UNITED STATES DEPARTMENT OF THE INTERIOR

GEOLOGICAL SURVEY

User's Guide to Five VAX FORTRAN Programs for Manipulating HYPOINVERSE Summary and Archive Files: SELECT, EXTRACT, SUMLIST, ARCPRINT and FORCON

by
Fred W. Klein

Open File Report 89-313

8/89 version

INTRODUCTION

The HYPOINVERSE earthquake location program (USGS Open File Reports 85-515 and 89-000) produces three types of output files suitable for manipulation by other computer programs: (1) a summary location file containing event data (one line per event), (2) an archive data file containing one line per station with the summary line as an event header, and (3) a magnitude data file containing a summary event header and detailed magnitude data on stations reporting amplitudes or coda durations. The archive file contains essentially all of the raw data and calculated results necessary to reprocess or conduct other studies of the earthquakes. Archive files may also contain "shadow" records which begin with a "$" and follow every line in the file. Shadow records contain additional data such as coda envelope amplitudes and seismogram access information.

The SELECT program selects subsets of earthquake summary files. It is a powerful program with a command interface and many options. SELECT applies any or all of geographic region, depth, magnitude, error, time and other numeric or alphabetic criteria in selecting events. It operates on multiple input or output files and in a variety of formats.

The EXTRACT program extracts subsets of archive files. The earthquakes to extract are in a summary file, for example one produced by SELECT.

The SUMLIST program reformats summary files into a very readable form suitable for publication in an earthquake bulletin.

The ARCPRINT program reformats archive files into a very readable form similar to the print output of the original HYPOINVERSE run.

The FORCON program converts summary files from one format to another. The formats supported are HYPOINVERSE, HYPO71 and a condensed binary format.

In Menlo Park, all 5 programs reside on the VAX 785 in WE:[KLEIN.SUM] and the Seismology Branch VAX 750 in WE:[KLEIN]. You may use them conveniently by putting lines like:

```
$ SELECT :== RUN WE:[KLEIN.SUM]SELECT.EXE
$ EXTRACT :== RUN WE:[KLEIN.SUM]EXTRACT.EXE    etc.
```

in your LOGIN.COM file, then run the program by typing SELECT, EXTRACT, etc.

SELECT

INTRODUCTION TO THE SELECT PROGRAM

The SELECT program reads summary files (one line per event) written by an earthquake location program, and produces similar files for which events meet some limiting criteria. SELECT reads every record in a file to find the desired events.

SELECT features multiple input files which are read sequentially. If the input data is in chronological order, you may also get multiple output files, each containing a certain time period. Epicenter selection may be done by five methods: (1) a latitude - longitude box, (2) containment within a closed polygon whose verticies are specified, (3) by one or more predefined regions in Hawaii or Northern California, (4) a series of latitude - longitude - depth boxes, or (5) epicentral distance from a geographic point. If more than one epicentral test is defined, the event must pass each test. Further selection may be by depth, time, amplitude and/or duration magnitude, location error, RMS travel time residual, number of stations and number of first motions. Selection may also be on any real number in the data record. Selection or rejection may also be done on the presence of one or more 1-3 letter remarks at one or two preset places on the data records. It is possible to output both selected and rejected events and of course rerun SELECT on its own output. Thus selection of events based on complex and/or criteria is thus possible.

SELECT is command driven. Commands are 3-letters long. A command line may set parameters or names if these follow the command on the same line. If no parameters appear on the command line, SELECT will prompt for them. You then will see the current parameter and either default to it (by pressing return) or selectively redefine it. Default values are set for all parameters and you need issue commands only for the parameters you want to change. Parameters supplied on a command line are in free format: they are separated by spaces or commas, and character strings must be enclosed within apostrophes. For example, the input file is set by a command like FIL 'INFILE.SUM'. Complete rules for free format input are given in the appendix.

Commands fall into three general types: (1) define numerical or other parameters used for subsequent selection, (2) take some action such as selecting events using the current files and criteria, and (3) invoke conveniences such as showing you a brief definition of commands, listing the input filenames or executing an operating system (DCL) command from within SELECT.

Commands may be executed either by typing them in (with or without supplying parameters on the command line), or by executing them from a disk file. The commands in a control file are executed as if they were typed. If a file called "SELINST." is in your directory, it is read as a command file when SELECT starts: it thus may be used to set up your own defaults. If the commands in the file have parameters, the file executes by itself. If the commands in the file do not have parameters, SELECT prompts and waits for you to input the parameters, the resumes execution of the control file.

SOME SIMPLE EXAMPLES OF SELECT CONTROL FILES

```
1) FIL 'INPUT.SUM'                  Set the input file name.
   OUT 'OUTPUT.SUM'                 Set the output file name.
   LAT 19 0  19 30                  Set the min & max latitude.
   LON 154 50  155 20               Set the min & max longitude.
   DEP 5 13                         Set the depth range.
   SEL                              Select earthquakes.


2) FIL 'INPUT.SUM'
   OUT 'OUTPUT.SUM'
   REG 'B'  'VERTICIES.'            Read the verticies of a polygon labeled B
                                    from the file VERTICIES.
   ERH 5                            Set the maximum horizontal error to 5 km.
   SEL


3) SEQ '[KLEIN]' '.SUM' 82 84       Set the 3 input files [KLEIN]82.SUM
                                    through [KLEIN]84.SUM.
   OUT 'PERIOD1.SUM'                Set the first output file.
   OFL 'PERIOD2.SUM' 82 12 2 0 0    Set 2nd output file & its beginning date.
   OFL 'PERIOD3.SUM' 83 5 6 0 0     Set 3rd output file & its beginning date.
   SEL                              Split the file into 3 sequential files.
```

This example selects events which are either between 120 and 121 deg longitude or 36 and 37 deg latitude.

```
4) FIL 'INPUT.SUM'
   OUT 'OUT1.SUM'                   Output file for first selected events.
   NIX 'REJECT.SUM'                 Output file for rejected events.
   LON 120 0 121 0                  Separate events of 120 deg lon from all
                                    the other longitudes.
   SEL
   FIL 'REJECT.SUM'                 Select from events previously rejected.
   OUT 'OUT2.SUM'                   Output file for second event group.
   NIX 'NONE'                       Stop output of rejected events.
   LON 0 0 180 0                    Select all longitudes.
   LAT 36 0 37 0                    Select this latitude range from remaining
                                    events.

   SEL                              The aggregate of files OUT1.SUM and
                                    OUT2.SUM now contain events in the "+"
                                    shaped region desired.
```

EXPLANATION OF SELECT COMMANDS
(The examples show the defaults)

---- INPUT FILES AND DATA FORMAT ----------

FIL Set the input filename. This will be either the only input file, or the first of a sequence of filenames if others are set with the IFL command.
EXAMPLE: FIL 'IN.SUM'

IFL Set the file number and name of the second or later input file. The number must be in the range 2-30.
EXAMPLE: IFL 2 'FILE2.' (But the default is for only one input file.)

SEQ Set a whole sequence of input files to be of the form <DIR>MM<EXT> through <DIR>NN<EXT>. The files are read sequentially. <DIR> is the character string for the first part of the filename including an optional pathname, MM and NN are integers such that MM $\leq$ NN, and <EXT> is the character string for the last part of the filename including the extension.
EXAMPLE: SEQ '[KLEIN]E' '.SUM' 1982 1983
establishes the two input files [KLEIN]E1982.SUM and [KLEIN]E1983.SUM.

FOR Set the summary format type. Use 1 for HYPOINVERSE, 2 for HYPO71, and 3 for Fred Klein's binary format. The binary format contains date, time, hypocenter, magnitude, horizontal and vertical errors, the RMS residual and number of readings used in the solution.
EXAMPLE: FOR 1

LIS List the current input files at your terminal.

---- OUTPUT FILES ----------

OUT Set the output filename. This will be either the only output file, or the first of a sequence if others are set with the OFL command. Use OUT 'NONE' to omit a selected file, as when writing only a reject file or obtaining event statistics.
EXAMPLE: OUT 'OUT.SUM'

OFL Set the second or later output filename, and the date on which output will begin. A series of OFL commands may be given in any order, and are automatically arranged in chronological order by the dates given. The first output filename must be set with the OUT command and receives any data before the beginning date of the second output file. All input data must be in chronological order. Dates are given by the five integers year, month, day, hour and minute. There is a maximum of 30 output files.

EXAMPLE: OFL 'OUT2.SUM' 82 6 1 12 30 (Default is only one output file)

CUT Remove an output file from the list by giving its name.
EXAMPLE: CUT 'OUT2.SUM'

KIL Cancel the entire list of output files except the first.

SEE Display the output files and their beginning dates on your terminal.

NIX Specify the filename for output of rejected events. The NIX command plus the OUT command let you split a file into two populations. Use a filename of NONE to cancel output of rejected events.
EXAMPLE: NIX 'REJECT.SUM' (The default is for no reject output)

CNT Set logical flags (T or F) to turn on the tabulation of count statistics for input and output files separately. Also supply the name of the file to write statistics to. If the filename is omitted, the name SELSTATS. is used. For each input and output file the following are given:

filename,
date and time of first and last event in the file,
number of events in the file (always written to the screen anyway),
counts of events in 100 intervals 1.0 in size for:
  depth
  horizontal error
  vertical error
  number of stations
  geographic regions (classes)
counts of events in 100 intervals 0.1 in size for:
  duration magnitude
  amplitude magnitude
  depth
  horizontal error
  vertical error
counts of events in 100 intervals 0.02 in size for:
  RMS residual

EXAMPLE: CNT F F 'SELSTATS.'

---- GEOGRAPHIC SELECTION CRITERIA ----------

Geographic selection may be made in four different ways: (1) Simple latitude, longitude and depth limits may be set with the LAT, LON and DEP commands. (2) The verticies (lat and lon) of a polygon may be set with the REG command. (3) Predefined geographic classes may be selected with the NET and CLS commands. The QPLOT documentation and a series of maps available on request show the 30 Hawaiian classes and 103 northern California classes. Classes for northeren California are depth independent; those for Hawaii include depth as a determining factor. (4) A series of lat-lon-depth boxes may be set with the BOX command such that an event will be selected if in any of the boxes.

Only one of methods 1, 2, or 4 may be used at any one time depending on which type of command was given most recently. The geographic selection by class number is separate from and in addition to the other possibilities: an event is selected if it is both in a requested class and within the specified box or polygon. The startup default is for no geographic selection.

BOX Set the number of latitude-longitude-depth boxes such that an event is selected if it is in any box. There may be from 1-10 boxes. IMPORTANT: if the number of boxes is more than 1, the BOX command must be given before the LAT, LON and DEP commands. These commands will require the box number as the first parameter if the number of boxes is more than 1, but will not need a box number if there is only 1 box.
EXAMPLE: BOX 1

LAT Set minimum and maximum latitude, each in degrees and minutes (positive north). If the number of boxes is more than 1, precede the min and max latitudes by the box number you are defining.
EXAMPLES: LAT -90 0 90 0 (NBOX = 1) or LAT 2 20 30 22 0 (NBOX > 1)

LON Set minimum and maximum longitude, each in degrees and minutes (positive west). If the number of boxes is more than 1, precede the min and max longitudes by the box number you are defining.
EXAMPLES: LON -180 0 180 0 (NBOX = 1) or LON 2 119 0 120 30 (NBOX > 1)

DEP Set minimum and maximum depth. If the number of boxes is more than 1, precede the min and max depth by the box number you are defining.
EXAMPLES: DEP 0 999 (NBOX = 1) or DEP 2 1.2 9.8 (NBOX > 1)

REG Set the 1-letter label and filename in which to read the verticies of a polygon. The polygon may be concave. An earthquake will be selected if in the polygon. Vertex files may be easily digitized on a map plot by the QPLOT program using its CURS command. Several rules must be followed by the vertex file:

1) Several polygons may be in each file, but each polygon must have a unique 1-letter label. Each vertex must be labeled. The 1-letter label given in the REG command will be searched for in the file and an error given if it is not found.

2) The format of the file is one vertex per line, with a 1-letter polygon label, latitude (deg & min) and longitude (deg & min) in (A1, 4X, I2, 1X, F5.2, 1X, I3, 1X, F5.2) format. This is QPLOT's cursor output format.

3) The polygon may be concave (interior angles more then 180 degrees), but must simply connected (no figure 8's).

4) The last vertex specified is assumed identical to the first, thus N+1 points specify an N-sided polygon. The last point is redundant but must be present in the file. This so the same file may be used by QPLOT for plotting the complete polygon with closure back to the first point.

5) At present, the polygon may have no more than 14 verticies.

EXAMPLE: REG 'A' 'VERTICIES.DAT'

NET Network number for determining earthquake classes (see also CLS). The net number is not used unless classes are being selected. The nets defined at present are 1=Hawaii, 2=Northern California.
EXAMPLE: NET 1

CLS Number of earthquake classes to select, and list of class numbers. The classes are predefined within SELECT, but are identical with those used by QPLOT and HYPOINVERSE. An event is selected if in any of the listed classes. Specifying the number of classes as 0 selects all classes.
EXAMPLE: CLS 0 (the default) or CLS 3 1 2 3

DIS Set a geographic point and range of epicentral distances from it of events to select. Give the latitude (degrees and minutes) and longitude (degrees and minutes) of the point. Also specify the minimum and maximum epicentral distances from this point in km.
EXAMPLE: DIS 37 0 122 0, 0 999

---- MISCELLANEOUS SELECTION CRITERIA ----------

ERH Set maximum horizontal error.
EXAMPLE: ERH 100

ERZ Set maximum vertical error.
EXAMPLE: ERZ 100

MAG Set minimum and maximum magnitude.
EXAMPLE: MAG 0 9

TYP Set type of magnitude to select on (HYPOINVERSE format only). The codes are: 1=average of amplitude and duration magnitudes; 2=duration mag; 3=amplitude mag; and 4=greater of amplitude or duration mag. Magnitudes of 0 are treated as undefined and are not averaged.
EXAMPLE: TYP 1

DAT Set beginning and ending date, both as yr, mon, day, hr and min. If either or both years are set to 0, no date cutoff is used.
EXAMPLES: DAT 10*0 (the default), or
DAT 1 1 1 0 0, 84 1 23 12 0 selects all events before the date given.

DEP Set minimum and maximum depth.
EXAMPLE: DEP 0 999

RMS Set maximum RMS travel time residual.
EXAMPLE: RMS 100

NUM Set minimum number of stations used in solution.
EXAMPLE: MIN 0

NFM Set minimum number of first motions (HYPOINVERSE only).
EXAMPLE: NFM 0

RAN Set flag to select a user-defined real number field (T or F). Also give format for reading and the minimum and maximum values.
EXAMPLES: RAN F (the default) or RAN T '(T81,F4.2)' 0 5

---- COMMANDS TO SELECT OR REJECT 1-3 LETTER REMARKS ----------

All remarks may be 1, 2 or 3 letters long. To use this feature you must give the COL command and one or both of the RMK and NOR commands.

COL Set the number of remarks to read from the data and the format string for reading the remark(s). The format could look like (Tn,Am) where n is the first column of the remark to be read and m is the length of the remark.

EXAMPLES: COL 2 '(T77,2A1)' (the default reads the two auxiliary remarks in the HYPOINVERSE format where qualifiers such as quarry blasts and felt earthquakes are indicated), or COL 1 '(A2)'

RMK Set the number of remarks to select on and give the list of remarks. An event will be selected if it contains any one of the remarks you define with this command. If the number of remarks is 0, no remark selection takes place. There is a maximum of 10 remarks.
EXAMPLES: RMK 0 (the default), or RMK 2 'A' 'B' (selects A & B events).

NOR Set the number of remarks to reject and give the list of remarks. An event will be rejected if it contains any one of the remarks you define with this command. If the number of remarks is 0, no remark rejection takes place. There is a maximum of 10 remarks.
EXAMPLES: NOR 0 (the default), or NOR 2 'C' 'D' (rejects C & D events).

---- COMMANDS WHICH TAKE ACTION ----------

SEL Select earthquakes using the current files and parameters. Several selects may be run in one session. If the same output filename is used twice, a new version is created without destroying the old.

STO Stops the program.

---- CONVENIENCE AND CONTROL COMMANDS ----------

HEL Gives a brief listing of the basic commands on the screen.

MOR Gives a brief listing of more advanced commands on the screen.

LIS Lists the current input files on the terminal.

SEE Displays the current output files and dates on the terminal.

SHO Displays some of the basic selection parameters on the terminal. All parameters may also be examined without changing them by typing the command and defaulting to the current values by just pressing the return key.

SAV Give the filename into which commands specifying all current parameters are to be saved. This preserves the current state of the program in a file which may be loaded back with the HOP command. The default is the last file SAVed or HOPed to.
EXAMPLE: SAV 'SET1.SEL'

HOP Transfer control to a command file whose name you give. The file may be built with the editor or previously SAVed. If a command file called "SELINST." is present in your directory, it is executed automatically when SELECT starts.
EXAMPLE: HOP 'SET1.SEL'

@ Note: @filename is equivalent to HOP 'filename'

SPA Spawn a DCL (operating system) command by giving it as a string. When the command finishes, control returns to SELECT where you left off.
EXAMPLES: SPA 'EDT SELINST.' or SPA 'DIR *.SUM'

Note: #command is equivalent to SPA 'command'

* Any command line beginning with a * or which is all blank is treated as a comment and is ignored.

EXTRACT

HOW TO USE THE EXTRACT PROGRAM

EXTRACT is much simpler than SELECT. It prompts for the information it needs rather than being command driven. EXTRACT operates only on archive and magnitude output data in HYPOINVERSE format, and all events must be in chronological order. EXTRACT handles archive files with or without shadow records. Only the origin time is used to match events, and thus only the date and time fields are read from summary records. It is therefore possible to edit by hand a list of times of events to extract.

EXTRACT first asks for the "tolerance in seconds for matching archive and summary events". All events in the archive file which are closer than this specified time to events in the summary file will be written to the output file. If a second or third event in the archive file matches one in the summary file, a warning will be given at the terminal. If an archived event can't be found to match one in the summary file, a warning message will be given. If the summary and archive files were produced in the same HYPOINVERSE location run, the origin times should match exactly and a small tolerance, say 0.03 second, is appropriate. If the summary times are approximate, as when building a list by hand, a tolerance of 60 or 100 seconds may be needed to be sure of getting all events, but may occaisionally extract an extra event.

Next the program asks for the names of the summary and archive input files, whether the archive file has shadow records, and then the archive output file and whether you want it to hav eshadow records. You may thus use EXTRACT to filter out shadow records. Only warning messages should appear while the program is reading the files. Before it exits, EXTRACT advises you of the number of events extracted and file they were written to.

SUMLIST

HOW TO USE THE SUMLIST PROGRAM

SUMLIST reformats a HYPOINVERSE summary file into a legible form suitable for printing in a bulletin. It offers some flexibility in the number of lines per printed page and number of characters across the page. SUMLIST prints events in groups of 5 lines with a blank line between. It first asks for the "number of printed groups of 5 events". Ten is good on a line printer, and 11 or 13 work with a laser printer running at 8 lines per inch. The program next asks for the "number of characters per line". Shortening the line from its maximum value of 111 characters cuts off the least essential information about each event and fits the data onto narrower pages. SUMLIST also reads HYPOINVERSE archive and magnitude data files, recognizing only the summary header records. The following is a guide to what is printed and at what character position:

Last character position	Abbrev.	
20		Origin time (year, month, day, hour, min, sec).
45		Depth
48	NRD	Number of readings used in solution (P & S).
50	NS	Number of S readings used in solution.
54	RMS	Root mean square travel time residual.
62	ERH,ERZ	Horizontal and vertical errors.
65	NFM	Number of first motions.
71	REMKS	3-letter region code and auxiliary remarks.
75	DUR MAG	Duration or coda magnitude (median).
78	NDU	Number of stations used in coda magnitude.
82	DUR RMS	RMS of coda magnitudes from median.
86	AZ GAP	Lagest azimuthal gap between stations.
89	MIN DS	Distance to closest station.
93	CRU MOD	Crust model code.
97	SOURCE	Data source codes (P= P&S picks, F= dur mag, X= amp mag
101	AMP MAG	Amplitude magnitude (median).
104	NA	Number of stations used in amplitude magnitude.
109	AMP RMS	RMS of amplitude magnitudes from median.
111	NV	Number of valid (assigned weights >0) P & S readings.

Use this table as a guide for deciding how many characters to print. For example, if you don't have amplitude magnitudes but want the minimum distance, use a length of 89 characters in SUMLIST. If you are using a larger letter size on the laser printer, you may need to limit lines to 75 characters to keep all of the data on one line. I've tested these print commands:

Line printer: 10 groups, any number of characters, VAX command PRINT.

Laser printer:

Vertical page:
(7.167 lines/inch) 11 groups, 89 characters, VAX command
ASCPS/TOP=48/BOT=0/LEFT=24
(9 lines/inch) 13 groups, 111 characters, VAX command
ASCPS/SIZE=8/TOP=48/LEFT=24
(8 lines/inch) 13 groups, up to 99 characters, VAX command LW/LIST.
(6 lines/inch) 9 groups, 75 characters, VAX command LW/LET.

Horizontal page:
(7.167 lines/inch) 8 groups, 111 characters, VAX command
ASCPS/LAND/1OP=48/BOT=0
(8 lines/inch) 8 groups, up to 132 characters, Unix command enscript -r -fCourier9 filename.

ARCPRINT

ARCPRINT reads all the HYPOINVERSE output data in an archive file and rewrites it in a format similar to that of the original HYPOINVERSE print output. The HYPOINVERSE documentation (USGS Open File Report 89-314) explains the various fields of event and station data. The output of ARCPRINT does not include the iteration history, eigenvalue or covariance data because these are not written to the archive file. In addition, the identification number appears at the end of an event's station listing.

ARCPRINT runs interactively. The program prompts for the name of the input archive file and output print file. It ignores shadow records beginning with a "$".

FORCON

FORCON converts a summary file's format from HYPOINVERSE, HYPO71 or Fred Klein's binary format to any one of these. The number of bytes taken up by records in these three formats are 113, 80 and 26. In addition, FORCON can read (but does not write) the CAT.BIN binary files produced by the CUSP system and write the data in another format. The 62-byte CAT.BIN format contains fields for ID numbers, a flag mask, quality, and source net and device that are not used by the other formats. The two binary formats may be plotted directly by QPLOT. The binary formats contain date, time, hypocenter, magnitude, horizontal and vertical errors, the RMS residual and number of readings used in the solution. If converting from binary or HYPO71 to HYPOINVERSE format, fields undefined by the smaller formats will take on values of zero in the more complete formats.

The FORCON program is run interactively and prompts for what it needs. It first asks for the formats of the input and output files, then the input and output filenames. After one format conversion is complete, FORCON asks the same questions again with the previous responses as defaults if you press the return key. You may thus efficiently convert several files in one session. Use ctrl-Y to stop the program.