

CHAPTER 8.

DATA ANALYSIS

1. procedure histo (A, B, c, d); array A, B; real c, d;  
will update a histogram defined by the one-dimensional  
arrays A and B according to the observation c with the weight d.  
A [i] is increased by d, where i is the smallest integer such  
that  $c \leq B [i]$ . It is assumed that the length of A is one  
greater than that of B. The last element of A corresponds  
to those observations which are greater than all elements  
of B. The procedure will accept parameters of any combi-  
nation of real and integer types.
2. procedure accum (a, b, c, d); value d; real a, b;  
begin a := a + c X (time-b);  
    b := time; c := c + d end;

This procedure serves to accumulate the system time integral  $a = \int c dt$ . The parameters a, b, c are called by name. The corresponding actual parameters must be variables. a and b must be of type real, c and d can be real or integer.

Statements of the form

    accum(ax,tx,x,dx)

will update the variable x by adding dx, as well as the

integral  $ax$ .  $dx$  can be any arithmetic expression. The variable  $tx$  serves to record the system time at which  $ax$  and  $x$  were last updated. During the period of integration no updating of  $ax$ ,  $tx$ , and  $x$  should take place, except through the procedure. The initial value of  $tx$  should be the system time at which the integration is started, say  $t_1$ . The variable  $ax$  can be updated to the current value of the integral at any time by the statement

`accum(ax,tx,x,0).`

The average value of  $x$  is equal to  $ax/(time-t_1)$ .

Since the procedure refers to the current system time, it will only be available within a SIMULA block.

3. procedure `hprint (Y,X,LB,UB,YO,YS);`

$Y$  and  $X$  are arrays of type integer or real.  $LB,UB,YO,YS$  are simple variables of type integer or real.

A call to this procedure will give a printout of a histogram. The columns are printed as 3 lines of asterisks (\*).

The parameters have the following meaning:

$Y$	Column length
$X$	Upper interval limits (See procedure <code>histo</code> )
$LB$	Index of first $Y$ -column to be printed
$UB$	Index of last $Y$ -column to be printed
$YO$	Starting point of $Y$ -axis
$YS$	Scale factor for column length

The number of asterisks in the lines of the K-th (column) is

$$\frac{Y(k) - Y_0}{Y_S}$$

rounded to the nearest integer.

Example:  $Y_0 = 500, Y_S = 5$

```
-----  
***  
***      512.5 ≤ Y [k] < 517.5  
***  
-----  
                Y < 502.5  
-----
```

Parameters may be omitted by

1. Shortening the parameter list by cutting out from the right, or
2. By using as actual parameter a variable of another type than the one specified above.

Examples: Boolean Q

hprint (Y,Q,Q,UB) : X, LB, Y<sub>0</sub> and Y<sub>S</sub> omitted

hprint (Y) : X, LB, UB, Y<sub>0</sub> and Y<sub>S</sub> omitted

hprint (Y,Q,Q,Q,Q): X, LB, UB, Y<sub>0</sub> and Y<sub>S</sub> omitted

The effects of omitting parameters are:

1. X : The intervals are instead given by the subscripts of Y.
2. LB : The printout of Y [k] -columns will start with the first Y(k) > 0.

3. UB : The printout of Y[k]-columns will stop after the last Y(k) > 0
4. YO : YO is put equal to 0
5. YS : YS is automatically adjusted to make max (Y[k]) consist of (3) columns of 100 asterisks.

The first and last columns of the Y-array will contain observations out of range. If these are empty, the mean (interval or subscript) value is printed after the histogram together with its standard deviation and the sum of the Y-columns.

At the extreme right of the printout the value of Y[k] is printed as a floating number together with the percentage which Y[k] represents of the sum of all Y[k]-s.

4. procedure arrinfo (text, code, A, COLMEAN, COLSD, COLMAX, COLMIN, COLRANGE, COLCORR, b);

text is a string, code and b integer, A is a one- or two-dimensional real array, COLMEAN, COLSD, COLMAX, COLMIN and COLRANGE are one dimensional arrays of type real. COLCORR is a two dimensional array of type real.

A call on arrinfo will compute and print the user's choice among the following items:

- 1) The mean of the elements in each column of A.
- 2) The standard deviation from the mean of the elements in each column of A.
- 3) The maximum value of the elements in each column of A.
- 4) The minimum value of the elements in each column of A.
- 5) The range of the elements in each column of A.
- 6) The correlation coefficients between the columns of A.

The parameters have the following meaning:

text : heading to precede any printout

code : integer code specifying which of the six different items should be printed. The code is an additive binary code defined as follows:

32 : print COLMEAN (item 1)  
16 : print COLSD (item 2)  
8 : print COLMAX (item 3)  
4 : print COLMIN (item 4)  
2 : print COLRANGE(item 5)  
1 : print COLCORR (item 6)

A : the input array for which information should be computed.

COLMEAN : one-dimensional output array for item 1.

COLSD : one-dimensional output array for item 2.

COLMAX : one-dimensional output array for item 3.

COLMIN : one-dimensional output array for item 4.

COLRANGE: one-dimensional output array for item 5.

COLCORR : two-dimensional output array for item 6.

b : the number of columns to be considered by the subroutine.

Parameters may be omitted as follows:

- 1) Shortening the parameter list from the right.
- 2) Shortening the parameter list from the left. If the parameter code is not present, a zero value will be assumed.

NOTE: the input array A must be present.

- 3) Using the integer constant 0 in place of any one of the six output arrays.

If COLMEAN is present, then COLSD must also be present.

Computation and printout is selected as follows:

- 1) If the parameter code indicates that an item should be printed or if the corresponding output array has not been omitted, then the item will be computed.
- 2) Printout will be given only for those items indicated by the parameter code.

The parameter A may be a one-dimensional array. In this case, COLCORR is undefined.

The following conditions will cause a run-time error message:

1. Each column of A contains only one element
2. The first parameter is a string and the second is not an integer.
3. The output parameters have incorrect type, dimension or length (except case integer 0).

Examples:

arrinfo ('ARRAY INFORMATION', 62, ABC, 0, 0, MA, MI, RA, CO);  
will cause the values of item 1,2,3,4 and 5 to be printed  
while the values of items 3,4,5 and 6 are returned to the user  
program through the array parameters MA, MI, RA and CO.

arrinfo (ABC, 0, 0, MA, MI, RA, CO);

As preceding example, but no printouts. This call is equivalent to

arrinfo ('', 0, ABC, 0, 0, MA, MI, RA, CO);

5. Off-line plotting facilities

Software have been written for the Norwegian Computing Center by Björn Kirkerud to produce paper tapes from SIMULA acceptable for off-line drawing by the KINGMATIC 1215 automatic drawing machine (produced by Kongsberg Våpenfabrikk, Kongsberg, Norway.)

A short summary of some of the subroutines available are given below:

1. Basic Procedures

DEFPAGE	is used to define an origin, the scale and the size of the drawing.
XPEN and YPEN	are functions defining the present X or Y coordinate of the pen.
MOVE TO	move the pen without drawing.
DLINE	draw a straight line.
DCIRK	draw a circle.
DOTON, DOTOFF	are used to control dotting of figures.
STARTPLOT, STOPPLOT	is used to initialize and stop the drawing.

## 2. High level procedures

The "write" subroutine will accept "PLOTTER" as device. The text of the edited lines will be drawn instead of printed. The size of the text and the angle between the X-axis of the drawing machine and the text may be determined by the user. In addition to the symbols usually found on a high speed printer, special character sets containing small letters and foreign (Greek, Russian) alphabets will be provided.

The procedures SAVEPLOT, UNSAVEPLOT and DRAWSAVED form a triple that may be used to construct a library of pre-defined drawings that may be inserted in an actual drawing. The pre-defined drawing may be scaled to the required size, turned an angle and reflected.

The procedures XAKSE and YAKSE will draw axes for a coordinate system. More than one coordinate system may be used in a drawing.

LKURV1 and GKURV1 are used to draw curves. LKURV1 will use straight lines to connect successive points on the curve, while GKURV1 will use a parabolas.

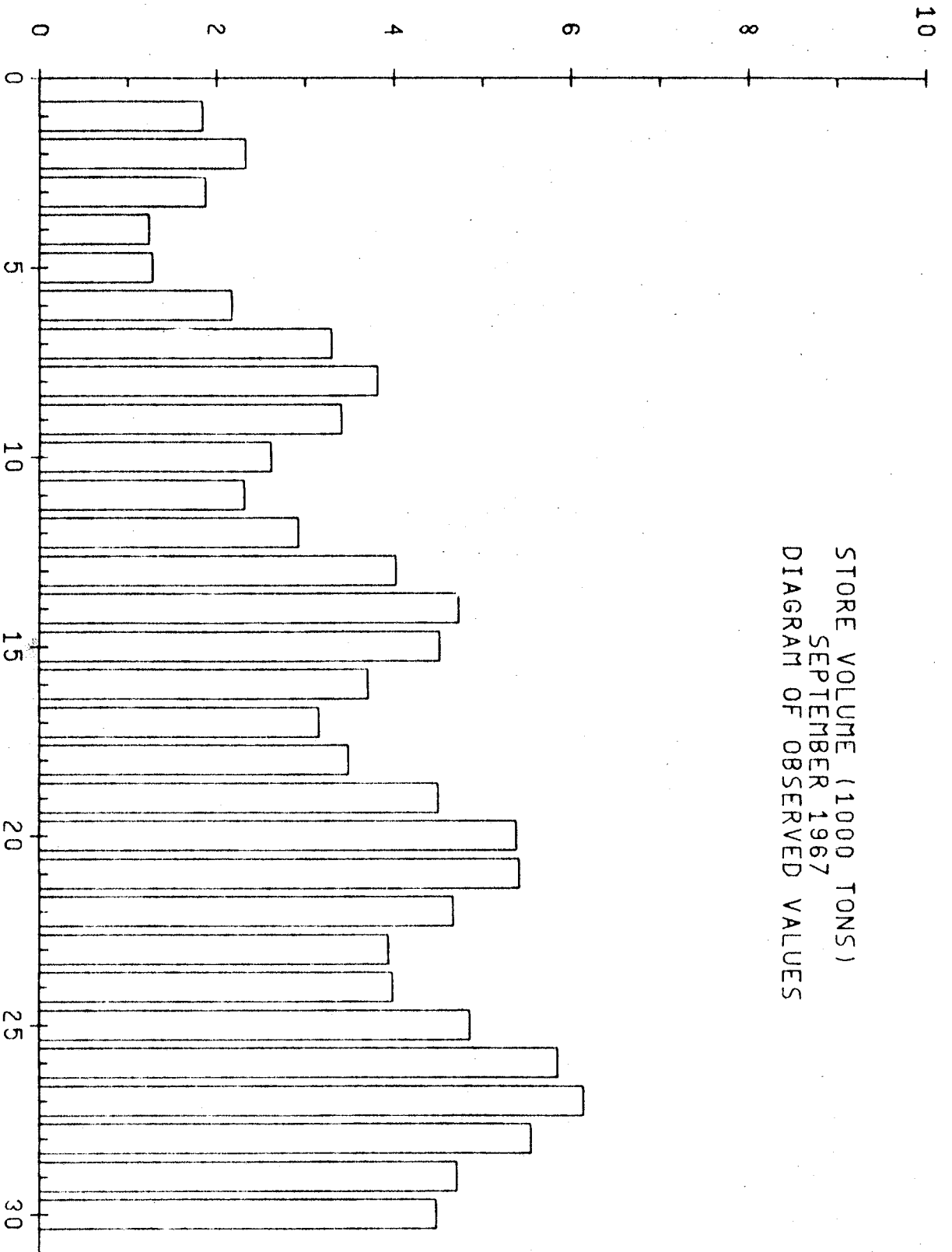
DSTPL1 will draw a steeple diagram while DHIST1 will draw a histogram.

Examples on some drawings made using these routines are found at the end of this chapter.

The routines are now in restricted use at the NCC and full publications will be given in a separate report to be published in the near future.



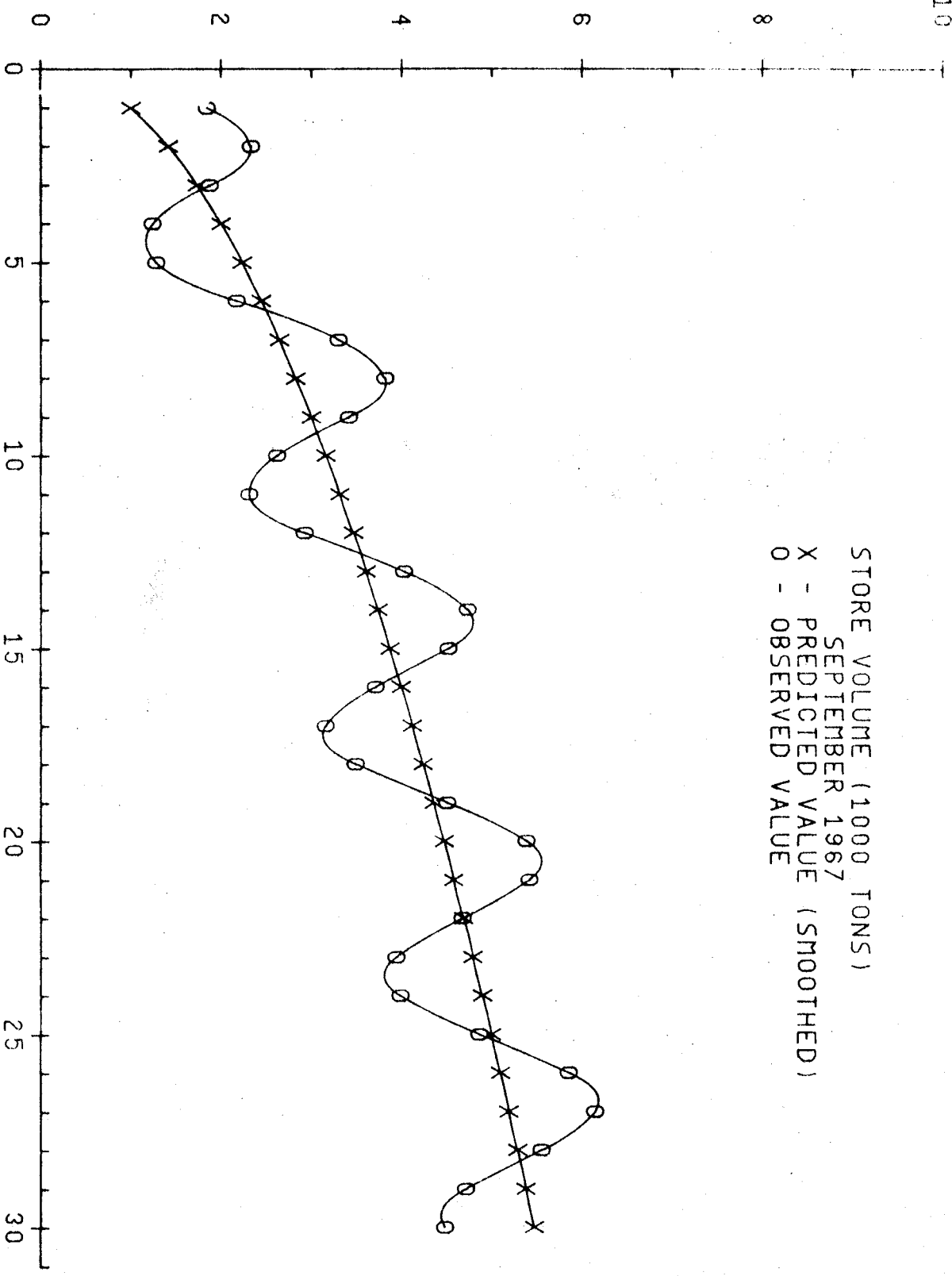
STORE VOLUME (1000 TONS)  
SEPTEMBER 1967  
DIAGRAM OF OBSERVED VALUES



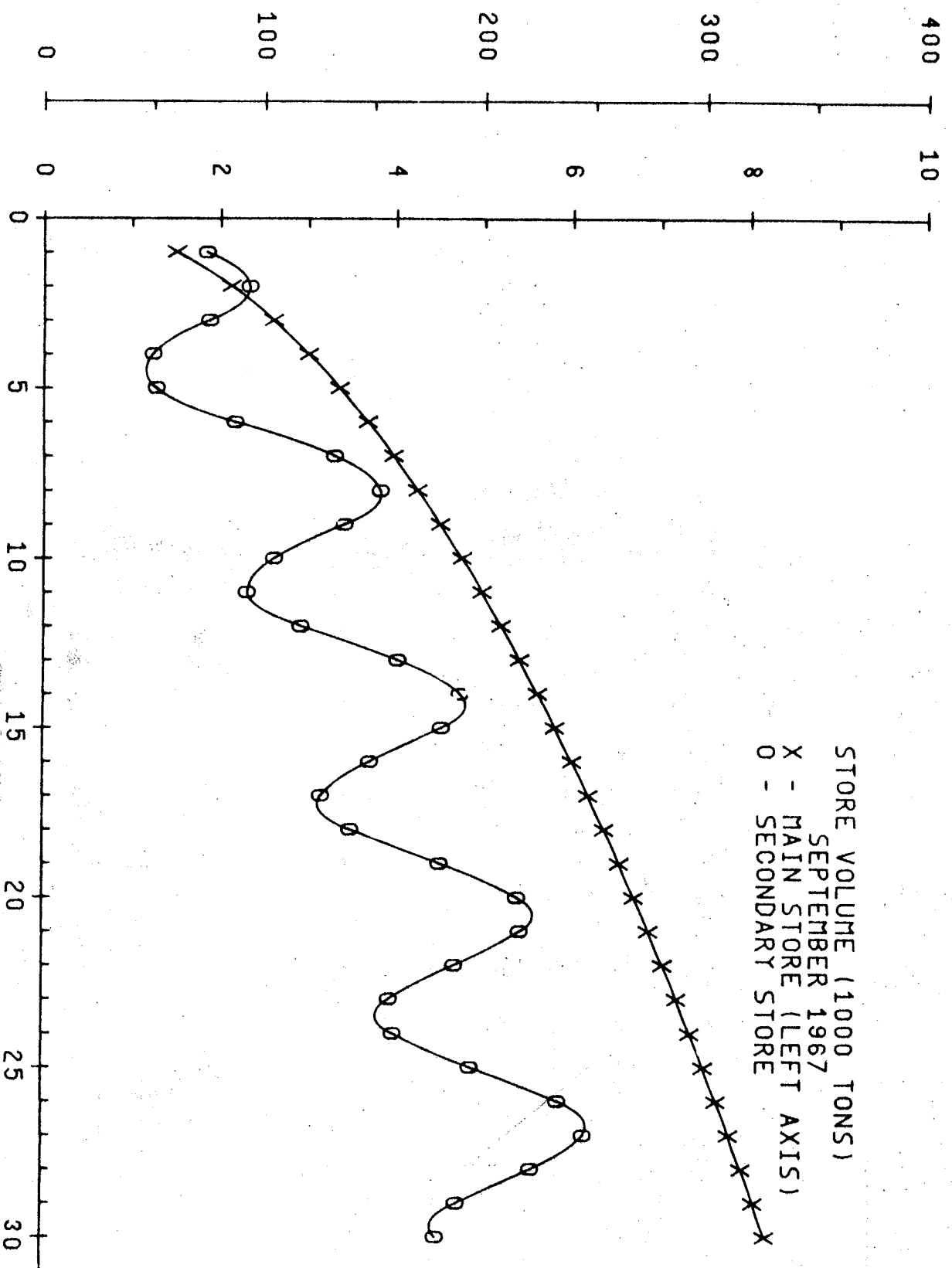
SCANDINAVIAN COAL COMPANY A/S

10

STORE VOLUME (1000 TONS)  
SEPTEMBER 1967  
X - PREDICTED VALUE (SMOOTHED)  
O - OBSERVED VALUE



STORE VOLUME (1000 TONS)  
 SEPTEMBER 1967  
 X - MAIN STORE (LEFT AXIS)  
 O - SECONDARY STORE



SCANDINAVIAN COAL COMPANY A/S