work required to interpret a particular conflict sequence.

(5). The fifth and final stage of the approach must consider ways in which the drawing out programme might be further refined. Such refinements might include for example the use of colours to distinguish responding from driving variables on the basis of a count of the number of lines of influence projected by or entering into particular variables.