Five Computer Programs for Testing Weighted Residuals and Calculating Linear Confidence and Prediction Intervals on Results from the Ground-Water Parameter-Estimation Computer Program MODFLOWP

by Mary C. Hill

U.S. GEOLOGICAL SURVEY

Open-File Report 93-481

Denver, Colorado
1994

U.S. DEPARTMENT OF THE INTERIOR

Bruce Babbitt, Secretary

U.S. GEOLOGICAL SURVEY

Robert M. Hirsch, Acting Director

For additional information
write to:

Chief, Branch of Regional Research
U.S. Geological Survey
Water Resources Division
Box 25046, Mail Stop 418
Denver Federal Center
Denver, CO 80225-0046

Copies of this report can be
purchased from:

U.S. Geological Survey
Earth Science Information Center
Open-File Reports Section
Box 25286, Mail Stop 517
Dever Federal Center
Denver, CO 80225-0425

CONTENTS

FIGURES

TABLES

CONVERSION FACTORS

Multiply	By	To obtain
foot (ft)	0.3048	meter
foot per day (ft/d)	0.3048	meter per day

FIVE COMPUTER PROGRAMS FOR TESTING WEIGHTED RESIDUALS AND CALCULATING LINEAR CONFIDENCE AND PREDICTION INTERVALS ON RESULTS FROM THE GROUND-WATER PARAMETER-ESTIMATION COMPUTER PROGRAM MODFLOWP

Mary C. Hill

ABSTRACT

This report documents five computer programs, YR, NORM, BCINT, YCINT, and BEALEP, that use results from the ground-water flow and parameter-estimation computer program MODFLOWP. MODFLOWP uses nonlinear regression to estimate parameter values. Programs YR and NORM produce data sets for two commonly used graphs of weighted residuals that are important in analyzing regression results. Programs BCINT and YCINT calculate linear confidence intervals on parameter estimates, and linear confidence and prediction intervals on simulated hydraulic heads and simulated flows along head-dependent boundaries. These intervals are essential for displaying the reliability of simulated results, and can be used to make sensitivity analyses much more accurate than is allowed by traditional methods. Program BEALEP computes a measure of model linearity, which is needed to determine the accuracy of the linear intervals calculated with programs BCINT and YCINT. This report includes a discussion of the purposes of all the programs and how results can be used, brief descriptions of the theory behind each program, illustrative examples of how the programs may be used, descriptions of program input and output, and listings of the FORTRAN source codes.

INTRODUCTION

In the U.S. Geological Survey computer program MODFLOWP, the values of parameters of ground-water flow systems can be estimated by nonlinear regression using observations of hydraulic heads and flows along head-dependent boundaries, and prior information on parameter values (Hill, 1992). Nonlinear regression can be used to effectively estimate parameter values of ground-water flow systems, but each regression must be evaluated to determine whether the regression is valid. MODFLOWP prints some statistics needed in this evaluation, but does not support two commonly used graphical methods of evaluating weighted residuals.

Use of nonlinear regression also allows for the calculation of very useful measures of the uncertainty with which the parameter values are estimated, and the uncertainty in the simulated hydraulic heads and flows along head-dependent boundaries that results from the uncertainty in the estimated parameter values. The measures are called linear confidence and prediction intervals, and they are discussed below. Although MODFLOWP prints some information regarding the uncertainty of estimated parameter values, no information regarding the uncertainty of simulated hydraulic heads and flows along head-dependent boundaries is printed. MODFLOWP does not calculate linear confidence and prediction intervals.

Purpose and Scope

This report presents and describes five computer programs, YR, NORM, BCINT, YCINT, and BEALEP, that use results from MODFLOWP to (1) produce data sets that can be used to create two graphs commonly used to test weighted residuals, (2) calculate linear confidence and prediction intervals, and (3) test the validity of one of the assumptions of the method used to calculate the linear confidence and prediction intervals. The results of the programs need to be used together--for example, the validity of the calculated confidence intervals depends on the conclusions drawn from the graphical analyses of weighted residuals. Because of this interdependence, the purposes of all the programs and how results can be used are discussed in the following sections of the introduction. The remainder of this report includes additional discussion about the theory behind each program, how the calculations are performed, illustrative

examples of how the programs may be used, and descriptions of program input and output. Listings of FORTRAN source codes for all of the programs are included at the end of the report.

Testing Weighted Residuals

Evaluation of the regression requires an investigation of the differences between the observed and simulated hydraulic heads and flows and between the prior information and estimated parameter values (Draper and Smith, 1981; Cooley and Naff, 1990; Hill, 1992, p. 61-65). These differences are called residuals. If weighted regression is used (Hill, 1992, p. 38-50), weighted residuals need to be used in the evaluation, and they are calculated as (Hill, 1992, p. 39):

$$\text{weighted residual} = \omega_i^{1/2}(y_i - \hat{y}_i) \qquad (1)$$

where $\omega_i^{1/2}$ = the square root of the weight of the ith observation or prior estimate (Hill, 1992, p. 5);

y_i = the ith observation of hydraulic head or flow along a head-dependent boundary, or a prior estimate of a parameter value.

$\hat{y}_i$ = a simulated equivalent of the ith observation, or a regression estimate of a parameter value.

The purpose of programs YR and NORM is to produce data sets for two types of graphs commonly used to evaluate weighted residuals. The two graphs are described below and also are described by Draper and Smith (1981, p. 147-148) and Cooley and Naff (1990, p. 168-170).

The first graph is of the weighted residuals (eq. 1) against the weighted simulated values (calculated as $\omega_i^{1/2}\hat{y}_i$). Examples of this graph for five simulations are shown in figure 1; the simulations differ in how various features of a ground-water system are represented. For a regression to be valid, the weighted residuals need to be randomly distributed above and below a weighted residual value of zero for all weighted simulated values (Draper and Smith, 1981, p. 147-148). Problems are indicated if, for example, weighted residuals are grouped above or below zero for a distinct range of weighted simulated values or the spread of the weighted residuals about zero increases or decreases systematically with increasing or decreasing values of the weighted simulated values.

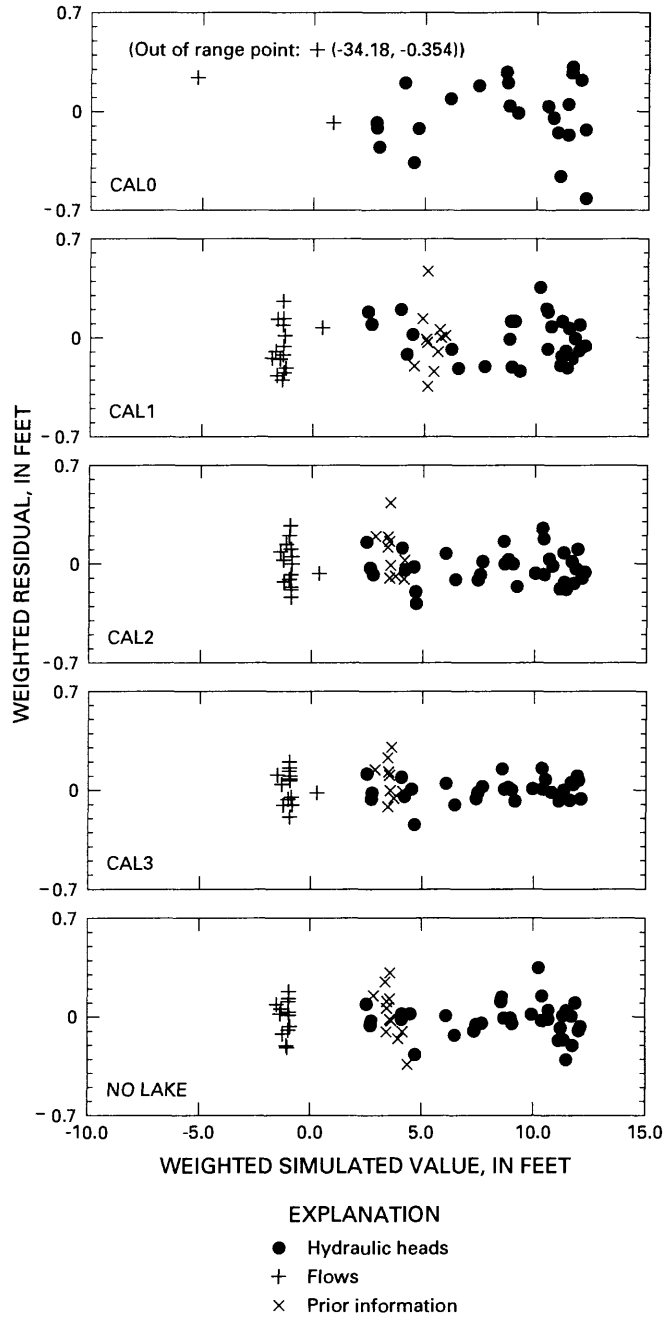


Figure 1.--Example graphs of weighted residuals and weighted simulated values for five simulations named CAL0, CAL1, CAL2, CAL3, and NO LAKE produced using output from program YR.

Examples showing common nonrandom graphs are presented by Cooley and Naff (1990, p. 170-171); nonrandom graphs may indicate the model and calibration problems discussed by Hill (1992, p. 66-69). In figure 1, the graph for CAL0 is nonrandom in that for weighted simulated values between 5 and 10 ft, all six weighted residuals are greater than zero. The other

graphs are more randomly distributed. The name of program YR is derived from the fact that simulated values often are referred to as $\hat{y}$ (eq. 1), and the word residuals begins with an r.

The second graph is a normal probability graph of the weighted residuals. An example of this graph is shown in figure 2. In a normal probability graph, the weighted residuals approximately fall on a straight line if they are independent and normally distributed. Helsel and Hirsch (1992, p. 30-33) show normal probability graphs characterized by several common problems. Benjamin and Cornell (1970, p. 449-450) show probability graphs with small samples of points generated from a probability distribution that is consistent with the graph, so that the values would be expected to fall on a straight line. Although the triangular probability distribution used in these examples differs from the normal probability distribution discussed in this report, the examples illustrate variations that might be expected with small samples. The name of program NORM is derived from the normal probability distribution.

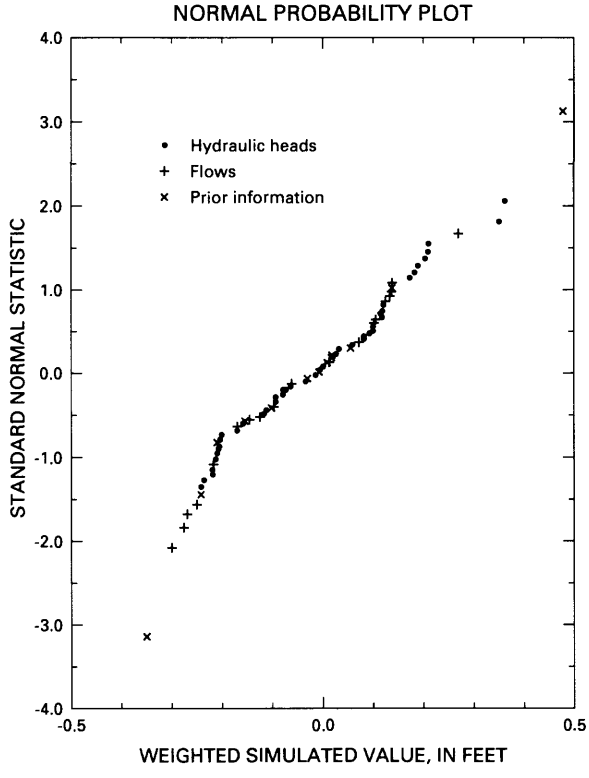


Figure 2.--Example normal probability graph of weighted residuals from simulation CAL1 produced using output from program NORM.

Whether the weighted residuals are independent and normally distributed is important if confidence and prediction intervals are to be calculated. Normality of the weighted residuals indicates that the true errors in the observations and prior information (Hill, 1992, p. 33) are normally distributed. If the true errors are normally distributed, the estimated parameter values and simulated hydraulic heads and flows are normally distributed, as generally is assumed to construct confidence and prediction intervals on these quantities (Draper and Smith, 1981, p. 22-31, 92-94; Helsel and Hirsch, 1992, p. 225).

Independence of the weighted residuals is not necessarily expected (Cooley and Naff, 1990, p. 168; Hill, 1992, p. 65), but any correlation that exists should be consistent with the correlation expected from the regression given that the true errors are independent. As discussed by Cooley and Naff (1990, p. 168-170), this can be tested by plotting normal deviates generated by using the expected correlation on normal probability graphs and by comparing these graphs to the graph of weighted residuals produced using NORM. For applications of MODFLOWP, normal deviates with the expected correlation can be generated with the computer program RESANP (called RESAN.MODP in Hill, 1992, p. 224).

If the graphs produced using YR show no problem with the model or the calibration, and the weighted residuals indicate that the true errors are normal and independently distributed, then confidence intervals on estimated parameter values and confidence and prediction intervals on simulated values may be calculated using a normal probability distribution.

Linear Confidence and Prediction Intervals

Programs BCINT and YCINT are designed to calculate linear confidence intervals of estimated parameter values and linear confidence or prediction intervals of simulated values (such as hydraulic heads and flows) using a normal probability distribution. The names of programs BCINT and YCINT were derived by noting that parameter values and simulated values are often represented mathematically using b and $\hat{y}$, respectively, and abbreviating 'confidence interval' as cint. Use of linear intervals is an application of first-order error analysis, and has been discussed by

many authors, including Graybill (1976), Cooley (1977, 1977a), Dettinger and Wilson (1981), Yeh (1986), Donaldson and Schnabel (1987), and Vecchia and Cooley (1987).

Confidence intervals are used to quantify the precision with which parameter values or simulated values are determined with the available data. When nonlinear regression is used in model calibration, the uncertainty with which the parameter values are estimated with the available data is quantified as variances on the parameters; the interdependence between parameter values is quantified as covariances on the parameter values (Hill, 1992, p. 57-59). The variances can be used to calculate confidence intervals on estimated parameter values so that the modeler can make statements such as: "With the available data, the value of the first parameter was estimated accurately enough that the probability that it lies between 10 and 30 ft/d is greater than 95 percent." The variances and covariances can be used to calculate confidence intervals on simulated values so that the modeler can make statements such as: "With the available data, the parameter values were estimated accurately enough that the probability that the true hydraulic head at well B at time five lies between 85 and 110 ft is greater than 95 percent." In this example, 85 and 110 ft are the limits of a 95-percent confidence interval which reflects all parameter variances and covariances, and, therefore, all parameter uncertainty and interdependence.

Prediction intervals reflect all of the uncertainty included in confidence intervals. In addition, prediction intervals include the uncertainty related to measurement error, and are, therefore, generally larger than the associated confidence interval. Prediction intervals need to be used to compare a measured value to an interval. For example, in a statement such as, "With the available data, the parameter values were estimated accurately enough that the probability that the measured hydraulic head at well B at time five lies between 84 and 111 ft is greater than 95 percent", 84 and 111 ft are the limits of a 95 percent prediction interval which reflects all parameter variances and covariances, and the error with which the measured hydraulic head is measured. The distinction between confidence and prediction intervals also is discussed in the section "Theory for YCINT" of this report.

The assumptions behind the construction of confidence and prediction intervals for sensitivity analyses are that: (1) the true errors in the available data are distributed as needed for the construction of the interval (usually a normal probability distribution is assumed); (2) the model correctly represents the relevant features of the ground-water flow system, so that the uncertainty in the simulated values is only produced by parameter uncertainty; and (3) if linear confidence intervals are calculated, the model is assumed to be effectively linear for parameter values close to the parameter values produced by the regression.

If a normal probability distribution is used, the first assumption can be tested using summary statistics printed by MODFLOWP and the graph produced using the computer program NORM, which is presented in this report. A normal probability distribution has been found to be valid for many ground-water model calibrations performed using nonlinear regression.

The second assumption is clearly untrue in some ways. The second assumption can be tested using the graphical methods suggested by Cooley and Naff (1990, p. 170-171); the first of these graphs can be produced using the computer program YR, which is presented in this report. In nonlinear regression, how much of the uncertainty in the simulated values is produced by parameter uncertainty and thus reflected in the confidence and prediction intervals on simulated values depends on how the parameters are defined. If they are defined such that they are influential in the calculation of all heads and flows in the ground-water flow system, the calculated confidence intervals probably reflect the model uncertainty accurately.

The third assumption can be tested using the modified Beale's measure (Cooley and Naff, 1990, p. 187-189), which can be calculated using computer program BEALEP, which is presented in this report.

Program BCINT produces an output file that can be used with a separate plotting routine to produce graphs as shown in figure 3. This graph shows estimated parameter values and their associated linear confidence intervals for four simulations.

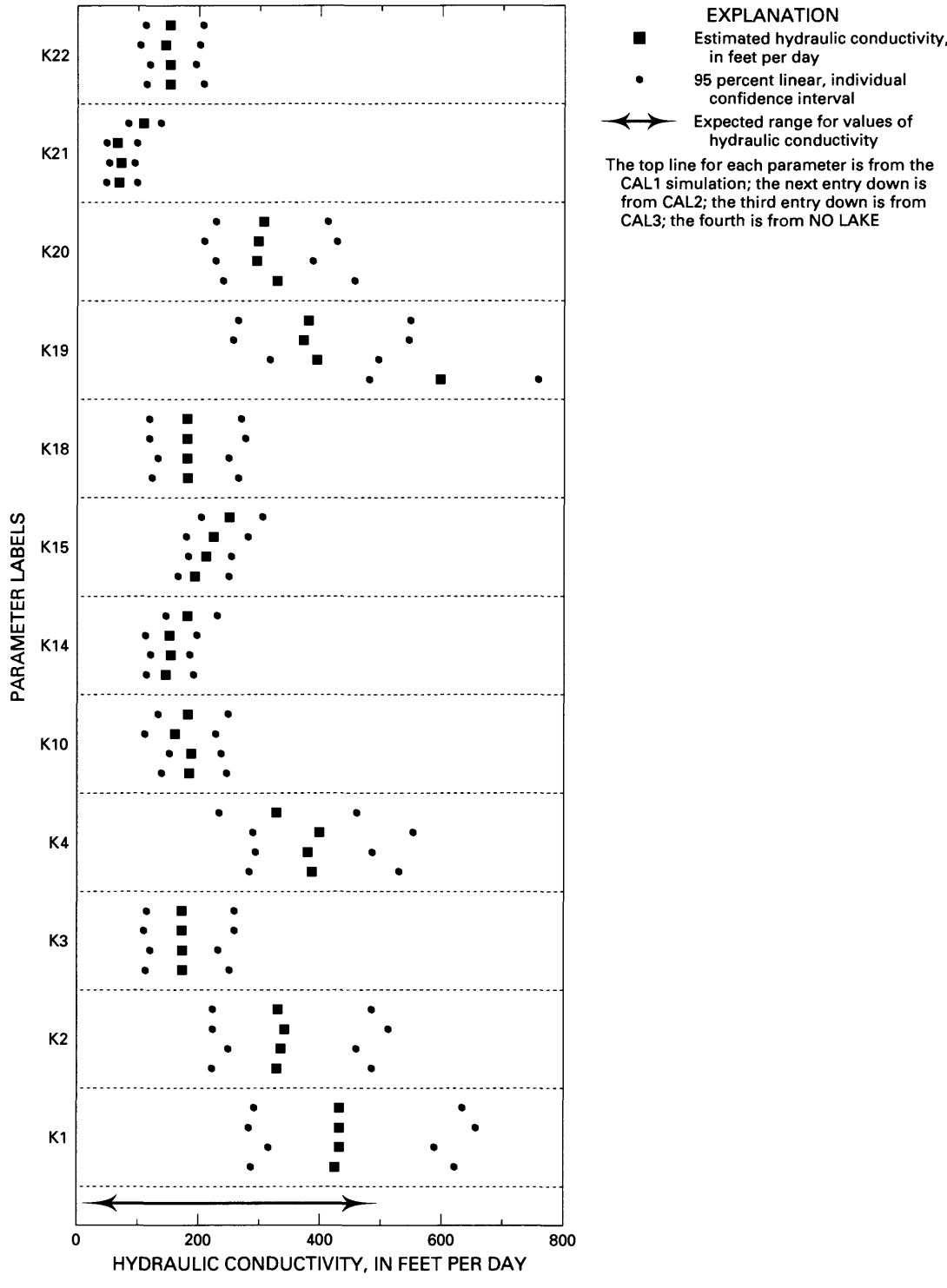


Figure 3.--Example graph of parameter estimates and their linear, individual confidence intervals for four simulations named CAL1, CAL2, CAL3, and NO LAKE produced using output from program BCINT.

Program YCINT calculates confidence or prediction intervals on simulated hydraulic heads and flows along head-dependent boundaries. Confidence and prediction intervals on the changes in these values between two simulations also can be calculated. This is needed, for example, to calculate a confidence or prediction interval on the drawdown caused by pumpage. The confidence or prediction intervals are printed in tabular form in the output file and are not ready for plotting.

Using Confidence Intervals for Sensitivity Analyses

Confidence intervals on simulated values can be used to replace the traditional procedure used to perform sensitivity analyses. According to Anderson and Woessner (1992, p. 246), "The purpose of a sensitivity analysis is to quantify the uncertainty in the calibrated model caused by uncertainty in the estimated parameter values", and in the procedure traditionally followed to fulfill this purpose, "calibrated values for hydraulic conductivity, storage parameters, recharge, and boundary conditions are systematically changed within the previously established plausible range." The results of several traditional sensitivity analyses are shown in Anderson and Woessner (1992, p. 247-254). The major weaknesses of the traditional procedure are that:

(1) the 'plausible range' usually is determined subjectively prior to model calibration (Anderson and Woessner, 1992, p. 231), so that it does not reflect the significant increase in the certainty with which some parameters are estimated based on model calibration. In this circumstance, the uncertainty in the calibrated model, as determined by the sensitivity analysis, would be unrealistically large.
(2) Coordinated changes in two or more parameter values are rarely considered, though they are often important in ground-water flow problems. A typical example is that if hydraulic conductivity and areal recharge are both increased, in many ground-water flow simulations the hydraulic heads change very little. Anderson and Woessner (1992, p. 248) suggest that in traditional sensitivity analyses hydraulic conductivity and areal recharge values be changed in opposite directions to display the full range of possible model uncertainty. In some cases this is appropriate, however, if model calibration has shown that this combination of parameter values

produces results that are contrary to the available data, including such a simulation as part of the sensitivity analysis will produce an exaggerated impression of model uncertainty.

Confidence intervals can be used to fulfill the purpose of sensitivity analyses more effectively than the traditional approach. Confidence intervals can be used to construct graphs that convey information analogous to that conveyed in the graphs presented by Anderson and Woessner (1992, p. 247-254), or, for example, confidence intervals could be calculated for hydraulic heads at many grid locations and the length of the confidence intervals could be contoured to produce a map of model uncertainty (for more ideas for using confidence intervals, see Hill, 1989, p. 188-189). When predictive simulations are constructed, confidence intervals can be included to quantify the likely uncertainty in the prediction.

A Problem with using parts of the MODFLOWP Output File as Input for YR, NORM, and BCINT

The input files for computer programs YR, NORM, and BCINT are at least partially composed of sections of the main MODFLOWP output file -- the modeler uses an editor to extract pieces from the output file and puts the pieces into the input files. The formats used to read these lines in YR, NORM, and BCINT are coordinated with the formats used to print them from MODFLOWP. A problem will occur if the first character in each line is used as a FORTRAN format character, as occurs on some computers in some situations. When this occurs, all lines of the main MODFLOWP output file are shifted one column to the left, and can not be read correctly by YR, NORM, and BCINT.

This problem can be detected by comparing the relevant sections of the main MODFLOWP output file to the data formats presented later in this report, or by inspection of one of the output files from YR, NORM, or BCINT -- look for missing minus signs in the first column. If the problem occurs, it can be fixed in one of the following ways: (1) Add a blank to the beginning of each line taken from the main MODFLOWP output file; or (2) change the format statements in YR, NORM, and BCINT -- in formats 505, 506, and 507 of YR and NORM, and in formats 510 and 520 of BCINT,

eliminate one of the skipped spaces preceding the fist occurrence of F11.0 (YR and NORM), or A4 or F10.0 (BCINT).

COMPUTER PROGRAM YR FOR CALCULATING VALUES FOR GRAPHS OF WEIGHTED RESIDUALS AGAINST WEIGHTED SIMULATED VALUES

Theory for YR

Construction of plots of weighted residuals against weighted simulated values and the underlying theory generally is simple and is covered in the introduction of this report. There are, however, three common problems with graphs of weighted residuals and weighted simulated values. These can be corrected as described below.

The first problem occurs when the weighted simulated values differ from each other enough that a convenient scaling of the horizontal axis on graphs such as those in figure 1 is not possible. To resolve this problem, the weighted simulated values of outlying points can be multiplied by a factor so they will plot closer to the other weighted simulated values. To ensure that the graph can still indicate whether the weighted residuals vary systematically with weighted simulated values for any one type of data, the factor needs to be applied carefully. It is usually a good idea to apply the same factor to all weighted simulated values for a given type of data--for example, the same factor could be applied to all flow weighted simulated values. Factors can be applied as described in the section "Input for YR".

The second problem occurs when the weights are calculated using the observed values as, for example, sometimes is done for streamflow (Hill, 1992, p. 49-50). In this circumstance, the standard deviation of the observation error is estimated as a constant times the observed value, and the weight, ω, equals $\sigma^2/(c \times y)^2$, where σ^2 is the common error variance (Hill, 1992, p. 39), c is the constant, and y represents an observed value of hydraulic head or flow or a prior information value. The constant, c, ideally equals the coefficient of variation (the standard deviation divided by the true, but unknown value of y). Using streamflow as an example, the weighted residual is calculated as:

$$\text{weighted residual} = [\omega_j^{1/2}] \times (q_j - \hat{q}_j) = [\sigma/(c_j \times q_j)] \times (q_j - \hat{q}_j), \quad (2)$$

where j = the identification number of the flow observation;
$\sigma/(c_j \times q_j)$ = the square root of the weight, also represented by $\omega_j^{1/2}$;
q_j = the observation; and
$\hat{q}_j$ = the simulated value.

The weighted, simulated value is calculated as:

$$\text{weighted simulated value} = [\omega_j^{1/2}] \times \hat{q}_j = [\sigma/(c_j \times q_j)] \times \hat{q}_j. \quad (3)$$

Normally, the values calculated by equations (2) and (3) would be used in graphs of weighted residuals and weighted simulated values. However, multiplying through, equations (2) and (3) can be rewritten as:

$$\text{weighted residual} = \sigma/c_j - (\sigma/c_j) \times (\hat{q}_j/q_j) \quad (4)$$

$$\text{weighted, simulated value} = (\sigma/c_j) \times (\hat{q}_j/q_j). \quad (5)$$

The second term of equation (4) equals equation (5), and σ/c_j is a constant. If c_j is the same for multiple observations, the related values will plot on a straight line with a slope of -1. This graph would give no indication about whether the weighted residuals vary systematically with the size of the simulated value. To rectify the situation, the weighted residual needs to be graphed against a value that excludes the observed value used to calculate the weight.

This second problem can be resolved by multiplying the weighted simulated values by q_j. Thus, equation (5) would become:

$$\text{modified weighted, simulated value} = \omega_j^{1/2} q_j \hat{q}_j = (\sigma/c_j) \times \hat{q}_j \quad (6)$$

To accomplish this, YR must be modified: The statement Q=Q*WT*FAC, in the DO 20 loop needs to be changed to Q=Q*WT*FAC*ABS(QOBS), where QOBS equals q_j, which already has been read by YR from QCALi (see section "Input for YR"). This change fixes the second problem, but it often exacerbates the first problem discussed above.

The third problem occurs when estimated parameters that have prior information are scaled using the value of the prior information. Scaling is sometimes convenient because it transforms all prior information to equal 1.0, and the percentage change between the estimate and the prior value is obvious from the estimate: for example, an estimate of 1.5 indicates a 50-percent increase over the prior value. In this circumstance, the weighted residuals are calculated as:

$$\text{weighted residual} = \omega_k^{1/2} \times (1.0 - b_k) \tag{7}$$

and the weighted, simulated values are calculated as:

$$\text{weighted, simulated value} = \omega_k^{1/2} \times b_k \ , \tag{8}$$

where ω_k = the weight for the prior value of b_k, where b_k is the kth estimated parameter.

If ω_k is the same for multiple prior parameters, a graph of weighted residuals and weighted simulated values is a straight line with a slope of -1. As for the second problem discussed above, the graph would not indicate whether the weighted residuals vary systematically with the size of the simulated value. A meaningful graph can be obtained by calculating a modified weighted, simulated value as:

$$\text{modified weighted, simulated value} = \omega_k^{1/2} \times b_k \times P_k \ , \tag{9}$$

where P_k = the original, unscaled value of the prior information. This can be accomplished by listing the original P_k values in the input file YR.DAT, which is described below. No program modifications are required to use equation (9) instead of equation (8); YR uses P_k = 1.0 in equation (9) unless input file YR.DAT is used as described in the section 'Input for YR'.

The second and third problems discussed above result from using the observed values and prior parameter estimates, which are considered to be random variables within the context of regression, to calculate weights and scaling factors, respectively, which are considered to be constant, known values within the context of regression. Although this usage is nonstandard and should be avoided if possible, there are circumstances in

the complicated problems that MODFLOWP is meant to represent in which use of these methods produce the best possible estimates of the correct weighting of observed hydraulic heads and flows, and the most convenient scaling of the parameters. Care should be taken when using these nonstandard methods so that, for example, graphs of weighted residuals and weighted simulated values are modified as discussed above to display what they are intended to display.

Input for YR

Often in model calibration, several calibrated models are compared. To accommodate this comparison, YR is constructed to read data from as many as 10 simulations. These data need to be in files with specific names, as discussed below. The value of NSD, which is defined immediately after the format statements in YR, must be equal to or greater than the number of simulations being considered, but may not exceed 10.

Most of the input for computer program YR consists of parts of the main MODFLOWP output file (See section of introduction "A problem with using parts of the MODFLOWP output file as input for YR, NORM, and BCINT"). Examples of the MODFLOWP output file are included in appendix A of Hill (1992, p. 174-192 and 205-221). For each of the simulations considered, the input for YR consists of three files that contain information on hydraulic heads, flows, and prior parameter estimates. The files need to be named HCAL0, HCAL1,....; QCAL0, QCAL1,...; and PCAL0, PCAL1,.... The numbers, 0, 1,..., refer to as many as 10 simulations, so the sequences may continue up to HCAL9, QCAL9, and PCAL9. YR will try to read files for i=0, 1,...NSD-1, where NSD is discussed above. For each of the HCALi, QCALi, and PCALi, i=0,1,...NSD-1 files that do exist, the following statement is printed to ISCREEN (see last paragraph of this section):

NUMBER OF NON-BLANK LINES READ FROM FILE filename = n,

where n is the number of lines. If a file does not exist, the following statement is printed to unit ISCREEN:

FILE filename DOES NOT EXIST

and execution continues. This message, therefore, does not necessarily indicate a problem. Examples of HCALi, QCALi, and PCALi input files used to produce figure 1 are shown in figure 4.

A

```
Column 1 of input file HCALi
|
    1 1         1   3  18      0     0.00 0.00 0.00      12.050      12.146 -0.957E-01   1.00      -0.957E-01
    2 1         3   3  18      0     0.00 0.00 0.00      11.850      11.887 -0.366E-01   1.00      -0.366E-01
    3 2         1   4  11      0     0.00 0.00 0.00      11.630      11.790 -0.160       1.00      -0.160
    4 2         3   4  11      0     0.00 0.00 0.00      11.350      11.565 -0.215       1.00      -0.215
    5 3         1   7  21      0     0.00 0.00 0.00      12.150      12.051  0.992E-01   1.00       0.992E-01
    6 3         3   7  21      0     0.00 0.00 0.00      11.830      11.830  0.399E-03   1.00       0.399E-03
    7 4         1  13  23      0     0.00 0.00 0.00      11.440      11.323  0.117       1.00       0.117
    8 4         3  13  23      0     0.00 0.00 0.00      11.430      11.311  0.119       1.00       0.119
    .
    .
    .
   34 17        3  38  23      0     0.00 0.00 0.00       6.220       6.295 -0.753E-01   1.00      -0.753E-01
   35 18        1   3  23      0     0.00 0.00 0.00      12.220      12.283 -0.635E-01   1.00      -0.635E-01
   36 18        3   3  23      0     0.00 0.00 0.00      12.120      12.038  0.820E-01   1.00       0.820E-01
   37 19        1   7   9      0     0.00 0.00 0.00      11.290      11.462 -0.172       1.00      -0.172
   38 19        3   7   9      0     0.00 0.00 0.00      11.100      11.321 -0.221       1.00      -0.221
   39 20        1  10  15      0     0.00 0.00 0.00      11.710      11.651  0.586E-01   1.00       0.586E-01
   40 20        3  10  15      0     0.00 0.00 0.00      11.300      11.396 -0.958E-01   1.00      -0.958E-01
   41 21        1  22  17      0     0.00 0.00 0.00       9.140       9.019  0.121       1.00       0.121
   42 21        3  22  17      0     0.00 0.00 0.00       8.950       8.830  0.120       1.00       0.120
   43 22        1  29  18      0     0.00 0.00 0.00       6.430       6.641 -0.211       1.00      -0.211
   44 22        2  29  18      0     0.00 0.00 0.00       6.410       6.631 -0.221       1.00      -0.221
```

B

```
Column 1 of input file QCALi
|
    45   S1     0     -0.521E+05 -0.546E+05  0.250E+04  0.543E-04  0.136      1.E-5
    45   S2     0     -0.521E+05 -0.546E+05  0.250E+04  0.543E-04  0.136      1.E-5
    46   S3     0     -0.443E+05 -0.409E+05 -0.342E+04  0.638E-04 -0.218      1.E-5
     .
     .
     .
    61   S17    0     -0.408E+05 -0.423E+05  0.145E+04  0.693E-04  0.101      1.E-5
    62   S18    0     -0.466E+05 -0.421E+05 -0.446E+04  0.607E-04 -0.271      1.E-5
    63   SSL    0      0.335E+05  0.318E+05  0.170E+04  0.423E-04  0.718E-01  1.E-5
```

C

```
Column 1 of input file PCALi
|
      1     0.100E+01  0.984E+00  0.165E-01   1.00       0.165E-01
      2     0.100E+01  0.991E+00  0.889E-02   1.00       0.889E-02
      3     0.100E+01  0.103E+01 -0.314E-01   1.00      -0.314E-01
      4     0.100E+01  0.946E+00  0.550E-01   1.00       0.550E-01
      5     0.100E+01  0.621E+00  0.477E+00   1.00       0.477
      6     0.100E+01  0.142E+01 -0.353E+00   1.00      -0.353
      7     0.100E+01  0.128E+01 -0.244E+00   1.00      -0.244
      8     0.100E+01  0.101E+01 -0.761E-02   1.00      -0.761E-02
      9     0.100E+01  0.983E+00  0.171E-01   1.00       0.171E-01
     10     0.100E+01  0.111E+01 -0.103E+00   1.00      -0.103
     11     0.100E+01  0.123E+01 -0.206E+00   1.00      -0.206
     12     0.100E+01  0.872E+00  0.137E+00   1.00       0.137
```

Figure 4.--Examples of (A) HCALi, (B) QCALi, and (C) PCALi (i=0,1,..., or 9) input files for programs YR and NORM.

Each file HCALi (i=0,1,...or 9) needs to contain the final repetition of the table 'DATA AT HEAD LOCATIONS' from the MODFLOWP output file of the ith simulation. Examples of this table are shown in Hill (1992, p. 189 and 216). In the first example (p. 189), HCALi would contain the lines for observations 1 through 32. Each line of an HCALi file is read using the format A10, 51X, F11.0, 11X, 2F11.0. The first 10 characters are used

to identify blank lines. The numbers read by the F11.0 formats are the values from the columns labeled "CALC. HEAD", "WEIGHT**.5" and "WEIGHTED RESIDUAL".

Each file QCALi (i=0,1,...or 9) needs to contain the final repetition of the table 'DATA FOR FLOWS' from the MODFLOWP output file of the ith simulation. An example of the table is shown in Hill (1992, p. 189). In that example, QCALi would contain the lines for observations 33 through 35. Each line of a QCALi file is read using the format A10, 11X, 2F11.0, 11X, 2F11.0. The first 10 characters are used to identify blank lines. The numbers read by the F11.0 formats are the values from the columns labeled "MEAS. FLOW", "CALC. FLOW", "WEIGHT**.5", and "WEIGHTED RESIDUAL".

Each file PCALi (i=0,1,...or 9) needs to contain the final repetition of the tables 'PARAMETERS WITH PRIOR INFORMATION, BY GROUP' if DATA SET 11 was used to designate prior information and 'PARAMETER SUMS WITH PRIOR INFORMATION' if DATA SET 12 was used to designate prior information. Examples of these tables are shown in Hill (1992, p. 217). In that example, PCALi would contain seven lines from the first table and two lines from the second table. Each line of a PCALi file is read using the format A10, 11X, F11.0, 11X, 2F11.0. The first 10 characters are used to identify blank lines. The numbers read by the F11.0 formats are the values from the columns "VALUES--CALC.", "WEIGHT**5" and "WEIGHTED RESIDUAL".

An additional file, YR.DAT is needed if one or both of two circumstances occur. The first circumstance is that any parameter with prior information is estimated as a log-transformed value (This option is activated using the variable LN of DATA SET 2 of Hill, 1992, p. 136). The second circumstance is that values of P_k need to be provided to rectify the third problem discussed in the last section. The YR.DAT file used to produce figure 1 is shown in figure 5. Each line of YR.DAT contains two numbers in F10.0, I5 format (Note that real numbers with digits to the right of the decimal point can be read using F10.0 format). The first line contains the value of P_k and LN for the parameter to which the first line of PCALi, i=0,1,...9 refers (fig. 4); the second line contains the value of P_k and LN for the parameter to which the second line of PCALi, i=0,1,...9 refers; and so on. The number of lines in YR.DAT must equal

the maximum number of lines in any PCALi file. If a file named YR.DAT does not exist, the following statement is printed to unit ISCREEN (see last paragraph of this section):

FILE YR.DAT DOES NOT EXIST

and execution continues with P_k = 1.0 in equation (9) for all k and LN = 0 for all parameters.

```
Column 1 of input file YR.DAT
|
437.          1
329.          1
165.          1
341.          1
288.          1
126.          1
193.          1
173.          1
380.          1
268.          1
82.           1
163.          1
169.          1
218.          1
195.          1
182.          1
```

Figure 5.--Example of a YR.DAT input file for YR.

YR is designed to read only one YR.DAT file. Separate YR.DAT files for each simulation are necessary if the differences between PCALi files are more than simply the addition of new prior values as would occur, for example, if the parameter definitions used in a simulation were reorganized so that the second parameter of one simulation was different than the second parameter of another simulation. Simulations requiring different YR.DAT files can not be processed with a single execution of program YR.

The input and output unit numbers used by YR are defined in the six lines following the definition of NSD, which follows the format statements in YR. Set ISCREEN to the FORTRAN unit number of the screen or a file; set IUH, IUQ, and IUP to the FORTRAN unit numbers for the input files HCALi, QCALi, and PCALi, respectively (the files for each simulation, i,

are opened, read, and closed sequentially); set IUM to the unit number for YR.DAT; and set IOUT to the unit number for the output files YRCALi, which are described below. All files are opened within the program.

Output for YR

YR produces output files named YRCALi for every simulation i for which at least one of the three input files exists. Each line in each YRCALi file contains two numbers written using format 2G16.8: the first is the weighted, simulated value (or modified weighted, simulated value); the second is the weighted residual. Most available plotting routines accept data in this format.

It is often helpful to have the three different types of data plotted with different symbols, as in figure 1. To accomplish this, the output from program YR needs to be coordinated with the plotting package used. Program YR is designed to insert blank lines between the different types of data (statement numbers 12 and 22). These blank lines provide the needed coordination with the nonproprietary plotting package used to produce figure 1 (Turner, 1992). To coordinate with other plotting packages, changes can be made to the WRITE statements in the DO 10, DO 20, and DO 30 loops of YR.

COMPUTER PROGRAM NORM FOR CALCULATING VALUES FOR NORMAL PROBABILITY GRAPHS OF WEIGHTED RESIDUALS

Theory for NORM

Normal probability graphs are used as discussed in the introduction of this report. Normal probability graphs are constructed by ordering the weighted residuals from smallest to largest and plotting them against the cumulative probability that would be expected for each value if they were independent and normally distributed. The cumulative probabilities can be calculated in a number of ways, as discussed by Draper and Smith (1981, p. 178) and Looney and Gulledge (1985). In the computer program NORM, they are calculated as (k-0.5)/n (Hazen, 1914), where n equals the number of weighted residuals and k equals one for the smallest weighted residual, two for the next largest residual, and so on. For the largest residual, k

equals n. Calculating the cumulative probabilities in this way makes the normal probability graphs consistent with how the statistic R_N^2 is calculated by MODFLOWP (Hill, 1992, p. 63), so R_N^2 is the correlation coefficient for the probability graph.

Normal probability graphs often are plotted on normal probability paper, on which the sample cumulative probabilities are graphed along a nonarithmetic scale (Benjamin and Cornell, 1970, p. 453). Because available plotting packages rarely provide a normal probability scale as an option, an arithmetic scale is preferable. A graph with an arithmetic scale can be created using the fact that in NORM the sample cumulative probabilities are calculated as (k-0.5)/n, as discussed above. Note that a sample cumulative probability is the probability that a randomly selected weighted residual (X) is less than the value of the ith smallest residual (x_i), or $P(X \leq x_i)$. To produce a graph with an arithmetic axis, the sample cumulative probabilities can be used to calculate abscissa values (the values along the horizontal axes of graphs of probability distributions) of a standard normal probability distribution. The abscissa values plot on an arithmetic scale, and are used in the graph shown on figure 2. In NORM, the abscissa values are determined using subroutine UNORM from MODFLOWP (Hill, 1992, p. 238 and 357-358). The axis is labeled 'standard normal statistic'.

Input for NORM

The input files for NORM are identical to the input files for YR, except that YR.DAT is omitted. NSD is defined in the PARAMETER statement in the beginning of the program.

The input and output unit numbers used by NORM are defined in the seven lines following the format statements in NORM. Set ISCREEN to the FORTRAN unit number of the screen or a file; set IUH, IUQ, and IUP to the FORTRAN unit numbers for the input files HCALi, QCALi, and PCALi, respectively; and set IOUTH, IOUTQ, and IOUTP to the unit numbers for the output files NORMHi, NORMQi, and NORMPi, respectively. The output files are described below. All files are opened within the program.

The PARAMETER statement in the beginning of NORM may need to be changed to allow enough spaces in the dimensioned arrays for the problem being considered. NDD must equal or exceed the number of hydraulic-head and flow observations; NPRD must equal or exceed the number of values of prior information.

Output for NORM

NORM produces output files named NORMHi, NORMQi, and NORMPi (i=0,1,...or 9). The NORMHi files contain the weighted residuals and the standard normal statistics related to hydraulic-head observations and are produced for each simulation, i, for which there is a file HCALi. NORMQi files contain data for graphs of the weighted residuals related to hydraulic-head and flow observations and are produced for each i for which there is a file QCALi. NORMPi files contain data for graphs of the weighted residuals related to hydraulic-head observations, flow observations, and prior information and are produced for each i for which there is a file PCALi. In all files, each line contains a weighted residual and associated standard normal statistic, printed using format 2G15.6.

It often is helpful to have the three different types of data plotted with different symbols, as was done in figure 2. To accomplish this, the output from program NORM needs to be coordinated with the plotting package used. Program NORM is designed to insert blank lines between the different types of data (statement number 300 of subroutine ORDER). These blank lines provide the needed coordination with the nonproprietary plotting package used to produce figure 2 (Turner, 1992). To coordinate with other plotting packages, changes can be made to the WRITE statement in the DO 290 loop in subroutine ORDER of NORM.

COMPUTER PROGRAM BCINT FOR CALCULATING VALUES FOR GRAPHS OF ESTIMATED PARAMETER VALUES AND THEIR LINEAR CONFIDENCE INTERVALS

Theory for BCINT

The computer program BCINT calculates linear confidence intervals on estimated parameter values and creates a file of these values which can be graphed as shown in figure 3.

The equations for the confidence intervals are developed based on the assumptions that (1) the parameter values are normally distributed, (2) the model is correct, and (3) the model is roughly linear for parameter values close to the optimal parameter values. These assumptions need to be tested as discussed in the introduction of this report.

The confidence intervals calculated for parameter estimates are usually individual linear confidence intervals, which are calculated as (Draper and Smith, 1981, p. 94):

$$b'_\ell \pm (v_{\ell\ell})^{1/2} \times t_s(ND+NPR-NP', 1.0-\alpha/2) \qquad (10)$$

where b'_ℓ = the estimate of the ℓth parameter;

$v_{\ell\ell}$ = the estimated variance of the ℓth parameter from the variance-covariance matrix on the parameter values;

$t_s(.,.)$ = the critical value of a random variable from the student-t probability distribution in which the first argument equals the degrees of freedom, the second argument equals the probability that the true parameter value occurs within the confidence interval, and α equals the probability that the confidence interval does not contain the true value;

ND = the number of observations used in the regression;

NPR = the number of prior estimates used in the regression; and

NP' = the number of parameters estimated in the regression (equals NP of line 3 of the Parameter-Estimation Package input file minus one for each parameter with a negative group number in data set 9 of that input file).

Alternatively, simultaneous confidence intervals, or individual or simultaneous prediction intervals, could be calculated if needed based on how the intervals are to be used. The different intervals are defined and discussed in the section 'Computer Program YCINT'.

Input for BCINT

Often in model calibration, several calibrated models are compared. To accommodate this circumstance, BCINT was constructed to read data from as many as NSD simulations, where NSD is defined in the PARAMETER

statement at the beginning of BCINT, and may not exceed 10. The data need to be in files named PAR0, PAR1,..., where the numbers 0, 1,... refer to as many as 10 simulations so the sequence may continue up to PAR9. BCINT will try to read files for i=0,1,...NSD-1. If a file does not exist, the following statement is printed to unit ISCREEN (see below):

FILE filename DOES NOT EXIST

and execution continues. This message, therefore, does not necessarily indicate a problem.

Each file PARi (i=0,1,...or 9) needs to contain the data sets listed in table 1. As an example, one of the PARi files used to produce figure 3 is shown in figure 6.

The input and output unit numbers used by BCINT are defined in the three lines following the format statements in BCINT (see section "Listing of computer program BCINT"). ISCREEN is the FORTRAN unit number of the screen or a file; IPAR is the unit number for the PARi files; and IOUT is the unit number for the output file, CI.OUT, which is described in the next section.

Output for BCINT

BCINT produces an output file named CI.OUT, which is composed of two sets of lines separated by one blank line. Each non-blank line of the output file contains two numbers in format 2G13.3: the first number is an estimated parameter value or, after the blank line, one of the limits of a confidence interval; the second number is an ordinate (y-axis) plotting position. The ordinate plotting position is the same for the estimated parameter value and the two ends of the confidence interval for a single parameter from one simulation so the values graph along the same horizontal line as in figure 3. The ordinate plotting positions are incremented so parameters from different simulations having the same value of IY (see table 1) are plotted together.

Table 1.-- Data Sets of the PARi (i=0,1,...or 9) input files for BCINT

Data set	Number of lines	Format	Variable	Definition
1	1	I5,F10.0	NP',STAT[1]	Number of parameters and the statistic used to calculated the confidence interval.
2	1 or more	10I5	LN(I), I=1,NP'	LN from DATA SET 2 of MODFLOWP for each parameter.
3	1 or more	10I5	IY(I), I=1,NP'	Confidence intervals are calculated for parameters I for which IY(I)>0, and parameters from different simulations with the same value of IY(I) are grouped.[2]
4	1 or more	10F5.0	BX(I), I=1,NP'	Multiplicative factors used to scale parameter values.[3]

The remainder of the file consists of the table 'PARAMETER SUMMARY' from the MODLFOWP output file[4], starting with the line that begins with 'PARAMETER ID' and ending after the parameter standard deviations, which are labeled 'STD. DEV'. The first part of the table is read using formation 16X, 10(6X,A4); the characters read are PIDs, which are described by Hill (1992, p. 113 and 136). Next a line is skipped and the estimated parameter values are read using format 17X, 10F10.0. Finally, the program skips to the line following STD.DEV., and the standard deviations are read using format 17X, 10F10.0.

---

[1] STAT equals t_s(ND+NRP-NP', 1.0-α/2) of equation (10) to calculate individual linear confidence intervals. For other types of intervals, STAT may be set to other critical values. See instructions for FSTAT of program YCINT.

[2] The IY were used to construct figure 3 so all the values related to the parameter labeled K1 are grouped together, all the values related to K2 are grouped together, and so on.

[3] The BX were used to construct figure 3 so that plotted values were hydraulic-conductivity values. In these simulations, the estimated parameter values had been scaled in MODFLOWP so a value of 1.0 indicated that the value of the estimated hydraulic conductivity was equal to the prior information (see the third problem discussed in the section ' Computer program YR, Input'). Thus, the BX were the prior estimates of hydraulic conductivity.

[4] See Hill (1992, p. 191 and 220) for examples of this table, and see the section of the introduction "A problem with using parts of the MODFLOWP output file as input for YR, NORM, and BCINT".

```
Column 1 of input file PARi
|
   18       2.0                                                    data set 1
    1    1    1    1    0    1    1    1    1    1                data set 2
    1    1    1    1    1    1    1    1
    0    0    0    0    0    0    1    2    3    4                data set 3
    5    6    7    8    9   10   11   12
   1.   1.   1.   1.   1.   1. 437. 329. 165. 341.                data set 4
 288. 126. 193. 173. 380. 268.  82. 163.
 PARAMETER ID :       GHB       KRB       KRB       KRB       RCH       KV        T         T         T         T
                      T         T         T         T         T         T         T         T
 FINAL VALUES
                  0.174E+01 0.101E+02 0.111E+02 0.118E+02 0.356E-02-0.465E+01-0.165E-01-0.894E-02 0.314E-01-0.552E-01
                 -0.477E+00 0.353E+00 0.244E+00 0.760E-02-0.172E-01 0.103E+00 0.206E+00-0.137E+00
   EXPONENTIAL OF LN PARAMETERS
   (0.0 FOR UNTRANSFORMED PARAMETERS)
                  0.567E+01 0.239E+05 0.693E+05 0.135E+06 0.000E+00 0.959E-02 0.984E+00 0.991E+00 0.103E+01 0.946E+00
                  0.621E+00 0.142E+01 0.128E+01 0.101E+01 0.983E+00 0.111E+01 0.123E+01 0.872E+00
 STD. DEV.
                  0.607E+00 0.114E+00 0.234E+00 0.284E+00 0.615E-04 0.539E+00 0.197E+00 0.196E+00 0.198E+00 0.174E+00
                  0.164E+00 0.115E+00 0.945E-01 0.199E+00 0.187E+00 0.155E+00 0.123E+00 0.158E+00
```

Figure 6.-- Example of a PARi (i=0,1,...,or 9) input file for program BCINT. The data sets are described in table 1. [The section 'EXPONENTIAL OF LN PARAMETERS' is not printed by MODFLOWP if no parameters are log transformed. In this circumstance, all values in data set 2 of the PARi input file would equal zero]

Generally, it is helpful to have the parameter values plotted with different symbols than the confidence interval limits, as in figure 3. To accomplish this, the output from program BCINT needs to be coordinated with the plotting package used. Program BCINT inserts a blank line between the parameter estimates and the limits of the confidence intervals to coordinate with the nonproprietary plotting package by Turner (1992). To coordinate with another plotting package, changes can be made to the WRITE statements in and between the DO 120 and DO 130 loops.

COMPUTER PROGRAM YCINT FOR CALCULATING LINEAR CONFIDENCE INTERVALS ON SIMULATED HYDRAULIC HEADS AND FLOWS ALOGN HEAD DEPENDENT BOUNDARIES

Theory for YCINT

The program YCINT calculates linear confidence or prediction intervals on simulated values. Both types of intervals are discussed in the introduction of this report. Additional discussion is provided here to further clarify the distinction between confidence and prediction intervals.

Confidence intervals produced by YCINT are a function of the uncertainty in the parameters estimated using MODFLOWP, as expressed in the variance-covariance matrix on the parameters (Hill, 1992, p. 57) and the sensitivities of the simulated values to the estimated parameters (Hill, 1992, p. 90); uncertainties about other aspects of the system, such as parameters not estimated by MODFLOWP and the location of boundary conditions, are not reflected in the confidence intervals.

Confidence intervals have a specified probability of including the true, not the measured, value to which they relate; they do not account for measurement error. For example, a confidence interval on a hydraulic head value simulated for a certain time and place has a specified probability of including the true value of hydraulic head at that time and place. To test whether or not the true value is included within the calculated confidence interval, it might seem reasonable to measure the hydraulic head at the specified time and place and compare it to the confidence interval. Because any measurement contains measurement errors, however, the measured hydraulic head would have a smaller probability of

being included in the confidence interval than the probability specified for the confidence interval. A prediction interval with the same probability would usually be larger than the confidence interval. Measured values need to be compared against prediction intervals, not confidence intervals.

The term 'predict' is used in two ways in this report. As above, this term is used to specify 'prediction' intervals, which differ from confidence intervals in that they account for measurement errors. The term 'predict' also is used to specify (1) 'prediction' period, (2) 'prediction' conditions, and (3) 'predicted' values, which refer to (1) a period of time other than, and usually following, the calibration period, (2) the hydrologic conditions of the prediction period, and (3) the values (hydraulic heads and flows) simulated for the prediction period.

The confidence and prediction intervals calculated using YCINT are accurate only if (1) the parameters are normally distributed, (2) the model is correct, and (3) the model is roughly linear. See the comments about testing these assumptions in the section "Linear confidence and prediction intervals" in the introduction of this report.

Confidence and prediction intervals vary in size depending on how many intervals are to be considered simultaneously. For example, consider a situation in which management criteria state that drawdown at a certain location should not exceed 2 feet and that the flow to a stream along a certain reach should not be decreased by more than 20 percent. Confidence or prediction intervals can be constructed on the simulated drawdown and the change in flow to the stream to indicate the reliability of the simulated values. The intervals can be defined in two ways: (1) each of the two true values (for confidence intervals) or measured values (for predictive intervals) have a specified probability of being included in their respective intervals, regardless of whether the other is included in its interval; or (2) both of the true or measured values have a specified probability of being included in their respective intervals simultaneously. The first definition produces individual confidence or prediction intervals, and the second produces simultaneous confidence or prediction intervals.

The two types of intervals are different, and care needs to be taken to select the interval most appropriate for a given situation. Individual intervals generally are smaller than simultaneous intervals because one predicted value is considered without regard to what else is simulated. For example, if the only simulated value of interest is the drawdown at a certain location, an individual confidence interval could be used to indicate the uncertainty with which the drawdown is calculated.

Individual confidence intervals are calculated as (Draper and Smith, 1981, p. 29; Miller, 1981, p. 49; Hill, 1992, p. 58):

$$\hat{y}_m \pm s_{y_m} \times t_s(ND+NPR-NP', 1.0-\alpha/2) \qquad (11)$$

where $\hat{y}_m$ = the mth simulated value;

s_{y_m} = the standard deviation of $\hat{y}_m$, calculated as described below; and

$t_s(.,.)$ = the critical value of a random variable from the student-t probability distribution (see table in Hill, 1992, p. 60, or any basic statistics book), in which the first argument is degrees of freedom and α equals the probability that the confidence interval does not contain the true value.

The other variables are defined after equation (10).

The standard deviation of $\hat{y}_m$ is calculated as (after Cooley and Naff, 1990, p. 176):

$$s_{y_m} = \left[\sum_{i=1}^{NP'} \sum_{j=1}^{NP'} \frac{\partial \hat{y}_m}{\partial b_i} \underline{\underline{V}}(\underline{b}')_{ij} \frac{\partial \hat{y}_m}{\partial b_j} \right]^{1/2} \qquad (12)$$

where $\partial \hat{y}_m / \partial b_i$ = the partial derivative of $\hat{y}_m$ with respect to the ith parameter, b_i, evaluated at $\underline{b}'$, the optimal parameter values (also referred to as the sensitivity of $\hat{y}_m$ with respect to b_i); and

$\underline{\underline{V}}(\underline{b}')_{ij}$ = the element in the ith row and jth column of the variance-covariance matrix on the parameters (Hill, 1992, p. 57).

The two methods of calculating simultaneous confidence intervals presented in this report are named Bonferroni and Scheffé, after people involved in their development (Miller, 1981). Both methods are conservative with respect to α, the stated probability that the calculated interval does not contain the true value--that is, the stated probability is greater than or equal to the actual probability. The method that produces the smallest interval generally should be used, and guidelines for when each is smallest are provided below. However, Scheffé-simultaneous intervals need to be used when the number of intervals being considered can not be defined. For example, this would occur if a management criteria stated that drawdown over a certain area of an aquifer should not exceed 2 feet, and the location of maximum drawdown could not be identified based on considerations such as pumping well location. In this situation, intervals could theoretically be calculated at each location in the area involved, so that there would be an infinite number of intervals. Note that the imposition of a grid of a numerical model on the area can not be used to reduce the number of locations in the defined area to the number of grid nodes in the area.

Bonferroni-simultaneous confidence intervals are calculated as (Miller, 1981, p. 67-69):

$$\hat{y}_m \pm s_{y_m} \times t_B(ND+NPR-NP',\ 1.0-\alpha/2k) \qquad (13)$$

where k = the number of simultaneous confidence intervals (this would have been two in the example discussed previously if the management criteria on drawdown and decrease in flow to a stream were considered simultaneously); and

$t_B(.,.)$ = the critical value of a random variable from the Bonferroni-t probability distribution (table 2) in which the first argument is degrees of freedom and is designated as ν in table 2, and α equals the probability that the confidence interval does not contain the true value.

Table 2.--Critical values of the Bonferroni t statistic $t_B(\nu, 1.0-\alpha/2k)$

[Modified from Miller (1981, p. 238) with permission from publisher]

$\alpha = 0.5$																	
ν \ k	2	3	4	5	6	7	8	9	10	15	20	25	30	35	40	45	50
5	3.17	3.54	3.81	4.04	4.22	4.38	4.53	4.66	4.78	5.25	5.60	5.89	6.15	6.36	6.56	6.70	6.86
7	2.84	3.13	3.34	3.50	3.64	3.76	3.86	3.95	4.03	4.36	4.59	4.78	4.95	5.09	5.21	5.31	5.40
10	2.64	2.87	3.04	3.17	3.28	3.37	3.45	3.52	3.58	3.83	4.01	4.15	4.27	4.37	4.45	4.53	4.59
12	2.56	2.78	2.94	3.06	3.15	3.24	3.31	3.37	3.43	3.65	3.80	3.93	4.04	4.13	4.20	4.26	4.32
15	2.49	2.69	2.84	2.95	3.04	3.11	3.18	3.24	3.29	3.48	3.62	3.74	3.82	3.90	3.97	4.02	4.07
20	2.42	2.61	2.75	2.85	2.93	3.00	3.06	3.11	3.16	3.33	3.46	3.55	3.63	3.70	3.76	3.80	3.85
24	2.39	2.58	2.70	2.80	2.88	2.94	3.00	3.05	3.09	3.26	3.38	3.47	3.54	3.61	3.66	3.70	3.74
30	2.36	2.54	2.66	2.75	2.83	2.89	2.94	2.99	3.03	3.19	3.30	3.39	3.46	3.52	3.57	3.61	3.65
40	2.33	2.50	2.62	2.71	2.78	2.84	2.89	2.93	2.97	3.12	3.23	3.31	3.38	3.43	3.48	3.51	3.55
60	2.30	2.47	2.58	2.66	2.73	2.79	2.84	2.88	2.92	3.06	3.16	3.24	3.30	3.34	3.39	3.42	3.46
120	2.27	2.43	2.54	2.62	2.68	2.74	2.79	2.83	2.86	2.99	3.09	3.16	3.22	3.27	3.31	3.34	3.37
∞	2.24	2.39	2.50	2.58	2.64	2.69	2.74	2.77	2.81	2.94	3.02	3.09	3.15	3.19	3.23	3.26	3.29

Table 2.--Critical values of the Bonferroni t statistic $t_B(\nu, 1.0-\alpha/2k)$--continued

$\alpha = .01$																	
ν \ k	2	3	4	5	6	7	8	9	10	15	20	25	30	35	40	45	50
5	4.78	5.25	5.60	5.89	6.15	6.36	6.56	6.70	6.86	7.51	8.00	8.37	8.68	8.95	9.19	9.41	9.68
7	4.03	4.36	4.59	4.78	4.95	5.09	5.21	5.31	5.40	5.79	6.08	6.30	6.49	6.67	6.83	6.93	7.06
10	3.58	3.83	4.01	4.15	4.27	4.37	4.45	4.53	4.59	4.86	5.06	5.20	5.33	5.44	5.52	5.60	5.70
12	3.43	3.65	3.80	3.93	4.04	4.13	4.20	4.26	4.32	4.56	4.73	4.86	4.95	5.04	5.12	5.20	5.27
15	3.29	3.48	3.62	3.74	3.82	3.90	3.97	4.02	4.07	4.29	4.42	4.53	4.61	4.71	4.78	4.84	4.90
20	3.16	3.33	3.46	3.55	3.63	3.70	3.76	3.80	3.85	4.03	4.15	4.25	4.33	4.39	4.46	4.52	4.56
24	3.09	3.26	3.38	3.47	3.54	3.61	3.66	3.70	3.74	3.91	4.04	4.1[1]	4.2[1]	4.3[1]	4.3[1]	4.3[1]	4.4[1]
30	3.03	3.19	3.30	3.39	3.46	3.52	3.57	3.61	3.65	3.80	3.90	3.98	4.13	4.26	4.1[1]	4.2[1]	4.2[1]
40	2.97	3.12	3.23	3.31	3.38	3.43	3.48	3.51	3.55	3.70	3.79	3.88	3.93	3.97	4.01	4.1[1]	4.1[1]
60	2.92	3.06	3.16	3.24	3.30	3.34	3.39	3.42	3.46	3.59	3.69	3.76	3.81	3.84	3.89	3.93	3.97
120	2.86	1.99	3.09	3.16	3.22	3.27	3.31	3.34	3.37	3.50	3.58	3.64	3.69	3.73	3.77	3.80	3.83
∞	2.81	2.94	3.02	3.09	3.15	3.19	3.23	3.26	3.29	3.40	3.48	3.54	3.59	3.63	3.66	3.69	3.72

[1] Obtained by graphical interpolation.

Scheffé-simultaneous confidence intervals are calculated as (Miller, 1981, p. 50; Cooley and Naff, 1990, p. 175-176):

$$\hat{y}_m \pm s_{y_m} \times [d \times F_\alpha(d, ND+NPR-NP')]^{1/2} \quad (14)$$

where d = k or NP', whichever is less; and

$F_\alpha(.,.)$ = the critical value of a random variable from the F probability distribution (see table in Cooley and Naff, 1990, p. 46-47 or any basic statistics book) in which the first and second arguments are degrees of freedom and α equals the probability that the confidence interval does not contain the true value.

The Bonferroni-simultaneous confidence interval is smaller than the Scheffé-simultaneous confidence interval when the critical value $t_B(ND+NPR-NP', 1.0-\alpha/2k)$ is less than the value of $[d \times F_\alpha(d, ND+NPR-NP')]^{1/2}$. Generally, the Bonferroni critical value is smaller than $[d \times F_\alpha(d, ND+NPR-NP')]^{1/2}$ unless k (the number of simultaneous confidence intervals) is much larger than NP' (the number of parameters estimated in the regression) (Miller, 1981, p. 69).

Prediction intervals also can be calculated as individual, Bonferroni-simultaneous, or Scheffé-simultaneous intervals. Individual prediction intervals are calculated as (Draper and Smith, 1981, p. 29; Miller, 1981, p. 49; Hill, 1992, p. 58):

$$\hat{y}_m \pm [s^2_{y_m} + s^2/\omega_m]^{1/2} \times t_s(ND+NPR-NP', 1.0-\alpha/2) \quad (15)$$

where s^2 = the calculated error variance of the regression; and

ω_m = a weight that equals σ^2/σ^2_m, where σ^2 is the estimated common error variance of the regression that should be close to s^2 (Hill, 1992, p. 39), and where σ^2_m is the variance of the error with which y_m is measured.

The term s^2/ω_m approximates the variance of the measurement error. This approximation is from Cooley and Naff (1990, p. 176), and it is valid only if the error with which y_m is measured is independent of all other measurement errors.

Bonferroni-simultaneous prediction intervals are calculated as (Miller, 1981, p. 67-69):

$$\hat{y}_m \pm [s^2_{y_m} + s^2/\omega_m]^{1/2} \times t_B(ND+NPR-NP', 1.0-\alpha/2k). \quad (16)$$

Scheffé-simultaneous prediction intervals are calculated as (Miller, 1981, p. 50; Cooley and Naff, 1990, p. 175-176):

$$\hat{y}_m \pm [s^2_{y_m} + s^2/\omega_m]^{1/2} \times [k \times F_\alpha(k, ND+NPR-NP')]^{1/2}. \quad (17)$$

Note that for prediction intervals, k replaces d.

As with confidence intervals, the two procedures of calculating simultaneous prediction intervals are nearly identical, and the procedure that produces the smallest interval should be used. For $\alpha = 0.05$, this procedure always will be the Bonferroni procedure (Miller, 1981, p. 116).

The simulated values for which confidence and prediction intervals can be calculated with YCINT are presented in the following paragraphs. To improve clarity, only confidence intervals are mentioned, however, prediction intervals also can be calculated for the simulated values included in the discussion.

With YCINT, linear confidence intervals can be calculated for the simulated values of hydraulic head at arbitrary times and locations and flows along head-dependent boundaries that are described in Hill (1992, p. 19-31). For example, figure 7A shows values of steady-state hydraulic heads and flows into a river simulated under calibration conditions and under two pumping scenarios, and figure 7B shows hydraulic heads simulated under transient calibration and predictive conditions. Linear confidence intervals can be calculated for any of these simulated values.

In many management situations, the value of interest is how much hydraulic heads or flows along head-dependent boundaries would change under certain conditions. For example, figure 7A shows a cross section through a ground-water flow system with hydraulic heads and flow to a

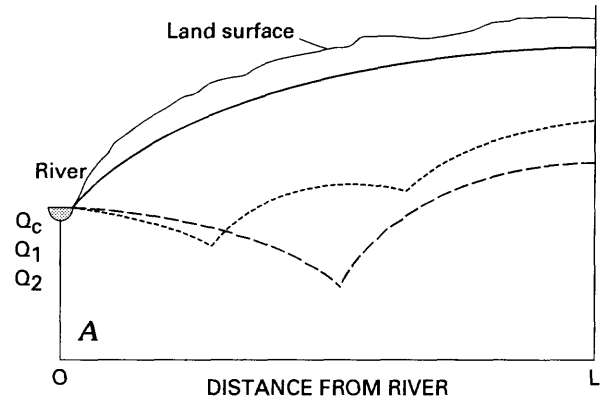
river that were simulated for steady-state calibration conditions and for steady-state conditions representing two pumping scenarios.

If the management criterion for the system in figure 7A is that drawdown cannot exceed 2 feet, simulated drawdown is of interest. Using the calibrated model, simulated hydraulic heads for the calibration conditions could be subtracted from simulated hydraulic heads for pumping scenarios 1 or 2 to produce simulated drawdowns. YCINT can be used to calculate the linear confidence interval on the maximum drawdown.

If the management criterion for the system in figure 7A is that the streamflow gain must not be decreased by more than 20 percent of an observed flow along a reach of the stream, the simulated change in streamflow is of interest. Using the calibrated model, the streamflow gain simulated for the calibration conditions could be subtracted from the streamflow gain simulated for pumping scenarios 1 and 2. YCINT can be used to calculate the linear confidence interval on the difference between these simulated streamflow gains.

To provide a consistent terminology, changes in hydraulic heads and changes in streamflow are called 'differences', and the calculated confidence intervals are called 'confidence intervals on the differences'. The simulation that produces the values from which other values are subtracted is referred to as 'the predictive simulation'; the simulation that produces the values that are subtracted is referred to as the 'base simulation'.

Differences need not be between calibration conditions and a pumping scenario. For example, if the differences between hydraulic heads simulated under pumping scenarios 1 and 2 of figure 7A are of interest, the hydraulic heads for the two pumping scenarios can be calculated using the calibrated model, and the differences and related confidence intervals can be calculated with YCINT. Note, however, that the confidence intervals on the differences generally will get smaller as the base and predictive simulations become more similar. This decrease occurs because the sensitivities related to the quantities being subtracted get closer to each other. See the last paragraph of this section for a discussion of this situation.



EXPLANATION

Simulated hydraulic-head distribution for:

——— Calibration conditions

– – – – Pumping scenario 1

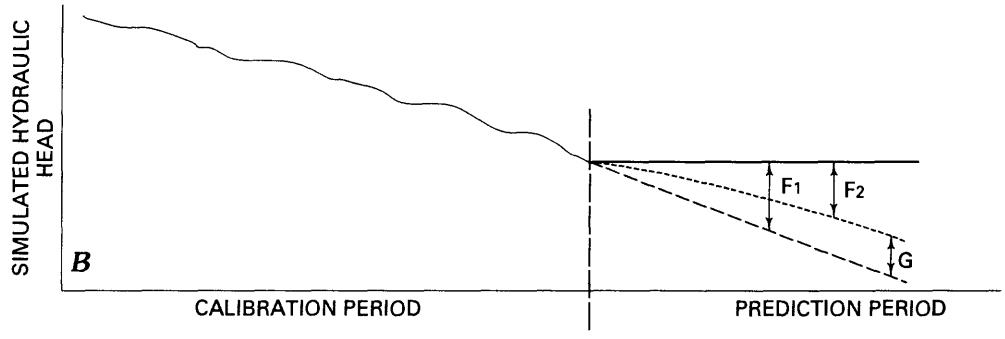
····· Pumping scenario 2

Simulated flow to the river for:

Q_c Calibration conditions

Q_1 Pumping scenario 1

Q_2 Pumping scenario 2



EXPLANATION

Simulated hydraulic head for calibration period

——— Simulated hydraulic head at end of calibration period

– – – – Simulated hydraulic head for scenario 1

····· Simulated hydraulic head for scenario 2

F1, F2, G Examples of the types of differences that may be of interest

Figure 7.-- Diagrams showing quantities for which confidence intervals can be calculated with program YCINT for (A) steady-state calibration and predictive simulations and (B) transient calibration and predictive simulations.

Transient calibration and prediction periods produce a variety of 'differences' that may be of interest. For example, a simulated hydrograph from a transient calibration period followed by a transient prediction period is shown in figure 7B. Differences which may be of interest are (1) decline in hydraulic head since the end of the calibration period (difference F_1 and F_2 of fig. 7B), or (2) the difference in the hydraulic head that would occur with and without proposed additional pumpage (difference G of fig. 7B).

When confidence or prediction intervals are calculated on differences, $\hat{y}_m$ of equations 11 through 17 equals:

$$\hat{y}_m = \hat{y}_{p_m} - \hat{y}_{q_m}, \tag{18}$$

where p represents the predictive conditions and q represents the base conditions. The standard deviation of $\hat{y}_m$, s_{y_m} of equations 11 through 17, is calculated as:

$$s_{y_m} = \left[\sum_{i=1}^{NP'} \sum_{j=1}^{NP'} \left(\frac{\partial \hat{y}_{p_m}}{\partial b_i} - \frac{\partial \hat{y}_{q_m}}{\partial b_i} \right) V(\underline{b})_{ij} \left(\frac{\partial \hat{y}_{p_m}}{\partial b_j} - \frac{\partial \hat{y}_{q_m}}{\partial b_j} \right) \right]^{1/2}, \tag{19}$$

where all terms were defined for equation (12) or (18). The approximation of the variance of the measurement error, s^2/ω_m of equation (15), (16), and (17), is calculated as:

$$s^2/\omega_m = s^2/\omega_{p_m} + s^2\omega_{q_m} \tag{20}$$

Care must be taken when calculating confidence and prediction intervals on differences because intervals on differences are calculated using differences in the sensitivities (eq. 19). If the sensitivities to each of the parameters are the same for the two subtracted values, s_{y_m} of equation (19) will equal 0.0 and the limits of calculated confidence intervals on differences will equal the the simulated difference. As a result, the width of these confidence intervals will equal zero and prediction intervals will reflect only the σ^2/ω_m measurement error term.

An unrealistic but illustrative example is if all conditions, including stresses, are the same in the two simulations for which differences are calculated. In this situation, all simulated differences would equal 0.0, and the confidence interval limits on differences would equal 0.0. This result indicates that it is certain that if nothing in the system changes, the hydraulic heads and flows will not change, which is a direct result of using the deterministic ground-water flow equation. These results also indicate that if the differences between two simulations are small, the confidence intervals on the differences also will tend to be small.

Another situation that produces s_{y_m}=0.0 occurs if the sensitivities related to all estimated parameters are independent of whatever is changed between the two simulations for which differences are calculated. For example, consider a simulation in which all model layers are confined and all boundary conditions are linear (this excludes some situations in which the River, Drain, and Evapotranspiration Packages of MODFLOW are used; McDonald and Harbaugh, 1988, ch. 6, 9, and 10). In such a simulation, a 1.0 increase in areal recharge produces the same increase in hydraulic head or flow at any specified location in the system regardless of the simulated pumpage (this can be verified with any simulation that satisfies the stated criteria). This situation indicates that the sensitivities related to areal recharge (which, in this situation, equal the changes in hydraulic heads or flows at the specified locations divided by the 1.0 change in areal recharge), are independent of pumpage. If areal recharge is the only parameter and confidence intervals are calculated for differences between two simulations that differ only in the pumpage, the difference between the sensitivities would equal zero and, therefore, s_{y_m} would equal zero. A more practical consequence of this result is that in a simulation in which many parameter values have been estimated by regression, for differences, s_{y_m} only will reflect the uncertainty and correlation of parameters for which the sensitivities are dependent on whatever is changed between the simulations for which the differences are calculated.

Input for YCINT

Two input files are always required by YCINT: RESANP.DAT and YCINT.DAT. An additional input file, YCINT.DAT2, is needed if intervals are to be calculated on differences. The three input files are produced by MODFLOWP when used as described in the following paragraphs.

If the first three input files are needed, it is suggested that RESANP.DAT be created first, then YCINT.DAT2, and, finally, YCINT.DAT. Creating the files in this order is more convenient because the needed changes to the MODFLOWP input files are progressive.

RESANP.DAT is identical to the data set required by the program RESANP (in Hill, 1992, p. 224 it is called RESAN.MODP). YCINT reads the variance-covariance matrix on the estimated parameters from RESANP.DAT. If RESANP.DAT was not produced in the final calibration run, it can be produced by executing MODFLOWP as follows [The LINE and DATA SET numbers cited refer to parts of the Parameter-Estimation Package input file (Hill, 1992, p. 131-150)]:

1. Substitute the final calibrated parameter values from file IOUB (LINE 7) into DATA SET 8. If these values are not available, set IOUB equal to a FORTRAN unit number and continue with the following instructions. Save the file referred to by the FORTRAN unit number IOUB.

2. Set IPAR=1 (LINE 6).

3. Set IOUR (DATA SET 13) equal to the FORTRAN unit number of RESANP.DAT.

4. In all other respects, the input file for the Parameter-Estimation Package and all other input files need to be the same as they were for calibration.

5. Execute MODFLOWP.

YCINT.DAT provides information about the simulated conditions for which confidence or prediction intervals are being calculated. If confidence or prediction intervals on the differences are to calculated, YCINT.DAT provides information about the final changed conditions. To create YCINT.DAT, MODFLOWP needs to be executed as follows [The LINE and DATA SET numbers cited refer to parts of the Parameter-Estimation Package input file (Hill, 1992, p. 131-150)]:

1. If not done in step 1 for RESANP.DAT, substitute the final calibrated parameter values from the file referred to by IOUB (LINE 7) into DATA SET 8.

2. Set IPAR=1 (LINE 6).

3. Modify DATA SETS 6 and 7 and related values on LINE 6 to define only the quantities for which confidence or prediction intervals are to be calculated. For example, if confidence intervals are to be calculated on five values of simulated hydraulic head and one flow, on LINE 6 NH needs to equal five, MOBS and MAXM need to be nonzero if any of the simulated hydraulic heads are multilayer (see instructions for DATA SETS 6 and 6A), NQ and NQT need to each equal 1, and NQC needs to equal the number of cells involved in calculating the flow (see instructions for DATA SET 7A). If confidence or prediction intervals on differences are to be calculated, DATA SETS 6 and 7 need to define the $\hat{y}_{p_m}$ of equation (18); see the discussion following the instructions for YCINT.DAT2. HOBS is ignored, so it does not matter what values are used; if prediction intervals are to be calculated, IWT needs to equal 0 and WT needs to equal s^2/ω_m (eqs. 15, 16, or 17) or, if differences are calculated, s^2/ω_{p_m} (eq. 20). If DATA SET 6C is used and ITT of DATA SET 6B equals 1, WT_h needs to equal s^2/ω_m, or if differences are calculated, s^2/ω_{p_m}. If ITT equals 2, WT_{DD} needs to equal s^2/ω_m; ITT generally should not equal 2 if differences are calculated. In these situations, IWT needs to equal 0.

4. A new variable, FSTAT, has been added to DATA SET 13. FSTAT needs to be placed on a new second line of DATA SET 13 and is read using F10.0 format. FSTAT needs to be set to the appropriate critical value: FSTAT = t_S(ND+NPR-NP', 1.0-α/2) to calculate individual confidence and prediction intervals, FSTAT = t_B(ND+NPR-NP', 1.0-α/2k) to calculate Bonferroni-simulataneous confidence and prediction intervals, FSTAT = [d × F_α(d, ND+NPR-NP')]$^{1/2}$ to calculate Scheffe'-simultaneous confidence intervals, and FSTAT = [k × F_α(k, ND+NPR-NP')]$^{1/2}$ to calculate Scheffe'-simultaneous prediction intervals. See the section 'Theory for YCINT' for a discussion of these critical values and the symbols used. Changing the value of FSTAT does not require reexecuting MODFLOWP: See point 9 of this list.

5. Set DMAX=1×10^{-6} and TOL=1×10^{12} (DATA SET 13) so the final calibrated parameter values in DATA SET 8 remain unchanged.

6. A new variable IOUE has been added to LINE 7; it follows IUNHEA in I5 format. Set IOUE equal to the FORTRAN UNIT number of YCINT.DAT; set IOUR (DATA SET 13) equal to zero.

7. Change files from other packages to represent the stresses, boundary conditions, and so on of the system for which predictions are being made.

8. Execute MODFLOWP to produce YCINT.DAT.

9. Changes need to be made to the first line of YCINT.DAT as follows:

   (a) Values for variables IDIF and IPRED need to be added to the first line of YCINT.DAT using 2I5 format. If confidence or prediction intervals for differences are to be calculated, set IDIF to 1; otherwise, set it to 0. If prediction intervals are being calculated, set IPRED to 1; otherwise, set it to 0.

   (b) The real number preceding IDIF on the first line of YCINT.DAT equals FSTAT, which was discussed in point 4 above. If the user wishes to change FSTAT, MODFLOWP need not be reexecuted. Up to

this point, FSTAT has not been used in any calculations, and it can simply be changed in YCINT.DAT. FSTAT occupies columns 41 to 50 of the first line of YCINT.DAT, and is read by YCINT using format F10.0.

If confidence or prediction intervals for differences are to be calculated (IDIF=1), a third input file, YCINT.DAT2, is needed. YCINT.DAT2 provides information about the base conditions. To create YCINT.DAT2, MODFLOWP needs to be executed as follows [The LINE and DATA SET numbers cited refer to parts of the Parameter-Estimation Package input file (Hill, 1992, p. 131-150)]:

1-2. Follow steps 1 and 2 for YCINT.DAT.

3. Modify DATA SETS 6 and 7 created for YCINT.DAT to define the $\hat{y}_{q_m}$ of equation (18); see the discussion below. HOBS is ignored, so it does not matter what values are used for HOBS; WT needs to equal s^2/ω_{q_m} (eq. 20) and IWT needs to equal 0. If DATA SET 6C is used and ITT of DATA SET 6B equals 1, WT_h needs to equal s^2/ω_{q_m} and IWT needs to equal 0. ITT generally should not equal 2 if diffeences are calculated.

4-5. Follow steps 4 and 5 for YCINT.DAT.

6. Set IOUR (DATA SET 13) equal to the FORTRAN UNIT number of YCINT.DAT2; set IOUE (LINE 7) equal to zero.

7. Change input files from other packages to represent the stresses, boundary conditions, and so on of the base conditions. If the base simulation represents calibration conditions the input files used in model calibration can be used.

8. Execute MODFLOWP.

Differences are calculated by subtracting the YCINT.DAT2 hydraulic heads and flows at head-dependent boundaries (the $\hat{y}_{q_m}$'s of eq. 18) from the YCINT.DAT values (the $\hat{y}_{p_m}$'s of eq. 18) in the order they are listed in DATA SETS 6 and 7 of those runs (see step 3). Therefore, the DATA SETS 6 and 7 need to be constructed so the values listed first in each run are to be subtracted from each other, the values listed second in each run are to be subtracted from each other, and so on. A problem occurs if, for example, the 15th value listed for YCINT.DAT is a flow along a head-dependent boundary, and the 15th value for YCINT.DAT2 is a hydraulic-head value. In this situation, the computer program YCINT will calculate the difference and related confidence interval as if the difference made sense. It is up to the user to discover that there was a mistake in how the two runs were constructed. To aid in this effort, the data identifiers (DID) from DATA SETS 6 and 7 are printed in the YCINT output file (fig. 8).

Once the input files have been created, YCINT can be executed after checking the FORTRAN to be sure that the values defined in the PARAMETER statement at the top of YCINT are large enough, and recomputing and loading the program if changes are made. NDD needs to equal or exceed NH+NQT (LINE 5); NPD needs to equal or exceed NP (LINE 3). Execution of YCINT produces YCINT.OUT, which is described below.

Output for YCINT

YCINT.OUT is composed of seven sections, which are briefly described in table 3. An example of the table of section G in which differences and their related confidence intervals are included is shown in figure 8. The first two columns show the sequential number and data identifiers from YCINT.DAT. The third column shows the simulated value under predictive conditions. The fourth column shows the standard deviation of the value simulated under predictive conditions, and the fifth and sixth columns show the upper and lower limits of the linear confidence or prediction interval on the simulated value.

Table 3.-- The seven sections of YCINT.OUT

[When differences and related intervals are calculated, base conditions are the conditions represented in the simulation that produces the values that are subtracted, and predictive conditions are the conditions represented in the simulation that produces the values from which subtraction occurs. When differences are not calculated, the 'predictive conditions' of sections E and F may be the calibration conditions, if confidence or prediction intervals on calibrated values are to be calculated]

Section	Description
A	The number of values for which confidence or prediction intervals are to be calculated, the number of estimated parameters, the value of FSTAT, a statement of whether differences and associated intervals are calculated, and a statement of whether or not confidence or prediction intervals are to be calculated. The value of FSTAT is the only indication in YCINT.OUT as to whether the calculated intervals are individual, Bonferroni-simultaneous, or Scheffe-simultaneous; it is also printed as part of section G.
B	The variance-covariance matrix for the estimated parameters.
(Sections C and D are omitted if differences are not calculated.)	
C	The values of the quantities defined by DATA SETS 6 and 7 calculated using optimum parameter values for the base conditions. Data identifiers also are listed.
D	Sensitivities for the optimum parameters for base conditions.
E	The values of the quantities defined by DATA SETS 6 and 7 calculated using optimum parameter values for predictive conditions. Data identifiers also are listed.
F	Sensitivities for the optimum parameters under predictive conditions.
G	A table that includes simulated values and their standard deviations and confidence or prediction intervals and data identifiers. FSTAT is printed just before this table.

To graph the confidence intervals from section G, the numbers need to be rearranged to be compatible with a graphing routine. Because it was anticipated that these numbers often would be used in tables instead of figures, and because a figure design that was expected to be useful in most circumstances was not apparent, a data set suitable for graphing is not produced by YCINT.

CRITICAL VALUE FOR THE INTERVALS = 1.7010

NO.	ID	SIMULATED VALUE	STD. DEV.	CONFIDENCE INTERVAL			ID	SIMULATED DIFFERENCE	STD. DEV.	CONFIDENCE INTERVAL		
1	1	9.64146	0.766377E-01	9.15194	;	10.1310	1	-2.48804	0.510111E-01	-2.16221	;	-2.81387
2	1	9.30883	0.652679E-01	8.89193	;	9.72573	1	-2.53597	0.498055E-01	-2.21784	;	-2.85410
3	2	9.47599	0.728768E-01	9.01049	;	9.94149	2	-2.33941	0.485737E-01	-2.02915	;	-2.64967
.												
.												
.												
48	24	8.13970	0.516148E-01	7.81001	;	8.46939	24	-1.92200	0.379360E-01	-1.67968	;	-2.16432
49	25	4.01142	0.436420E-01	3.73266	;	4.29018	25	-0.748320	0.266246E-01	-0.578256	;	-0.918385
50	25	4.01891	0.438386E-01	3.73889	;	4.29893	25	-0.751900	0.266767E-01	-0.581503	;	-0.922297
51	26	4.74146	0.874126E-01	4.18311	;	5.29981	26	-2.77515	0.642564E-01	-2.36471	;	-3.18559
52	26	4.73484	0.873998E-01	4.17657	;	5.29311	26	-2.77527	0.642628E-01	-2.36479	;	-3.18575
53	27	9.43885	0.822693E-01	8.91336	;	9.96434	27	-2.47345	0.562133E-01	-2.11439	;	-2.83251
54	27	8.60850	0.856080E-01	8.06168	;	9.15532	27	-3.01200	0.798011E-01	-2.50227	;	-3.52173
55	DD	9.43885	0.822693E-01	8.91336	;	9.96434	DD	-2.47345	0.562133E-01	-2.11439	;	-2.83251
56	S1	-72835.9	3747.87	-96775.4	;	-48896.4	S1	24313.0	1155.20	31691.8	;	16934.2
57	S2	-45853.4	3253.29	-66633.8	;	-25073.0	S2	16348.4	1086.12	23286.0	;	9410.84
58	S3	-51499.4	2979.84	-70533.1	;	-32465.7	S3	19792.5	735.779	24492.3	;	15092.7
.												
.												
.												
70	S16	-23599.4	1011.93	-30063.1	;	-17135.7	S16	-20276.2	463.469	-23236.6	;	-17315.8
71	S17	-24552.8	1082.98	-31470.3	;	-17635.3	S17	-19354.7	773.969	-24298.4	;	-14411.0
72	S18	-26632.5	1306.03	-34974.7	;	-18290.3	S18	-18604.2	1079.35	-25498.6	;	-11709.8

Figure 8.--An example of section G of YCINT.OUT.

The remaining columns are related to the calculated differences and are omitted if IDIF = 0 in YCINT.DAT (see item 9 in the instructions for YCINT.DAT). The seventh column shows the data identifier from YCINT.DAT2. The eight column shows the simulated difference; the ninth column shows the standard deviation of the difference; and the tenth and eleventh columns show the lower and upper limits of the linear confidence or prediction interval on the difference.

COMPUTER PROGRAM BEALEP FOR CALCULATING THE MODIFIED BEALE'S MEASURE OF MODEL LINEARITY

Theory for BEALEP

As mentioned in the introduction of this report, the modified Beale's measure can be used to indicate whether a ground-water flow model is roughly linear for parameter values close to the optimized parameter values. A model needs to be roughly linear for the linear confidence intervals produced by computer programs BCINT and YCINT, which are presented in this report, to be accurate.

The modified Beale's measure is calculated by generating sets of parameter values close to the optimized parameter values using the variances and covariances of the optimized parameter values. The number of parameter sets generated equals twice the number of parameters. In the generated sets, parameter values corresponding to parameters with small variances generally are relatively near the optimized values, while parameter values corresponding to parameters with larger variances are relatively far from the optimized values in at least some of the generated sets. The generated sets are calculated using equation 5.6-14 of Cooley and Naff (1990, p. 174)

Once the parameter sets are generated, they are used to calculate hydraulic heads and flows in two ways. First, each parameter set is read into MODFLOWP and the ground-water flow equation is solved. The simulated values of hydraulic heads and flows at observation points generated from this solution are referred to as $\tilde{y}_{\ell i}$, where i refers to the ith observed hydraulic head or flow used in the regression and ℓ indicates that the ℓth generated parameter set was used. A problem that can occur at this step

is that some parameter sets may create situations for which the ground-water flow equation is difficult or impossible to solve. If solving the flow equation is difficult, so that the information printed by the solver indicates that convergence is being reached slowly or that the convergence criteria have nearly been reached, a solution may be obtainable by increasing the number of solver iterations or by increasing the convergence criteria. Increasing the convergence criteria by as much as a factor of five generally will not have a major impact on the calculated value of the modified Beale's measure. Solving the flow equation is impossible if any of the generated parameter values produce negative values of aquifer transmissivity, vertical conductance, or conductance along head-dependent boundaries. To correct this problem, the parameter(s) involved need to be log-transformed (see instructions for DATA SET 2 of the Parameter-Estimation Package input file, Hill, 1992, p. 136) the regression needs to be reexecuted, and the parameter sets for the modified Beale's measure need to be regenerated. If the $\tilde{y}_{\ell i}$ can not be calculated, it must be assumed that the model is effectively nonlinear.

Each generated parameter set also is used to calculate hydraulic heads and flows using the assumption that the ground-water flow equation is linear. The ith linearized hydraulic head or flow, represented by the symbol $\tilde{y}^{o}_{\ell i}$, where the superscript o indicates that a linear model is used, is calculated as:

$$\tilde{y}^{o}_{\ell i} = \sum_{j=1}^{NP} (b'_j - \tilde{b}_{\ell j}) \left.\frac{\partial \hat{y}_i}{\partial b_j}\right|_{\underline{b}'} \qquad \ell=1,\ 2\times NP' \qquad (21)$$

where NP' equals the number of parameters, b'_j is the value of the jth optimized parameter value, $\tilde{b}_{\ell j}$ is the value of the jth parameter value from the ℓth generated parameter set, $\partial \hat{y}_i / \partial b_j$ is the partial derivative of the simulated value of the ith hydraulic head or flow with respect to the jth parameter, and the derivatives are calculated using the optimized parameter values which, in vector form, are expressed as $\underline{b}'$. The derivatives are sensitivities, and are calculated as discussed in Hill (1992, p. 90-94).

Once the above calculations have been completed, the modified Beale's measure is calculated using equation 6.2-3 of Cooely and Naff (1990, p. 188). Using the notation presented above, this equation is equivalent to:

$$\hat{N}_b = \frac{\sum_{\ell=1}^{2\times NP'} \sum_{i=1}^{ND} (\tilde{y}_{\ell i} - \tilde{y}^{o}_{\ell i})\, \omega_i (\tilde{y}_{\ell i} - \tilde{y}^{o}_{\ell i})}{\sum_{\ell=1}^{2\times NP'} \sum_{i=1}^{ND} (\tilde{y}^{o}_{\ell i} - \hat{y}_i)\, \omega_i (\tilde{y}^{o}_{\ell i} - \hat{y}_i)} \qquad (22)$$

where ND equals the number of observations of hydraulic head and flow, $\hat{y}_i$ are hydraulic heads and flows simulated using the optimized parameter values, and all other terms were defined above. The derivation of the modified Beale's measure and its critical values are discussed by Cooley and Naff (1990, p. 187-189).

In some situations, the modified Beale's measure may not adequately indicate whether or not a model is roughly linear. For example, if a model is calibrated using only hydraulic-head observations, the $\hat{y}_i$, $\tilde{y}_{\ell i}$, and $\tilde{y}^{o}_{\ell i}$ of equation (22) all relate to hydraulic heads, and the calculated value of the modified Beale's measure could not be used to indicate whether or not the model was roughly linear for the calculation of a linear confidence or prediction interval on a simulated flow. In general, if the quantities for which linear confidence or prediction intervals are to be calculated are very different than the observations used in the regression, the modified Beale's measure may be an inadequate measure of model linearity.

To calculate the modified Beale's measure of model linearity as described by Cooley and Naff (1990, p. 187-189), subroutine PAR1BE has been added to MODFLOWP. PAR1BE produces BEALE.DAT and BEALE.DAT2, input files for the computer program BEALEP. BEALEP is a modified version of a code presented by Cooley and Naff (1990, p.191-198). Instructions for producing the two input files follow. The LINE and DATA SET numbers cited refer to parts of the Parameter-Estimation Package input file (Hill, 1992, p. 131-150).

Input for BEALEP

First, the two variables added to the Parameter-Estimation Package input file and introduced in items 4 and 6 of the instructions for creating YCINT.DAT also are used to calculate the modified Beale's measure. IOUE is the FORTRAN UNIT number of BEALE.DAT and is read from the end of LINE 7 using I5 format. Therefore, LINE 7 now includes five integers read using I5 format.

The second new variable is FSTAT, which is added to a new second line of DATA SET 13 (Hill, 1992, p. 150) and is read using F10.0 format. FSTAT is the value of the F-distribution statistic required to calculate the different parameter sets that are then used to calculate the modified Beale's measure, F_{α}(NP',ND+NPR-NP') (Cooley and Naff, 1990, eqs. 5.6-12 and 5.6-14, p. 173-174). For a significance level of 0.05, values of FSTAT can be determined using table 2.10.3 of Cooley and Naff (1990, p. 46). For example, with seven parameters (NP'=7) and 127 observations (ND=127; NPR=0), FSTAT=2.02.

Two input files are needed by BEALEP: BEALE.DAT and BEALE.DAT2. The steps required to create input files BEALE.DAT and BEALE.DAT2 are as follows, starting with the files used for model calibration [The LINE and DATA SET numbers cited refer to parts of the Parameter-Estimation Package input file (Hill, 1992, p. 131-150)]:

1. Set IPAR=1, IOUE equal to the unit number for BEALE.DAT, and FSTAT equal to F_{α}(NP',ND+NPR-NP'). If the regression has been completed and the purpose of this run is just to create the input file for BEALEP.F77, the following procedures will make this run take less time.
   (a) Copy the final parameter values from the IOUB file (Hill, 1992, p. 134) into DATA SET 8.
   (b) If prior information is specified using DATA SETS 8 and 10, step (a) will have changed your prior parameter estimates. To restore them (which is needed to calculate the modified Beale's measure), set the entries in DATA SET 10 to 0.0, and specify the prior information using repetitions of DATA SET 12 (see Hill, 1992, p. 230).

2. Execute MODFLOWP. BEALE.DAT will be written on FORTRAN unit IOUE.

3. In BEALE.DAT, the generated parameter sets for the modified Beale's measure follow the statement 'THE PARAMETER SETS FOLLOW'. In the following steps, these parameter sets are read from BEALE.DAT and are used to solve for hydraulic heads and flows along head-dependent boundaries defined in DATA SETS 6 and 7. The simulated values of hydraulic heads and flows are written to BEALE.DAT2.

4. Set IPAR=-1 on LINE 6.

5. Configure the Output Control input file to avoid printing or saving heads or drawdowns. This is necessary because hydraulic heads will be calculated (and would be printed and saved if specified in the Output Control input file) for all the parameter sets in BEALE.DAT.

6. BEALE.DAT will be read instead of written to in step 8 below. On some computers, this means the file definition may need to be changed. The FORTRAN UNIT number of BEALE.DAT is still IOUE.

7. Make any arrangements necessary on your computer to write to BEALE.DAT2 as FORTRAN UNIT number IOUE+1.

8. Execute MODFLOWP.

9. Unless you plan to repeat the calculation of the modified Beale's measure, set IOUE to 0 and reset the Output Control input file to print heads and drawdowns as you prefer.

The input and output unit numbers used by BEALEP are defined in the three lines following the format statements. Set IIN equal to the FORTRAN unit number of BEALE.DAT, set IIN2 to the FORTRAN unit number of BEALE.DAT2, and set IOUT equal to the FORTRAN unit number of BEALE.OUT. All files are opened within the program.

BEALEP can be executed after checking to be sure that the values defined in the PARAMETER statement at the top of BEALEP are large enough. NDD needs to equal or exceed NH+NQT of LINE 5 of the Parameter-Estimation Package input file (Hill, 1992, p. 132), NPD needs to equal or exceed NPD of LINE 3, and MPD needs to equal or exceed MPR of LINE 5. Execution of BEALEP produces BEALE.OUT, which is described below.

Output for BEALEP

BEALE.OUT includes the modified Beale's measure statistic and its critical values and some background information by which the source of model nonlinearity can be identified. BEALE.OUT also includes the statistics from the top right column of Cooley and Naff (1990, p.174). BEALE.OUT is composed of 11 sections, which are briefly described in table 4. Examples of sections B, I, J, and K are shown in figure 9. As noted in table 4, sections B and I can be used to identify the source of model nonlinearity.

Table 4.-- The 11 sections of BEALE.OUT

Section	Description
A	Values of NP', NRES (usually equals NP'; see Cooley and Naff, 1990, p. 187-197, for uses; to change NRES to a smaller value, create BEALE.DAT using FSTAT=F_{α}(NRES,ND+NPR-NP') and proceed as above), NH+NQT, NPR (number of nonzero values in DATA SET 10; needs to equal 0), MPR, NPTS (2×NRES), calculated error variance from the regression, FSTAT (equals FSTAT of DATA SET 13).
B	Optimized parameter values from the regression.
C	Hydraulic heads and flows defined in DATA SETS 6 and 7 simulated using the optimized parameter values.
D	Observed values of the hydraulic heads and flows of DATA SETS 6 and 7.
E	Weights of the observations.
F	Sensitivities calculated using the optimized parameter values.
G	Prior information on the parameters.
H	LN for each of the parameters (from DATA SET 2).

Table 4.-- Continued

Section	Description
I	Each parameter set used to calculate the modified Beale's measure is listed with: (1) hydraulic heads and flows simulated using this parameter set and the ground-water flow model, (2) hydraulic heads and flows calculated using this data set and a linear approximation to the ground-water flow model (the linear model uses the sensitivities of F), (3) the sum of squared errors calculated using simulated hydraulic heads and flows from the ground-water flow model, and (4) the sum of squared errors calculated using the calculated hydraulic heads and flows from the linear model. Large differences between the last two values indicate that the model is nonlinear with respect to this parameter set. For the example output shown in figure 9, the model is approximately linear with respect to parameter set 1, and nonlinear with respect to parameter set 48. To determine the source of the nonlinearity, compare the parameter set in I with the optimized set in B to determine which parameter values are most different, and compare the two sets of hydraulic heads and flows presented in I to determine which are most different. For the example output shown in figure 9, parameter set 48 differs from the optimized parameter set most for parameters 23 and 24, and the dependent variables that are most different from the linearized dependent variables are those with DID's of SS1, which, for this example, are flows along reaches of a head-dependent boundary.
J	The modified Beale's measure and its critical values.
K	A table of the statistic described by Cooley and Naff (1990, p. 174) for each of the parameter sets listed in I, and an explanation of how to interpret the values.

OPTIMUM PARAMETERS

NO.	PID	BOPT	NO.	PID	BOPT	NO.	PID	BOPT
1	RCH	0.35405E-02	9	T	1.1264	17	T	0.89533
2	KRB	17715.	10	T	0.62690	18	T	0.83925
3	KRB	44793.	11	T	1.1375	19	T	0.81115
4	KRB	88626.	12	T	1.0156	20	T	0.68161
5	KV	0.86154E-02	13	T	0.99995	21	T	0.81799
6	T	0.96500	14	T	1.5677	22	T	1.5169
7	T	0.99012	15	T	1.2035	23	T	0.80766
8	T	1.0024	16	T	0.78975	24	T	0.95562

.
.
.

PARAMETERS FOR SAMPLE NO. 1

NO.	PID	B	NO.	PID	B	NO.	PID	B
1	RCH	0.31800E-02	9	T	1.0600	17	T	0.79900
2	KRB	15700.	10	T	0.57300	18	T	0.81700
3	KRB	36200.	11	T	1.0900	19	T	0.72700
4	KRB	62700.	12	T	0.95100	20	T	0.61900
5	KV	0.76100E-02	13	T	0.99300	21	T	0.77800
6	T	0.93300	14	T	1.3500	22	T	1.2000
7	T	0.96700	15	T	1.0900	23	T	0.83300
8	T	0.98900	16	T	0.69600	24	T	0.76200

DEPENDENT VARIABLES COMPUTED FOR SAMPLE NO. 1

NO.	DID	FC	NO.	DID	FC	NO.	DID	FC
1	1	12.174	25	13	7.6559	49	25	4.8080
2	1	11.886	26	13	7.6502	50	25	4.8181
3	2	11.895	27	14	2.6429	51	26	7.5808
4	2	11.572	28	14	2.6372	52	26	7.5750
5	3	12.060	29	15	4.5632	53	27	11.976
6	3	11.840	30	15	4.5585	54	27	11.681
7	4	11.456	31	16	2.7779	55	SS1	-50407.
8	4	11.454	32	16	2.7690	56	SS1	-38360.
9	5	10.859	33	17	6.2561	57	SS2	-55234.
10	5	10.572	34	17	6.2517	58	SS1	-61673.
11	6	10.857	35	18	12.290	59	SS1	-51444.
12	6	10.849	36	18	12.032	60	SS1	-46054.
13	7	9.2330	37	19	11.665	61	SS1	-42984.
14	7	9.2281	38	19	11.337	62	SS1	-41815.
15	8	8.8455	39	20	11.712	63	SS1	-41722.
16	8	8.8376	40	20	11.427	64	SS1	-41809.
17	9	8.7533	41	21	9.1876	65	SS1	-41234.
18	9	8.7462	42	21	8.9817	66	SS1	-40757.
19	10	10.542	43	22	6.5663	67	SS1	-40465.
20	10	10.538	44	22	6.5639	68	SS1	-40405.
21	11	4.2326	45	23	7.8729	69	SS1	-40235.
22	11	4.2291	46	23	7.8651	70	SS1	-39788.
23	12	4.2400	47	24	10.391	71	SS1	-39650.
24	12	4.2325	48	24	10.123	72	SS1	-40572.

LINEARIZED DEPENDENT VARIABLES COMPUTED FOR SAMPLE NO. 1

NO.		FL	NO.		FL	NO.		FL
1	1	12.130	25	13	7.6016	49	25	4.7792
2	1	11.842	26	13	7.5959	50	25	4.7903
3	2	11.852	27	14	2.6252	51	26	7.5269
4	2	11.530	28	14	2.6195	52	26	7.5211
5	3	12.015	29	15	4.5402	53	27	11.932
6	3	11.795	30	15	4.5356	54	27	11.637
7	4	11.409	31	16	2.7605	55	SS1	-49952.
8	4	11.407	32	16	2.7514	56	SS1	-38045.
9	5	10.817	33	17	6.2125	57	SS2	-54756.
10	5	10.531	34	17	6.2082	58	SS1	-61697.
11	6	10.809	35	18	12.245	59	SS1	-51622.
12	6	10.801	36	18	11.987	60	SS1	-46228.
13	7	9.1956	37	19	11.622	61	SS1	-43126.
14	7	9.1908	38	19	11.296	62	SS1	-41930.
15	8	8.8121	39	20	11.666	63	SS1	-41815.
16	8	8.8042	40	20	11.383	64	SS1	-41901.
17	9	8.7181	41	21	9.1395	65	SS1	-41335.
18	9	8.7109	42	21	8.9360	66	SS1	-40858.
19	10	10.493	43	22	6.5248	67	SS1	-40564.
20	10	10.488	44	22	6.5225	68	SS1	-40499.
21	11	4.2093	45	23	7.8385	69	SS1	-40299.
22	11	4.2050	46	23	7.8306	70	SS1	-39861.
23	12	4.2165	47	24	10.353	71	SS1	-39720.
24	12	4.2088	48	24	10.085	72	SS1	-40642.

SS((FC-FOPT)*W**.5) = 0.96349
SS((FL-FOPT)*W**.5) = 0.81862

.
.
.

Figure 9.-- Examples of sections B, I, J, and K of BEALE.OUT.

PARAMETERS FOR SAMPLE NO. 48

NO.	PID	B	NO.	PID	B	NO.	PID	B
1	RCH	0.35800E-02	9	T	1.2500	17	T	0.60500
2	KRB	18800.	10	T	0.87800	18	T	0.86000
3	KRB	51200.	11	T	1.1500	19	T	0.80400
4	KRB	0.11900E+06	12	T	0.80700	20	T	0.67600
5	KV	0.92500E-02	13	T	1.0000	21	T	0.98200
6	T	0.95000	14	T	1.4900	22	T	1.5400
7	T	0.98100	15	T	1.1400	23	T	0.20400
8	T	1.0100	16	T	0.77900	24	T	6.1700

DEPENDENT VARIABLES COMPUTED FOR SAMPLE NO. 48

NO.	DID	FC	NO.	DID	FC	NO.	DID	FC
1	1	11.457	25	13	5.9715	49	25	4.5014
2	1	11.188	26	13	5.9649	50	25	4.5114
3	2	11.205	27	14	2.4455	51	26	5.9430
4	2	10.906	28	14	2.4389	52	26	5.9363
5	3	11.285	29	15	3.8951	53	27	11.262
6	3	11.076	30	15	3.8900	54	27	10.987
7	4	10.566	31	16	2.7375	55	SS1	-52775.
8	4	10.564	32	16	2.7274	56	SS1	-39159.
9	5	10.168	33	17	4.6511	57	SS2	-58897.
10	5	9.8995	34	17	4.6462	58	SS1	-68884.
11	6	9.9805	35	18	11.548	59	SS1	-54946.
12	6	9.9714	36	18	11.304	60	SS1	-48105.
13	7	8.6873	37	19	10.980	61	SS1	-44076.
14	7	8.6817	38	19	10.676	62	SS1	-42189.
15	8	8.3204	39	20	10.976	63	SS1	-41667.
16	8	8.3116	40	20	10.710	64	SS1	-41328.
17	9	8.1539	41	21	8.4166	65	SS1	-39502.
18	9	8.1459	42	21	8.2186	66	SS1	-37717.
19	10	9.5299	43	22	5.4864	67	SS1	-36489.
20	10	9.5248	44	22	5.4837	68	SS1	-37038.
21	11	4.0212	45	23	7.4229	69	SS1	-44967.
22	11	4.0171	46	23	7.4141	70	SS1	-57756.
23	12	3.7107	47	24	9.7521	71	SS1	-69549.
24	12	3.7023	48	24	9.5006	72	SS1	-80147.

LINEARIZED DEPENDENT VARIABLES COMPUTED FOR SAMPLE NO. 48

NO.		FL	NO.		FL	NO.		FL
1	1	12.120	25	13	7.4577	49	25	4.7352
2	1	11.852	26	13	7.4509	50	25	4.7464
3	2	11.824	27	14	2.5626	51	26	7.4191
4	2	11.525	28	14	2.5563	52	26	7.4123
5	3	11.989	29	15	4.6993	53	27	11.910
6	3	11.780	30	15	4.6945	54	27	11.636
7	4	11.342	31	16	2.9971	55	SS1	-57233.
8	4	11.340	32	16	2.9886	56	SS1	-42576.
9	5	10.779	33	17	5.9254	57	SS2	-64359.
10	5	10.509	34	17	5.9199	58	SS1	-76147.
11	6	10.771	35	18	12.238	59	SS1	-61095.
12	6	10.762	36	18	11.995	60	SS1	-53770.
13	7	9.1706	37	19	11.583	61	SS1	-49548.
14	7	9.1650	38	19	11.279	62	SS1	-47818.
15	8	8.7934	39	20	11.634	63	SS1	-47535.
16	8	8.7844	40	20	11.368	64	SS1	-47248.
17	9	8.6960	41	21	9.2300	65	SS1	-45215.
18	9	8.6879	42	21	9.0347	66	SS1	-42955.
19	10	10.431	43	22	6.6162	67	SS1	-40965.
20	10	10.426	44	22	6.6138	68	SS1	-39776.
21	11	4.2342	45	23	7.8385	69	SS1	-40273.
22	11	4.2303	46	23	7.8297	70	SS1	-42209.
23	12	4.1165	47	24	10.318	71	SS1	-45670.
24	12	4.1072	48	24	10.063	72	SS1	-50460.

SS((FC-FOPT)*W**.5) = 38.634
SS((FL-FOPT)*W**.5) = 0.83138

Figure 9.-- Examples of sections B, I, J, and K of BEALE.OUT--Continued.

```
BEALES MEASURE =  2.5213
IF BEALES MEASURE IS GREATER THAN   0.59     , THE MODEL IS NONLINEAR.
IF BEALES MEASURE IS LESS THAN   0.53E-01, THE MODEL IS ROUGHLY LINEAR
AND LINEAR CONFIDENCE INTERVALS ARE FAIRLY ACCURATE IF THE RESIDUALS ARE
NORMALLY DISTRIBUTED.  INTERMEDIATE VALUES ARE INCONCLUSIVE.

THE FOLLOWING TABLE SHOWS VALUES OF THE STATISTIC DESCRIBED BY COOLEY AND
NAFF (1990,P.174,TOP OF RIGHT COLUMN). THE STATISTIC EQUALS THE NONLINEAR SUM
OF SQUARED ERRORS EVALUATED FOR EACH DATA SET (NSSE) MINUS THE SUM OF SQUARED
ERRORS FOR THE OPTIMUM PARAMETER VALUES (  1.30     ).
IF THE MODEL IS LINEAR, THE STATISTIC SHOULD BE CLOSE TO  0.829
IF THE CORRELATIONS BETWEEN PARAMETERS IS SMALL, THE TABLE SHOWS WHICH
INDIVIDUAL PARAMETERS ARE MOST NONLINEAR.
THE FIRST PAIR OF PARAMETER SETS ARE RELATED TO THE FIRST PARAMETER, THE
SECOND PAIR ARE RELATED TO THE SECOND PARAMTER, AND SO ON.

PARAMETER                          STATISTIC       PERCENT
  SET       NSSE     STATISTIC    - 0.829       DIFFERENCE

    1       2.27      0.976        0.147           17.72
    2       2.60       1.30        0.468           56.47
    .
    .
    .
   48       40.0       38.7         37.8          4566.83
```

Figure 9.-- Examples of sections B, I, J, and K of BEALE.OUT--Continued.

Description of Computer Program YR

Program YR reads the weighted residual, simulated value, and the square root of the weight from each line of the input files; calculates the weighted simulated value; and writes the weighted simulated value and the weighted residual to the output file.

The variables used in the program YR are described in the following table.

Variable	Definition
B	Values from the column 'CALC. VALUES' of the tables 'PARAMETERS WITH PRIOR INFORMATION, BY GROUP' and 'PARAMETER SUMS WITH PRIOR INFORMATION'. Also used to store the weighted simulated value, which is calculated with ln B if the associated LN ≠ 0.
CHECK	Character variable used to read the first 10 characters in lines of input files HCALi, QCALi, and PCALi. If these characters are all blank, it is assumed that this is a blank line in the file and the line is skipped.
FILE	Character variable used to store the names of files that are being opened or closed.
H	Values from the column 'CALC. HEAD' of the table 'DATA AT HEAD LOCATIONS'. Also used to store H*WT.
IH	Flag which is 1 if there are hydraulic-head observations for the simulation being considered, and 0 if there are not.
IP	Flag which is 1 if there is prior information for the simulation being considered, and 0 if there is not.
IQ	Flag which is 1 if there are flow observations for the simulation being considered, and 0 if there are not.
IS	The sequential number of the simulation being considered. IS progresses from 1 to NSD.
IUH	Fortran unit number of HCALi
IUQ	Fortran unit number of QCALi
IUP	Fortran unit number for PCALi
IUM	Fortran unit number for YR.DAT
IOUT	Fortran unit member for YRCALi
LN	If LN is not zero, the parameter was log-transformed for the regression.
NH	Counter for the number of nonblank lines in each file HCALi.

Variable	Definition
NPR	Counter for the number of nonblank lines in each file PCALi.
NQ	Counter for the number of nonblank lines in each file QCALi.
NSD	The number of simulations for which YR tries to read data. NSD needs to be less than or equal to 10.
P	Value of the prior information. Used for parameters with prior information that are scaled with this value (eq. 9).
Q	Values from the column 'CALC. FLOW' of the table 'DATA FOR FLOWS' and used to store weighted simulated values.
QOBS	Values from the column 'MEAS. FLOW' of the table 'DATA FOR FLOWS'.
WT	Value from the column 'WEIGHT**.5' of all the tables used to create the HCALi, QCALi, and PCALi input files.
WTR	Value from the column 'WEIGHTED RESIDUAL' of all the tables used to create HCALi, QCALi and PCALi input files.

Listing of Computer Program YR

```
C     FORTRAN PROGRAM YR
C
C     READS PARTS OF THE MODFLOWP OUTPUT FILE WHICH HAVE BEEN PUT IN SEPARATE
C     FILES NAMED HCAL0, HCAL1,...; QCAL0, QCAL1,...; AND PCAL0, PCAL1,....
C     THE FIRST SET (HCAL) CONTAIN THE TABLES 'DATA AT HEAD LOCATIONS' FROM
C     SIMULATION 0, 1,....  THE SECOND SET (QCAL) CONTAIN THE TABLES 'DATA
C     FOR FLOWS' FROM SIMULATIONS 0, 1,....  THE THIRD SET (PCAL)CONTAIN THE
C     TABLES 'PARAMETERS WITH PRIOR INFORMATION, BY GROUP', AND 'PARAMETER
C     SUMS WITH PRIOR INFORMATION' FROM SIMULATION 0, 1,....
C
C     DATA FILES (YRCAL0, YRCAL1,...), WHICH CONTAIN THE DATA NEEDED FOR
C     PLOTS OF WEIGHTED RESIDUALS VERSUS WEIGHTED SIMULATED VALUES ARE
C     PRODUCED BY YR.
C
C     MARY C HILL    01DEC1992
C
C        $Date: 1993/10/15 19:14:28 $
C        $Revision: 1.4 $
C
      CHARACTER*20 FILE
      CHARACTER*10 CHECK,BLANK
C
  505 FORMAT(A10,51X,F11.0,11X,3F11.0)
  506 FORMAT(A10,11X,2F11.0,11X,3F11.0)
  507 FORMAT(A10,11X,F11.0,11X,3F11.0)
  508 FORMAT(2G16.8)
  509 FORMAT(F10.0,I5)
  515 FORMAT(I5)
  520 FORMAT('HCAL',I1)
  530 FORMAT('QCAL',I1)
  540 FORMAT('PCAL',I1)
  545 FORMAT('YRCAL',I1)
  550 FORMAT(' NSD MUST BE LESS THAN 11 -- STOP')
  560 FORMAT(A10)
  580 FORMAT(/,' FILE ',A20,' DOES NOT EXIST')
  590 FORMAT(/,' NUMBER OF NON-BLANK LINES READ FROM FILE ',A20,'=',I5)
C
C-----NSD IS EQUAL TO OR GREATER THAN THE NUMBER OF SIMULATIONS FOR WHICH
C     DATA ARE BEING PLOTTED.  NSD MUST NOT EXCEED 10.
      NSD=5
C
      ISCREEN=6
      IUH=1
      IUQ=2
      IUP=3
      IUM=10
      IOUT=4
C
      IF(NSD.GT.10) THEN
        WRITE(ISCREEN,550)
        STOP
      ENDIF
```

```
C-----INPUT FILE OF SCALING FACTORS FOR PARAMETERS WITH PRIOR INFORMATION
      IM=1
      OPEN(UNIT=IUM,FILE='YR.DAT',STATUS='OLD',ERR=8)
      GO TO 9
    8 IM=0
      FILE='YR.DAT
      WRITE(ISCREEN,580) FILE
C
    9 BLANK='
C
C     MAIN LOOP
C
      DO 200 IS=1,NSD
        IH=1
        IQ=1
        IP=1
C
C-----FILE WITH FINAL HEADS AND RESIDUALS FROM MODFLOWP OUTPUT FILE
        WRITE(FILE,520) IS-1
        OPEN(UNIT=IUH,FILE=FILE,STATUS='OLD',ERR=1)
        GO TO 2
    1   WRITE(ISCREEN,580) FILE
        IH=0
C
C-----FILE WITH FINAL FLOWS AND RESIDUALS FROM MODFLOWP OUTPUT FILE
    2   WRITE(FILE,530) IS-1
        OPEN(UNIT=IUQ,FILE=FILE,STATUS='OLD',ERR=3)
        GO TO 4
    3   WRITE(ISCREEN,580) FILE
        IQ=0
C
C-----FILE WITH FINAL PARAMETER VALUES FOR PARAMETERS WITH PRIOR
C       INFORMATION AND RESIDUALS
    4   WRITE(FILE,540) IS-1
        OPEN(UNIT=IUP,FILE=FILE,STATUS='OLD',ERR=5)
        GO TO 6
    5   WRITE(ISCREEN,580) FILE
        IP=0
C
    6   IF(IH+IQ+IP.EQ.0) GO TO 200
C
C-----OUTPUT FILE WITH VALUES FOR PLOTTING
       WRITE(FILE,545) IS-1
       OPEN(UNIT=IOUT,FILE=FILE,STATUS='UNKNOWN')
C
C-----INITIALIZE VARIABLES
        NH=0
        NQ=0
        NPR=0
C-----HEADS
        IF(IH.EQ.1) THEN
          DO 10 I=1,10000
            READ(IUH,505,END=11) CHECK,H,WT,WTR,FAC
            IF (FAC.EQ.0) FAC=1.
            IF(CHECK.EQ.BLANK) GO TO 10
            NH=NH+1
            H=H*WT*FAC
```

```
            WRITE(IOUT,508) H,WTR
   10       CONTINUE
   11       IF(NH.EQ.0) CLOSE(IUH)
 12         IF(NH.NE.0) WRITE(IOUT,560) BLANK
            WRITE(ISCREEN, 590) FILE,NH
          ENDIF
C-----FLOWS
C       MULTIPLY THE WEIGHTED OBSERVED VALUE (Q) BY ABS(QOBS) IF WT
C       IS A MULTIPLE OF QOBS.
          IF(IQ.EQ.1) THEN
            DO 20 I=1,10000
              READ(IUQ,506,END=21) CHECK,QOBS,Q,WT,WTR,FAC
              IF (FAC.EQ.0.) FAC=1.
              IF(CHECK.EQ.BLANK) GO TO 20
              NQ=NQ+1
              WRITE(11,508) REAL(NQ),Q
              Q=Q*WT*FAC*ABS(QOBS)
C             Q=Q*WT*FAC
              WRITE(IOUT,508) Q,WTR
   20       CONTINUE
   21       IF(NQ.EQ.0)CLOSE(IUQ)
 22         IF(NQ.NE.0) WRITE(IOUT,560) BLANK
 23         IF(NQ.NE.0) WRITE(11,560) BLANK
            WRITE(ISCREEN, 590) FILE,NQ
          ENDIF
C-----PRIOR INFORMATION
C       THE WEIGHTED OBSERVED VALUES (B) ARE MULTIPLIED BY P, THE PRIOR
C       PARAMETER ESTIMATE, BECAUSE THE ESTIMATED PARAMETERS ARE SCALED BY
C       DIVIDING BY P.  THE LN IS TAKEN BECAUSE THE ESTIMATED PARAMETERS ARE
C       LOG TRANSFORMED
          IF(IP.EQ.1) THEN
            P=1.
            LN=0
            DO 30 I=1,10000
              READ(IUP,507,END=31) CHECK,B,WT,WTR,FAC
              IF (FAC.EQ.0) FAC=1.
              IF(IM.EQ.1) READ(IUM,509) P,LN
              IF(CHECK.EQ.BLANK) GO TO 30
              NPR=NPR+1
              B=P*B
              IF(LN.GT.0) B=ALOG(B)
              B=B*WT*FAC
              WRITE(IOUT,508) B,WTR
   30       CONTINUE
   31       IF(NPR.EQ.0)  CLOSE(IUP)
            WRITE(ISCREEN, 590) FILE,NPR
            REWIND(IUM)
          ENDIF
  200 CONTINUE
C
        STOP
        END
```

Description of Computer Program NORM

Program NORM consists of a MAIN and subroutines ORDER and UNORM. MAIN reads all of the weighted residuals from the input files and calls subroutine ORDER. Subroutine ORDER orders the residuals, calculates the cumulative probability, calls subroutine UNORM (which calculates the value of the order statistic), and writes the weighted residuals and the order statistics to the output file. This is repeated for each simulation for which at least one input file exists.

The variables and their definitions are listed below. Many of the variables are the same as the variables in YR; these are not repeated in this table.

Variable	Description
D	DIMENSION (NDD+NPRD,2), used to store the weighted residuals and a number equal to 1.0 for hydraulic-head residuals, 2.0 for flow residuals, and 3.0 for prior information residuals.
NN	Number of weighted residuals to be included in a normal probability graph.
NMIN	Used as temporary storage when ordering the weighted residuals.
RMIN	Used as temporary storage when ordering the weighted residuals.
RNORM	Calculated cumulative probability.
SMIN	Used as temporary storage when ordering the weighted residuals.

Listing of Computer Program NORM

```
C       FORTRAN PROGRAM NORM
C
C     READS PARTS OF THE MODFLOWP OUTPUT FILE WHICH HAVE BEEN PUT IN SEPARATE
C     FILES NAMED HCAL0, HCAL1,...; QCAL0, QCAL1,...; AND PCAL0, PCAL1,....
C     THE FIRST SET (HCAL) CONTAIN THE TABLES 'DATA AT HEAD LOCATIONS' FROM
C     SIMULATION 0, 1,....  THE SECOND SET (QCAL) CONTAIN THE TABLES 'DATA
C     FOR FLOWS' FROM SIMULATIONS 0, 1,....  THE THIRD SET (PCAL)CONTAIN THE
C     TABLES 'PARAMETERS WITH PRIOR INFORMATION, BY GROUP', AND 'PARAMETER
C     SUMS WITH PRIOR INFORMATION' FROM SIMULATION 0, 1,....
C
C     DATA FILES WHICH CONTAIN THE VALUES NEEDED FOR NORMAL PROBABILITY PLOTS
C     OF WEIGHTED RESIDUALS ARE PRODUCED BY NORM.
C     HEAD RESIDUALS ONLY ARE CONTAINED IN FILES NAMED NORMH0, NORMH1,...;
C     HEAD AND FLOW RESIDUALS ARE CONTAINED IN FILES NAMED NORMQ0, NORMQ1,...;
C     HEAD, FLOW, AND PRIOR INFORMATION RESIDUALS ARE CONTAINED IN FILES NAMED
C     NORMP0, NORMP1,....
C
C       READS PIECES OF THE MODFLOWP OUTPUT FILE AND PRODUCES INPUT FILES FOR
C       MORMAL PROBABILITY PLOTS
C
C
C       MARY C HILL     15DEC1992
C
C          $Date: 1993/05/18 15:47:02 $
C          $Revision: 1.2 $
C
C       NDD MUST EQUAL OR EXCEED THE NUMBER OF HEAD AND FLOW OBSERVATIONS.
C       NPRD MUST EQUAL OR EXCEED THE NUMBER OF PRIOR ESTIMATES ON THE
C         PARAMETERS.
C       NSD MUST EQUAL OR EXCEED THE NUMBER OF RUNS.  NSD MAY NOT EXCEED 10.
        PARAMETER (NDD=100,NPRD=25,NSD=10)
        CHARACTER*20 BLANK,FILE,CHECK
        DIMENSION D(NDD+NPRD,2),R(NDD+NPRD)
C
  505 FORMAT(A20,41X,10X,23X,F11.0)
  506 FORMAT(A20,1X,22X,22X,F11.0)
  507 FORMAT(A20,1X,11X,22X,F11.0)
  515 FORMAT(I5)
  520 FORMAT('HCAL',I1)
  530 FORMAT('QCAL',I1)
  540 FORMAT('PCAL',I1)
  545 FORMAT('NORMH',I1)
  546 FORMAT('NORMQ',I1)
  547 FORMAT('NORMP',I1)
  550 FORMAT(' IN PARAMETER STATEMENT IN BEGINNING OF PROGRAM,
     1 'NSD MUST BE LESS THAN 11 -- STOP')
  560 FORMAT(A20)
  580 FORMAT(/,' FILE ',A20,' DOES NOT EXIST')
C
      ISCREEN=6
      IUH=1
      IUQ=2
      IUP=3
      IOUTH=12
      IOUTQ=13
```

```
      IOUTP=14
C
      IF(NSD.GT.10) THEN
        WRITE(ISREEN,550)
        STOP
      ENDIF
      BLANK='
C
C     MAIN LOOP
C
      DO 200 IS=1,NSD
        IH=1
        IQ=1
        IP=1
C
C-----FILE WITH FINAL HEADS AND RESIDUALS FROM MODFLOWP OUTPUT FILE
        WRITE(FILE,520) IS-1
        OPEN(UNIT=IUH,FILE=FILE,STATUS='OLD',ERR=1)
        GO TO 2
   1    WRITE(ISCREEN,580) FILE
        IH=0
C
C-----FILE WITH FINAL FLOWS AND RESIDUALS FROM MODFLOWP OUTPUT FILE
   2    WRITE(FILE,530) IS-1
        OPEN(UNIT=IUQ,FILE=FILE,STATUS='OLD',ERR=3)
        GO TO 4
   3    WRITE(ISCREEN,580) FILE
        IQ=0
C
C-----FILE WITH FINAL PARAMETER VALUES FOR PARAMETERS WITH PRIOR
C      INFORMATION AND RESIDUALS
   4    WRITE(FILE,540) IS-1
        OPEN(UNIT=IUP,FILE=FILE,STATUS='OLD',ERR=5)
        GO TO 6
   5    WRITE(ISCREEN,580) FILE
        IP=0
C
C-----OUTPUT FILES WITH VALUES FOR PLOTTING
   6    IF(IH.EQ.1) THEN
          WRITE(FILE,545) IS-1
          OPEN(UNIT=IOUTH,FILE=FILE,STATUS='UNKNOWN')
        ENDIF
C
        IF(IQ.EQ.1) THEN
          WRITE(FILE,546) IS-1
          OPEN(UNIT=IOUTQ,FILE=FILE,STATUS='UNKNOWN')
        ENDIF
C
        IF(IP.EQ.1) THEN
          WRITE(FILE,547) IS-1
          OPEN(UNIT=IOUTP,FILE=FILE,STATUS='UNKNOWN')
        ENDIF
C
C-----INITIALIZE VARIABLES
        NH=0
        NQ=0
        NPR=0
```

```
        DO 8 I=1,NDD+NPRD
        D(I,1)=0.
    8   D(I,2)=0.
C-----HEADS
        IF(IH.EQ.1) THEN
          DO 10 I=1,10000
            READ(IUH,505,END=11) CHECK,WTR
            IF(CHECK.EQ.BLANK) GO TO 10
            NH=NH+1
            D(NH,1)=WTR
            D(NH,2)=1.
   10     CONTINUE
   11     IF(NH.EQ.0) CLOSE(IUH)
          IF(NH.NE.0) CALL ORDER(D,R,NH,IOUTH,NDD+NPRD)
        ENDIF
C-----FLOWS
        IF(IQ.EQ.1) THEN
          DO 20 I=1,10000
            READ(IUQ,506,END=21) CHECK,WTR
            IF(CHECK.EQ.BLANK) GO TO 10
            NQ=NQ+1
            D(NH+NQ,1)=WTR
            D(NH+NQ,2)=2.
   20     CONTINUE
   21     IF(NQ.EQ.0)CLOSE(IUQ)
          IF(NQ.NE.0) CALL ORDER(D,R,NH+NQ,IOUTQ,NDD+NPRD)
        ENDIF
C-----PRIOR INFORMATION
        IF(IP.EQ.1) THEN
          DO 30 I=1,10000
            READ(IUP,507,END=31) CHECK,WTR
            IF(CHECK.EQ.BLANK) GO TO 10
            NPR=NPR+1
            D(NH+NQ+NPR,1)=WTR
            D(NH+NQ+NPR,2)=3.
   30     CONTINUE
   31     IF(NPR.EQ.0) CLOSE(IUP)
          IF(NPR.NE.0) CALL ORDER(D,R,NH+NQ+NPR,IOUTP,NDD+NPRD)
        ENDIF
  200 CONTINUE
C
      STOP
      END


C*************************************************************************
        SUBROUTINE ORDER(D,R,NN,IPR,ND)
        DIMENSION D(ND,2),R(NN)
500     FORMAT(2G15.6)
510     FORMAT(A4)
        DO 100 N1=1,NN-1
          NMIN=N1
          RMIN=D(N1,1)
          DO 50 N2=N1+1,NN
            IF(D(N2,1).LE.RMIN) THEN
              RMIN=D(N2,1)
              SMIN=D(N2,2)
```

```
              NMIN=N2
            ENDIF
50        CONTINUE
          IF(NMIN.NE.N1) THEN
            D(NMIN,1)=D(N1,1)
            D(NMIN,2)=D(N1,2)
            D(N1,1)=RMIN
            D(N1,2)=SMIN
          ENDIF
100     CONTINUE
        DO 200 N=1,NN
          RNORM=(REAL(N)-.5)/(REAL(NN))
          CALL UNORM(R(N),RNORM,-1)
200     CONTINUE
        DO 300 I=1,3
        DO 290 N=1,NN
290       IF(INT(D(N,2)).EQ.I) WRITE(IPR,500) D(N,1),R(N)
300       WRITE(IPR,510) '    '
        RETURN
        END


C***********************************************************************
C     SUBROUTINE UNORM IS FROM MODFLOWP (HILL,1992)
      SUBROUTINE UNORM(U,RNORM,IP)                                      UTI0267
C-----VERSION 1000 01FEB1992                                            UTI0268
C     ***************************************************************UTI0269
C     FIND THE PROBABILITY RELATED TO A U (IP=1), OR A U RELATED TO A   UTI0270
C     PROBABILITY (IP=-1) FOR A STANDARD GAUSSIAN DISTRIBUTION          UTI0271
C     ***************************************************************UTI0272
C        SPECIFICATIONS:                                                UTI0273
C     ---------------------------------------------------------------UTI0274
      DIMENSION PNORM(2,41)                                             UTI0275
      DATA (PNORM(1,I),I=1,41)/0.0,.15,.20,.25,.30,.35,                 UTI0276
     1       .40,.45,.50,.55,.60,.65,.70,.75,.80,                       UTI0277
     2       .85,.90,.95,1.00,1.05,1.10,1.15,1.20,1.25,                 UTI0278
     3       1.30,1.35,1.40,1.45,1.50,1.55,1.60,1.65,                   UTI0279
     4       1.70,1.75,1.80,1.85,1.90,1.95,2.20,                        UTI0280
     5       4.00,5.5/                                                  UTI0281
      DATA (PNORM(2,I),I=1,41)/.5,.5596,.5793,.5987,.6179,.6368,        UTI0282
     1       .6554,.6736,.6915,.7088,.7257,.7422,.7580,.7734,.7881,     UTI0283
     2       .8023,.8159,.8289,.8413,.8531,.8643,.8749,.8849,.8944,     UTI0284
     3       .90320,.91149,.91924,.92647,.93319,.93943,.94520,.95053,   UTI0285
     4       .95543,.95994,.96407,.96784,.97128,.97441,.98610,          UTI0286
     5       .9999683,1.0/                                              UTI0287
C     ---------------------------------------------------------------UTI0288
C                                                                       UTI0289
C-----GIVEN U, GET THE CUMULATIVE PROBABILITY                           UTI0290
      IF(IP.EQ.1) THEN                                                  UTI0291
C-------FIND THE VALUES ABOVE AND BELOW U                              UTI0292
        AU=ABS(U)                                                       UTI0293
        IF(AU.GE.5.5) THEN                                              UTI0294
          RNORM=1.0                                                     UTI0295
          IF(U.LT.0.0) RNORM=0.0                                        UTI0296
          RETURN                                                        UTI0297
        ENDIF                                                           UTI0298
        DO 100 I=1,40                                                   UTI0299
```

```
  100     IF(AU.GE.PNORM(1,I).AND.AU.LT.PNORM(1,I+1)) GO TO 150          UTI0300
        STOP 'ERROR IN UNORM -- U NOT FOUND'                             UTI0301
C-------INTERPOLATE                                                     UTI0302
  150   FACTOR=(AU-PNORM(1,I))/(PNORM(1,I+1)-PNORM(1,I))                 UTI0303
        RNORM=PNORM(2,I)+FACTOR*(PNORM(2,I+1)-PNORM(2,I))                UTI0304
        IF(U.LT.0) RNORM=1.0-RNORM                                       UTI0305
        RETURN                                                           UTI0306
      ENDIF                                                              UTI0307
C-----GIVEN THE CUMULATIVE PROBABILITY, GET U                            UTI0308
      IF(IP.EQ.-1) THEN                                                  UTI0309
C-------FIND THE VALUES ABOVE AND BELOW RNORM                           UTI0310
        ARNORM=RNORM                                                     UTI0311
        IF(RNORM.LT..50) ARNORM=1.-RNORM                                 UTI0312
        IF(ARNORM.EQ.1.0) THEN                                           UTI0313
          U=5.5                                                          UTI0314
          IF(RNORM.LT..5) U=-5.5                                         UTI0315
          RETURN                                                         UTI0316
        ENDIF                                                            UTI0317
        DO 170 I=1,40                                                    UTI0318
  170     IF(ARNORM.GE.PNORM(2,I).AND.ARNORM.LT.PNORM(2,I+1))            UTI0319
     1      GO TO 190                                                   UTI0320
        STOP 'ERROR IN UNORM -- RNORM NOT FOUND'                         UTI0321
C-------INTERPOLATE                                                     UTI0322
  190   FACTOR=(ARNORM-PNORM(2,I))/(PNORM(2,I+1)-PNORM(2,I))            UTI0323
        U=PNORM(1,I)+FACTOR*(PNORM(1,I+1)-PNORM(1,I))                    UTI0324
        IF(RNORM.LT..50) U=-U                                            UTI0325
        RETURN                                                           UTI0326
      ENDIF                                                              UTI0327
      END                                                                UTI0328
```

Description of Computer Program BCINT

Program BCINT reads the input files for all of the simulations, calculates the confidence intervals and the plotting positions, and writes the parameter values and their plotting positions to the output file, followed by the confidence intervals and their plotting positions.

The variables used in the program BCINT are described in the following table.

Variable	Definition
B	DIMENSION (NPD,NSD), estimated parameter values. Also used to store the estimated parameter values times BX.
BL	DIMENSION (NPD,NSD), lower limits of confidence intervals.
BU	DIMENSION (NPD,NSD), upper limits of confidence intervals.
BX	DIMENSION (NPD), multiplicative factors used to scale parameter values.
CTMP	Character variable used to check location in the input file.
IY	DIMENSION (NPD,NSD), read from data set 3 of the PARi input files (table 1) for i = 1, NSD. Confidence intervals are calculated for parameters I for which IY(I) $\neq$ 0, and parameters from different simulations with the same value of IY(I) are grouped. See the definition for variable Y.
LN	DIMENSION (NPD), LN values (see definitions for DATA SET 2 of MODFLOWP) for each parameter.
NPD	Defined in the PARAMETER statement at the beginning of the program and used to dimension arrays. Must equal or exceed the maximum number of parameters in any of the simulations.
NSD	Defined in the PARAMETER statement at the beginning of the program and used to dimension arrays. Must equal or exceed the number of simulations, which may not exceed 10.
NSIM	Counter for the number of simulations for which confidence intervals are calculated.
PID	DIMENSION (NPD), same as PID in DATA SET 2 of MODFLOWP.
STD	DIMENSION (NPD), standard deviations of the estimated parameter values.
Y	Plotting position calculated by adding (NS-1)/(NSIM+1) to IY(I, NS), where NS is one for the first simulation for which confidence intervals are calculated, two for the second simulation for which confidence intervals are calculated and so on, and NSIM at this point in the program equals the total number of simulations for which confidence intervals are calculated. Y was used as the plotting position along the vertical axis in figure 3.

Listing of Computer Program BCINT

```
C     FORTRAN PROGRAM BCINT
C
C     WRITES DATA FILE (CI.OUT) WHICH CONTAINS THE DATA REQUIRED TO PLOT
C     ESTIMATED PARAMETER VALUES AND THEIR CONFIDENCE INTERVALS.
C     THE INPUT FILES FOR BCINT (PAR0, PAR1,...) CONTAIN THE TABLES
C     'PARAMETER SUMMARY FROM THE MODFLOWP OUTPUT FILE.
C
C     MARY C. HILL       15DEC1992
C
C         $Date: 1993/05/18 15:47:03 $
C         $Revision: 1.2 $
C
C     NSD MUST EQUAL OR EXCEED THE NUMBER OF RUNS FOR WHICH CONFIDENCE
C         INTERVALS ARE TO BE PLOTTED.  NSD MAY NOT EXCEED 10.
C     NPD MUST EQUAL OR EXCEED THE LARGEST NUMBER OF PARAMETER VALUES
C         IN ANY OF THE SIMULATIONS
      PARAMETER (NSD=5,NPD=50)
C
      DIMENSION BL(NPD,NSD),B(NPD,NSD),BU(NPD,NSD),BX(NPD),
     1     IY(NPD,NSD),STD(NPD),LN(NPD)
      CHARACTER*4 PID(NPD),CTMP,BLANK
      CHARACTER*20 FILE
500   FORMAT(10I5)
505   FORMAT (10F5.0)
510   FORMAT(16X,6X,A4,6X,A4,6X,A4,6X,A4,6X,A4,6X,A4,6X,A4,6X,A4,6X,A4,6X,A4,
     1   6X,A4)
520   FORMAT(17X,10F10.0)
530   FORMAT(A4)
540   FORMAT(I5,F10.0)
550   FORMAT('PAR',I1)
560   FORMAT(2G13.3)
570   FORMAT(/,' IN PARAMETER STATEMENT IN BEGINNING OF PROGRAM, ',
     1     'NSD MUST BE LESS THAN 11')
600     FORMAT(/,' FILE ',A20,' DOES NOT EXIST')
C
      ISCREEN=6
      IPAR=1
      IOUT=2
      BLANK='
C
      IF(NSD.GT.10) THEN
        WRITE(ISCREEN,570)
        STOP
      ENDIF
C
      DO 10 NS=1,NSD
      DO 10 IP=1,NPD
10    IY(IP,NS)=0.
C
      NPMAX=0.
      NSIM=0
C
      DO 100 NS=1,NSD
C         DEFINE AND OPEN DATA FILE
        WRITE(FILE,550) NS-1
```

```
      OPEN(UNIT=IPAR,FILE=FILE,STATUS='OLD',ERR=95)
      NSIM=NSIM+1
C       READ DATA
C         NUMBER OF PARAMETERS AND STATISTIC FOR CALCULATING CONFIDENCE
C         INTERVALS
      READ(IPAR,540) NP,STAT
      IF(NP.GT.NPMAX) NPMAX=NP
C         WERE THE PARAMETERS LOG-TRANSFORMED FOR THE REGRESSION?
      READ(IPAR,500) (LN(IP),IP=1,NP)
C         THE SEQUENCE IN WHICH THE PARAMETERS FOR THIS SIMULATION ARE
C         PLOTTED
      READ(IPAR,500) (IY(IP,NSIM),IP=1,NP)
C         MULTIPLICATIVE FACTORS
      READ(IPAR,505) (BX(IP),IP=1,NP)
C         LINES FROM MODFLOWP
      READ(IPAR,510) (PID(IP),IP=1,NP)
      READ(IPAR,510) CTMP
      READ(IPAR,520) (B(IP,NSIM),IP=1,NP)
      READ(IPAR,530) CTMP
      IF(CTMP.NE.' STD'.AND.CTMP.NE.'STD.') THEN
        READ(IPAR,510) CTMP
        READ(IPAR,520) (DUMB,IP=1,NP)
      ENDIF
C       STANDARD DEVIATIONS
      READ(IPAR,510) CTMP
      READ(IPAR,520) (STD(IP),IP=1,NP)
C       CALCULATE CONFIDENCE INTERVALS
      DO 20 IP=1,NP
        BU(IP,NSIM)=B(IP,NSIM)+STAT*STD(IP)
        BL(IP,NSIM)=B(IP,NSIM)-STAT*STD(IP)
20    CONTINUE
      DO 30 IP=1,NP
        IF(LN(IP).NE.0) THEN
          B(IP,NSIM)=EXP(B(IP,NSIM))
          BU(IP,NSIM)=EXP(BU(IP,NSIM))
          BL(IP,NSIM)=EXP(BL(IP,NSIM))
        ENDIF
30    CONTINUE
      DO 40 IP=1,NP
        BXX=BX(IP)
        B(IP,NSIM)=B(IP,NSIM)*BXX
        BU(IP,NSIM)=BU(IP,NSIM)*BXX
        BL(IP,NSIM)=BL(IP,NSIM)*BXX
40    CONTINUE
      CLOSE(IPAR)
      GO TO 100
C       FILE DOES NOT EXIST
95    WRITE(ISCREEN,600)FILE
C
100   CONTINUE
C
      OPEN(UNIT=IOUT,FILE='CI.OUT',STATUS='UNKNOWN')
      FRAC=1./REAL(NSIM+1)
      DO 120 NS=1,NSIM
        SUB=REAL(NS)-1.
      DO 120 IP=1,NPMAX
        Y=REAL(IY(IP,NS))-SUB*FRAC
```

```
120     IF(IY(IP,NS).NE.0) WRITE(IOUT,560) B(IP,NS),Y
      WRITE(IOUT,530) BLANK
      DO 140 NS=1,NSIM
      DO 130 IP=1,NPMAX
        IF(IY(IP,NS).EQ.0) GO TO 130
        SUB=REAL(NS)-1.
        Y=REAL(IY(IP,NS))-SUB*FRAC
        WRITE(IOUT,560) BL(IP,NS),Y
        WRITE(IOUT,560) BU(IP,NS),Y
130   CONTINUE
140   CONTINUE
      STOP
      END
```

Description of Computer Program YCINT

YCINT is modified from an unpublished program called RELIAB by R.L. Cooley (U.S. Geological Survey, written commun., 1992). The variables used in the program YCINT are described in the following table.

Variable	Definition
C	DIMENSION (NPD,NPD), variance-covariance matrix for the estimated parameters.
DID	DIMENSION (NDD), data identifiers (see descriptions for DATA SETS 6 and 7 of MODFLOWP) for the predictive conditions.
DID1	DIMENSION (NDD), data identifiers (see descriptions for DATA SETS 6 and 7 of MODFLOWP) for the base conditions.
DUMBC	A character variable used to skip lines in the input files.
H	DIMENSION (NDD), the values of the quantities defined by DATA SETS 6 and 7 calculated for predictive conditions using optimum parameter values.
H1	DIMENSION (NDD), the values of the quantities defined by DATA SETS 6 and 7 calculated for the base conditions using optimum parameter values.
NDD	Defined in the PARAMETER statement at the beginning of the program. NDD must equal or exceed NH+NQT (LINE 5 of MODFLOWP).
NPD	Defined in the PARAMETER statement at the beginning of the program. NPD must equal or exceed NP (LINE 3 of MODFLOWP).
NVAR	Equals NP of LINE 3 of MODFLOWP.
X	DIMENSION (NPD,NDD), sensitivities for the optimum parameters under predictive conditions.
X1	DIMENSION (NPD,NDD), sensitivities for the optimum parameters for the base conditions.

Listing of Computer Program YCINT

```
C       FORTRAN PROGRAM YCINT
C
C      CALCULATE LINEAR CONFIDENCE INTERVALS ON SIMULATED HYDRAULIC HEADS
C      AND FLOWS ALONG HEAD-DEPENDENT BOUNDARIES.
C
C      MARY C. HILL     15DEC1992
C          MODIFIED FROM AN UNPUBLISHED PROGRAM CALLED RELIAB BY
C          RICHARD L. COOLEY (WRITTEN COMMUN., 1992)
C
C        $Date: 1993/05/18 15:47:03 $
C        $Revision: 1.2 $
C
C      NDD MUST EQUAL OR EXCEED THE NUMBER OF HEADS AND FLOWS FOR WHICH
C           INTERVALS ARE CALCULATED
C      NPD MUST EQUAL OR EXCEED THE NUMBER OF PARAMETER VALUES
       PARAMETER(NDD=100,NPD=30)
       CHARACTER*4 DID(NDD),DID1(NDD),DUMBC
       DIMENSION X(NPD,NDD),C(NPD,NPD),H(NDD),H1(NDD),X1(NPD,NDD),
      1   V(NDD),V1(NDD)
       COMMON/ITP/IIN,IOUT
C**FORMAT LIST
     1 FORMAT (6I5,2F10.0,2I5)
     2 FORMAT (8F10.0)
     3 FORMAT (' NUMBER OF ESTIMATED PARAMETERS = ',I4/
      1        ' NUMBER OF INTERVALS..... = ',I4/
      2        ' CRITICAL VALUE..... = ',G11.5)
     4 FORMAT(/,' INTERVALS ARE CALCULATED ON DIFFERENCES')
     5 FORMAT (/,9X,'VALUES COMPUTED WITH OPTIMUM PARAME',
      1'TERS FOR PREDICTIVE CONDITIONS AND ASSOCIATED IDENTIFIERS',/
      1 1H ,3X,3(3HNO.,2X,'ID',3X,'VALUE',9X))
     6 FORMAT (/,9X,'VALUE COMPUTED WITH OPTIMUM PARAME',
      1'TERS FOR CALIBRATION CONDITIONS AND ASSOCIATED IDENTIFIERS',/
      1 1H ,3X,3(3HNO.,7X,'VALUE',8X))
     7 FORMAT(/,' INTERVALS ARE NOT CALCULATED ON DIFFERNCES')
     8 FORMAT (/,' SENSITIVITIES FOR OPTIMUM PARAMETERS FOR PREDICTIVE',
      1        ' CONDITIONS')
     9 FORMAT (/,' SENSITIVITIES FOR OPTIMUM PARAMETERS FOR CALIBRATION',
      1        ' CONDITIONS')
    16 FORMAT(16F13.0)
    17 FORMAT (6F13.0)
    18 FORMAT (/,' NDD AND NPD, DEFINED IN THE PARAMETER STATEMENT, ',
      1    'ARE NOT LARGE ENOUGH.',/,
      2    '    SET NDD >=',I5,' AND NPD >=',I5,' -- STOP EXECUTION')
   500 FORMAT(/,' VARIANCE-COVARIANCE MATRIX FOR ESTIMATED PARAMETERS')
   510 FORMAT(/,' CRITICAL VALUE FOR THE INTERVALS = ',G11.5,//,
      1/,12X,'SIMULATED',52X,'SIMULATED',/,
      1    '  NO.  ID     VALUE        STD. DEV.      ',
      1   'CONFIDENCE INTERVAL    ',
      1    '  ID     DIFFERENCE      STD. DEV.   CONFIDENCE INTERVAL')
   515 FORMAT(/,' CRITICAL VALUE FOR THE INTERVALS = ',G11.5,//,
      1/,12X,'SIMULATED',/
      1    '  NO.  ID     VALUE        STD. DEV.      ',
      1'CONFIDENCE INTERVAL')
   520 FORMAT(I4,3X,A4,G13.6,1X,G13.6,1X,G13.6,';',G13.6,A4,G13.6,1X,
      1     G13.6,1X,G13.6,';',G13.6)
```

```
  530 FORMAT(20A4)
  540 FORMAT(/,' PREDICTION INTERVALS ARE CALCULATED -- YCINTS.VAR',
     1   ' IS READ')
  545 FORMAT(/,' CONFIDENCE INTERVALS ARE CALCULATED')
  550 FORMAT(/,' CRITICAL VALUE FOR THE INTERVALS = ',G11.5,//,
     1/,12X,'SIMULATED',52X,'SIMULATED',/,
     1    '  NO.  ID     VALUE        STD. DEV.      ',
     1   'PREDICTION INTERVAL    ',
     1    '  ID     DIFFERENCE      STD. DEV.   PREDICTION INTERVAL')
  555 FORMAT(/,' CRITICAL VALUE FOR THE INTERVALS = ',G11.5,//,
     1/,12X,'SIMULATED',/,
     1    '  NO.  ID     VALUE        STD. DEV.      ',
     1'PREDICTION INTERVAL')
C
C**DEFINE INPUT FILE, OUTPUT FILE, AND ARRAY DIMENSION
      IIN=1
      IIN2=2
      IIN3=3
      IOUT=4
      IVAR=8
C
      OPEN(IIN,FILE='YCINT.DAT',STATUS='OLD')
C     UNIT IIN2 OPENED BELOW, IF NEEDED
      OPEN(IIN3,FILE='RESANP.DAT',STATUS='OLD')
      OPEN(IOUT,FILE='YCINT.OUT',STATUS='UNKNOWN')
C     UNIT IVAR OPENED BELOW, IF NEEDED
C**READ FIRST LINE OF YCINT.DAT
      READ(IIN,1) NVAR,NRES,NOBS,NPR,MPR,NPTS,VAR,FSTAT,IDIF,IPRED
      WRITE(IOUT,3) NVAR,NOBS,FSTAT
      IF(NVAR.GT.NPD.OR.NOBS.GT.NDD) THEN
        WRITE(IOUT,18)NVAR,NOBS
        STOP
      ENDIF
      IF(IDIF.EQ.1) WRITE(IOUT,4)
      IF(IDIF.NE.1) WRITE(IOUT,7)
      IF(IPRED.EQ.1) WRITE(IOUT,540)
      IF(IPRED.NE.1) WRITE(IOUT,545)
C**INITIALIZE V AND V1
      DO 60 N=1,NOBS
        V(N)=0.
   60   V1(N)=0.
C
C**READ RESANP.DAT - VARIANCE-COVARIANCE MATRIX ON THE PARAMETERS FROM
C  THE CALIBRATION.  MUST BE PRODUCED USING THE STRESSES AND DATA SETS
C  6 AND 7 FROM THE CALIBRATION.
C
      READ(IIN3,1) IDUMB
      DO 20 I=1,NVAR
        READ(IIN3,16) (C(I,J),J=I,NVAR)
        IF(I.LT.NVAR) THEN
          DO 10 J=I+1,NVAR
            C(J,I)=C(I,J)
   10     CONTINUE
        ENDIF
   20 CONTINUE
      WRITE(IOUT,500)
      CALL PRTOT(C,NVAR,NVAR,NPD)
```

```
C
C**READ YCINT.DAT2 - BASE CONDITIONS FOR INTERVALS ON DIFFERENCES.
C   PRODUCED WITH IPAR=1; IOUR>0 (DATA SET 13).
C
      IF (IDIF.EQ.1) THEN
        OPEN(IIN2,FILE='YCINT.DAT2',STATUS='OLD')
        READ(IIN2,1) IDUMB,IDUMB,NPR
        DO 30 I=1,NVAR
   30     READ(IIN2,16) (DUMB,J=I,NVAR)
        IF(IPRED.EQ.1) READ(IIN2,16)(V1(I),I=1,NOBS)
        IF(IPRED.NE.1) READ(IIN2,16)(DUMB,I=1,NOBS)
        DO 40 N=1,NOBS
   40     READ(IIN2,16) (X1(I,N),I=1,NVAR)
        IF(NPR.NE.0) THEN
          READ(IIN2,2) DUMB
          READ(IIN2,2) (DUMB,I=1,NVAR)
        ENDIF
        READ(IIN2,16) (H1(N),N=1,NOBS)
        WRITE(IOUT,6)
        READ(IIN2,530) (DID1(N),N=1,NOBS)
        CALL PRTOTB(H1,NOBS,DID1)
        WRITE(IOUT,9)
        CALL PRTOT(X1,NVAR,NOBS,NPD)
      ENDIF
C
C**READ YCINT.DAT - PREDICTIVE CONDITIONS FOR INTERVALS ON DIFFERENCES.
C   PRODUCED WITH IPAR=1; IOUE>0 (LINE 7).
C
      READ(IIN,17) (DUMB,J=1,NVAR)
      READ(IIN,530) (DUMBC,J=1,NVAR)
      READ(IIN,530) (DID(N),N=1,NOBS)
      READ(IIN,17) (H(N),N=1,NOBS)
      WRITE(IOUT,5)
      CALL PRTOTB(H,NOBS,DID)
      READ(IIN,17) (DUMB,N=1,NOBS)
      IF(IPRED.EQ.1) READ(IIN,2) (V(N),N=1,NOBS)
      IF(IPRED.NE.1) READ(IIN,2) (DUMB,N=1,NOBS)
      DO 50 N=1,NOBS
      READ(IIN,17) (X(I,N),I=1,NVAR)
   50 CONTINUE
      WRITE(IOUT,8)
      CALL PRTOT(X,NVAR,NOBS,NPD)
C**CALCULATE AND PRINT THE INTERVALS
      IF(IPRED.EQ.0) THEN
        IF(IDIF.EQ.1) WRITE(IOUT,510)FSTAT
        IF(IDIF.NE.1) WRITE(IOUT,515)FSTAT
      ELSE
        IF(IDIF.EQ.1) WRITE(IOUT,550)FSTAT
        IF(IDIF.NE.1) WRITE(IOUT,555)FSTAT
      ENDIF
      DO 80 N=1,NOBS
        S2=0.0
        S21=0.0
        DO 70 I=1,NVAR
        DO 70 J=1,NVAR
        IF(IDIF.EQ.1)
     1       S21=S21 + (X(I,N)-X1(I,N)) * C(I,J) * (X(J,N)-X1(J,N))
```

```
  70     S2=S2+X(I,N)*C(I,J)*X(J,N)
       S2=S2+V(N)
       S2=S2**.5
       IF(IDIF.EQ.1) THEN
         S21=S21+V(N)+V1(N)
         S21=S21**.5
       ENDIF
       HL=H(N)-FSTAT*S2
       HU=H(N)+FSTAT*S2
       IF(IDIF.EQ.1) THEN
         HD=H(N)-H1(N)
         HL1=HD-FSTAT*S21
         HU1=HD+FSTAT*S21
         WRITE(IOUT,520) N,DID(N),H(N),S2,HL,HU,DID1(N),HD,S21,HL1,HU1
       ELSE
         WRITE(IOUT,520) N,DID(N),H(N),S2,HL,HU
       ENDIF
  80   CONTINUE
       STOP
       END


C
       SUBROUTINE PRTOT(C,NR,NC,NRD)
C**PRINT MATRICES DIVIDED VERTICALLY INTO TEN-COLUMN BLOCKS
       DIMENSION C(NRD,NC)
       COMMON/ITP/IIN,IOUT
       DO 60 K=1,NC,10
       J10=K+9
       IF(J10.GT.NC) J10=NC
       WRITE(IOUT,70) (J,J=K,J10)
       WRITE(IOUT,90)
       DO 30 I=1,NR
    30 WRITE(IOUT,80) I,(C(I,J),J=K,J10)
    60 CONTINUE
    70 FORMAT(1H0,10(9X,I3))
    80 FORMAT (1H ,I3,1X,10(1X,G11.5))
    90 FORMAT (1H )
       RETURN
       END


C
       SUBROUTINE PRTOTB(VAL,NO,DID)
C**PRINT VALUES IN THREE GROUPS OF TWO COLUMNS
       CHARACTER*4 DID(NO)
       DIMENSION VAL(NO)
       COMMON/ITP/IIN,IOUT
       NR=NO/3
       IF(3*NR.NE.NO) NR=NR+1
       DO 10 K=1,NR
       WRITE(IOUT,20) (L,DID(L),VAL(L),L=K,NO,NR)
    10 CONTINUE
       RETURN
    20 FORMAT (1H ,2X,3(I3,2X,A4,1X,G11.5,3X))
       END
```

Description of Computer Program BEALEP

BEALEP is modified from a program documented by Cooley and Naff (1990, p. 190-193). All of the variables used in BEALEP were used in the previous version, or are described in the section 'Description of Computer Program YCINT' of this report.

Listing of computer program BEALEP

```
C       FORTAN PROGRAM BEALEP
C
C       CALCULATES BEALE'S MEASURE OF NONLINEARITY
C
C       MARY C. HILL   15DEC1992
C         MODIFIED FROM COOLEY AND NAFF (1990, P. 187-189)
C
C       $Date: 1993/05/18 15:47:04 $
C       $Revision: 1.6 $
C
C       MODIFIED BEALE'S MEASURE PROGRAM BY R. L. COOLEY, USGS, DENVER,
C       COLO.   SEE COOLEY AND NAFF (1990, P. 187-198)
C       MODIFIED FOR MODFLOWP BY M. C. HILL 01JUN1992
C         VARIABLES IN THE PARAMETER STATEMENT:
C          NDD = OR > SUM OF THE HEAD AND FLOW OBSERVATIONS USED IN THE
C                     REGRESSION (NH+NQT FOR LINE 5 OF MODFLOWP)
C          NPD = OR > NUMBER OF PARAMETERS ESTIMATED IN THE REGRESSION
C          MPD = OR > MPR OF DATA LINE 5 OF MODFLOWP
      PARAMETER(NDD=100,NPD=30,MPD=16)
      CHARACTER*4 DID(NDD),PID(NPD)
      DIMENSION BOPT(NPD),FOPT(NDD),B(NPD),FC(NDD),FL(NDD),X(NPD,NDD)
     1,W(NDD),WP(NPD+MPD),FOBS(NDD),PRM(NPD+1,MPD),B1(NPD),SUMY(2*NPD),
     2LN(NPD)
      COMMON/ITP/IIN,IOUT
      COMMON/FLT/X
C**FORMAT LIST
    1 FORMAT (6I5,2F10.0)
    2 FORMAT (8F10.0)
    3 FORMAT (' NP    = ',I4,/,' NRES  = ',I4,/,' NH+NQT= ',I4,
     1/,' NPR   = ',I4,/,' MPR   = ',I4,/,' NPTS  = ',I4,/,
     2' CALCULATED ERROR VARIANCE   = ',G11.5,/,
     3' FSTAT = ',G11.5)
    4 FORMAT (/,26X,18HOPTIMUM PARAMETERS
     1/1H ,3X,3(3HNO.,1X,'PID',5X,4HBOPT,8X))
    5 FORMAT (/,9X,'DEPENDENT VARIABLES COMPUTED WITH OPTIMUM PARAME',
     1'TERS',/1H ,3X,3(3HNO.,1X,'DID',5X,4HFOPT,8X))
    6 FORMAT (/,21X,26HPARAMETERS FOR SAMPLE NO. ,I3
     1/1H ,3X,3(3HNO.,1X,'PID',7X,1HB,9X))
    7 FORMAT (/,12X,44HDEPENDENT VARIABLES COMPUTED FOR SAMPLE NO. ,I3
     1/1H ,3X,3(3HNO.,1X,'DID',6X,2HFC,9X))
    8 FORMAT (/,37H SENSITIVITIES FOR OPTIMUM PARAMETERS)
    9 FORMAT (1H0,6X,55HLINEARIZED DEPENDENT VARIABLES COMPUTED FOR SAMP
     1LE NO. ,I3/1H ,3X,3(3HNO.,1X,'DID',6X,2HFL,9X))
   10 FORMAT (/,' USING FSTAT = ',G11.5,', BEALES MEASURE = ',G11.5,/,
     1' IF BEALES MEASURE IS ',
     1'GREATER THAN ',G10.2,', THE MODEL IS NONLINEAR.',/,
     2' IF BEALES MEASURE IS LESS THAN ',G10.2,', THE MODEL IS ',
     3'EFFECTIVELY LINEAR,',/,' AND LINEAR CONFIDENCE INTERVALS ',
```

```
     4'ARE FAIRLY ACCURATE IF THE RESIDUALS',
     5' ARE NORMALLY',/,' DISTRIBUTED.',/)
   11 FORMAT (/,23H SS((FC-FOPT)*W**.5) = ,G11.5
     1/23H SS((FL-FOPT)*W**.5) = ,G11.5)
   12 FORMAT (/,14X,42HRELIABILITY WEIGHTS FOR SAMPLE INFORMATION,
     1/1H ,3X,3(3HNO.,1X,'DID',6X,1HW,10X))
   13 FORMAT (16I5)
   14 FORMAT (/,12X,'STARTING PARAMETERS FROM DATA SET 8',
     1/1H ,3X,3(3HNO.,8X,3HIPR,10X))
   15 FORMAT (/,14X,40HSTANDARD DEVIATIONS OF PRIOR INFORMATION
     1/1H ,3X,3(3HNO.,10X,2HWP,9X))
   16 FORMAT (/,6H EV = ,G11.5)
   17 FORMAT (6F13.0)
   18 FORMAT (/,' DIMENSIONS EXCEED DIMENSIONS IN PARAMETER STATEMENT',
     1    ' -- STOP EXECUTION')
   19 FORMAT (/,9X,'OBSERVED VALUES OF THE DEPENDENT VARIABLES',
     1/1H ,3X,3(3HNO.,1X,'DID',5X,4HFOBS,8X))
  500 FORMAT(/,' MULTIPLE PRIOR NUMBER',I3,
     1    ' ESTIMATE(LOG IF LN>0)=',G10.3,', WEIGHT=',G10.3)
  510 FORMAT(' THE FOLLOWING TABLE SHOWS VALUES OF THE',
     1    ' STATISTIC DESCRIBED BY COOLEY AND',/,
     1    ' NAFF (1990,P.174,TOP OF RIGHT COLUMN).',
     1    ' THE STATISTIC EQUALS THE NONLINEAR SUM',/,
     1    ' OF SQUARED ERRORS EVALUATED FOR EACH DATA',
     1    ' SET (NSSE) MINUS THE SUM OF SQUARED',/,
     1    ' ERRORS FOR THE OPTIMUM PARAMETER VALUES (',G10.3,').',/,
     1    ' IF THE MODEL IS LINEAR, THE STATISTIC SHOULD BE CLOSE TO ',
     1   G10.3,'.',/,
     2    ' IF THE CORRELATIONS BETWEEN PARAMETERS IS SMALL, THE',
     2    ' TABLE SHOWS WHICH INDIVIDUAL',/,
     2    ' PARAMETERS ARE MOST NONLINEAR. THE FIRST PAIR OF PARAMETER',
     2    ' SETS ARE RELATED TO THE',/,
     2    ' FIRST PARAMETER, THE SECOND PAIR ARE RELATED TO THE SECOND',
     2    ' PARAMTER, AND SO ON.',
     3    //,' PARAMETER',17X,'      STATISTIC       PERCENT',/,
     4    '   SET       NSSE    STATISTIC   -',G10.3,'   DIFFERENCE',/)
  520 FORMAT(I6,5X,2G10.3,2X,G10.3,2X,F10.2)
  530 FORMAT(8I10)
  540 FORMAT(/,' PARAMETER',/,'  NUMBER     LN',/)
  550 FORMAT(I6,I9)
  560 FORMAT(20A4)
  600 FORMAT(' END OF BEALE.DAT2.',/,'  MODFLOWP SOLUTION MUST NOT',
     1 ' HAVE CONVERGED FOR NEXT SET OF PARAMETER VALUES.',/,
     2 ' CHECK PARAMETER VALUES; NEGATIVE VALUES CAN BE MADE TO',
     3 ' STAY POSITIVE BY ESTIMATING THE LOG TRANSFORM')
C**DEFINE INPUT AND OUTPUT UNIT NUMBERS
      IIN=1
      IIN2=2
      IOUT=3
C**OPEN FILES
      OPEN (IIN,FILE='BEALE.DAT',STATUS='OLD',ACCESS='SEQUENTIAL'
     1,FORM='FORMATTED')
      OPEN (IIN2,FILE='BEALE.DAT2',STATUS='OLD',ACCESS='SEQUENTIAL'
     1,FORM='FORMATTED')
      OPEN (IOUT,FILE='BEALE.OUT',STATUS='UNKNOWN',ACCESS='SEQUENTIAL'
     1,FORM='FORMATTED')
C**READ BASE DATA
```

```
      READ(IIN,1) NVAR,NRES,NOBS,NPR,MPR,NPTS,VAR,FSTAT
      WRITE(IOUT,3) NVAR,NRES,NOBS,NPR,MPR,NPTS,VAR,FSTAT
      IF(NVAR.GT.NPD.OR.NOBS.GT.NDD.OR.MPR.GT.MPD) THEN
        WRITE(IOUT,18)
        STOP
      ENDIF
      READ(IIN,17) (BOPT(J),J=1,NVAR)
      READ(IIN,560) (PID(J),J=1,NVAR)
      WRITE(IOUT,4)
      CALL PRTOTB(BOPT,PID,NVAR)
      READ(IIN,560) (DID(I),I=1,NOBS)
      READ(IIN,17) (FOPT(I),I=1,NOBS)
      WRITE(IOUT,5)
      CALL PRTOTB(FOPT,DID,NOBS)
      READ(IIN,17) (FOBS(I),I=1,NOBS)
      WRITE(IOUT,19)
      CALL PRTOTB(FOBS,DID,NOBS)
      READ(IIN,2) (W(I),I=1,NOBS)
      WRITE(IOUT,12)
      CALL PRTOTB(W,DID,NOBS)
      DO 20 J=1,NOBS
      READ(IIN,17) (X(I,J),I=1,NVAR)
   20 CONTINUE
      WRITE(IOUT,8)
      CALL PRTOT(X,NVAR,NOBS,NPD)
      IF(NPR.GT.0) THEN
        READ(IIN,17) (B1(I),I=1,NVAR)
        WRITE(IOUT,14)
        CALL PRTOTB(B1,PID,NVAR)
        READ(IIN,17) (WP(I),I=1,NVAR)
        WRITE(IOUT,15)
        CALL PRTOTB(WP,PID,NVAR)
      ENDIF
      IF(MPR.GT.0) THEN
        DO 30 IMP=1,MPR
          READ(IIN,2) (PRM(IP,IMP),IP=1,NVAR+1),WP(NVAR+IMP)
          WRITE(IOUT,500) IMP,PRM(NVAR+1,IMP),WP(NVAR+IMP)
          CALL PRTOTB(PRM(1,IMP),PID,NVAR)
   30   CONTINUE
      ENDIF
      READ(IIN,530) (LN(I),I=1,NVAR)
      WRITE(IOUT,540)
      WRITE(IOUT,550) (I,LN(I),I=1,NVAR)
      DO 40 IP=1,NVAR
   40 IF(LN(IP).NE.0) BOPT(IP)=ALOG(BOPT(IP))
C**READ DATA FOR EACH SAMPLE AND COMPUTE MODIFIED BEALE'S MEASURE, BN,
C**AND THE SATAISTIC FROM COOLEY AND NAFF(1990,P.174,TOP OF RIGHT COLUMN)
      SUMA=0.
      SUMB=0.
      DO 80 M=1,NPTS
      READ(IIN2,2,END=42) (B(J),J=1,NVAR)
      GO TO 43
   42 WRITE(IOUT,600)
      STOP
   43 CONTINUE
      WRITE(IOUT,6) M
      CALL PRTOTB(B,PID,NVAR)
```

```
      DO 45 I=1,NVAR
45    IF(LN(I).NE.0) B(I)=ALOG(B(I))
      READ(IIN2,17) (FC(I),I=1,NOBS)
      WRITE(IOUT,7) M
      CALL PRTOTB(FC,DID,NOBS)
      SUMC=0.
      SUMD=0.
      SUMY(M)=0.
      DO 60 J=1,NOBS
      SUM=FOPT(J)
      DO 50 I=1,NVAR
50    SUM=SUM+X(I,J)*(B(I)-BOPT(I))
      FL(J)=SUM
      TMP=FC(J)-SUM
      SUMA=SUMA+TMP*W(J)*TMP
      TMP=FC(J)-FOPT(J)
      SUMC=SUMC+TMP*W(J)*TMP
      TMP=SUM-FOPT(J)
      SUMD=SUMD+TMP*W(J)*TMP
      TMP=FOBS(J)-FC(J)
      SUMY(M)=SUMY(M)+TMP*W(J)*TMP
60    CONTINUE
      IF(NPR.GT.0) THEN
        DO 70 J=1,NVAR
          IF(WP(J).GT.0.) THEN
            TMP=B(J)-BOPT(J)
            TMP=TMP*WP(J)*TMP
            SUMC=SUMC+TMP
            SUMD=SUMD+TMP
            TMP=B1(J)-B(J)
            SUMY(M)=SUMY(M)+TMP*WP(J)*TMP
          ENDIF
70      CONTINUE
      ENDIF
      IF(MPR.GT.0) THEN
        DO 72 J=1,MPR
          TEMP=0.
          TEMP1=0.
          DO 71 I=1,NVAR
          TEMP1=TEMP1+PRM(I,J)*BOPT(I)
71        TEMP=TEMP+PRM(I,J)*B(I)
        TMP=TEMP-TEMP1
        TMP=TMP*WP(NVAR+J)*TMP
        SUMC=SUMC+TMP
        SUMD=SUMD+TMP
        TMP=PRM(NVAR+1,J)-TEMP
        SUMY(M)=SUMY(M)+TMP*WP(NVAR+J)*TMP
72      CONTINUE
      ENDIF
75    WRITE(IOUT,9) M
      CALL PRTOTB(FL,DID,NOBS)
      WRITE(IOUT,11) SUMC,SUMD
80    SUMB=SUMB+SUMD*SUMD
      TMP=NRES
      BN=TMP*VAR*SUMA/SUMB
      WRITE(IOUT,10) FSTAT,BN,1./FSTAT,.09/FSTAT
      SSTAT=VAR*NRES*FSTAT
```

```
      SSE=VAR*(NOBS+NPR+MPR-NVAR)
      WRITE(IOUT,510) SSE,SSTAT,SSTAT
      WRITE(IOUT,520) (I,SUMY(I),SUMY(I)-SSE,SUMY(I)-SSE-SSTAT,
     1       100.*(SUMY(I)-SSE-SSTAT)/SSTAT,I=1,NPTS)
      STOP
      END


      SUBROUTINE PRTOTB(VAL,VID,NO)
C**PRINT VALUES IN THREE GROUPS OF TWO COLUMNS
      CHARACTER*4 VID(NO)
      DIMENSION VAL(NO)
      COMMON/ITP/IIN,IOUT
      NR=NO/3
      IF(3*NR.NE.NO) NR=NR+1
      DO 10 K=1,NR
      WRITE(IOUT,20) (L,VID(L),VAL(L),L=K,NO,NR)
   10 CONTINUE
      RETURN
   20 FORMAT (1H ,2X,3(I3,2X,A4,1X,G11.5,3X))
      END


      SUBROUTINE PRTOT(C,NR,NC,NRD)
C**PRINT MATRICES DIVIDED VERTICALLY INTO TEN-COLUMN BLOCKS
      DIMENSION C(NRD,NC)
      COMMON/ITP/IIN,IOUT
      DO 60 K=1,NC,10
      J10=K+9
      IF(J10.GT.NC) J10=NC
      WRITE(IOUT,70) (J,J=K,J10)
      WRITE(IOUT,90)
      DO 30 I=1,NR
   30 WRITE(IOUT,80) I,(C(I,J),J=K,J10)
   60 CONTINUE
   70 FORMAT(1H0,10(9X,I3))
   80 FORMAT (1H ,I3,1X,10(1X,G11.5))
   90 FORMAT (1H )
      RETURN
      END
```

REFERENCES

Anderson, M.P. and Woessner, W.W., 1992, Applied groundwater modeling simulation of flow and advective transport: New York, Academic Press, 381 p.

Benjamin, J.R. and Cornell, C.A., Probability, statistics, and decision for civil engineers: New York, McGraw-Hill, 684 p.

Cooley, R.L., 1977, A method of estimating parameters and assessing reliability for models of steady state groundwater flow, 1, Theory and numerical properties: Water Resources Research, v. 13, no. 2, p. 318-324.

Cooley, R.L., 1979, A method of estimating parameters and assessing reliability for models of steady state groundwater flow, 2, Application of statistical analysis: Water Resources Research, v. 15, no. 3, p. 603-617.

Cooley, R.L. and Naff, R.L., 1990, Regression modeling of ground-water flow: U.S. Geological Survey Techniques of Water-Resources Investigations, book 3, chap. B4, 232 p.

Dettinger, M.D. and Wilson, J.L., 1981, First-order analysis of uncertainty in numerical-models of groundwater flow, 1, Mathematical development: Water Resources Research, vol. 17, no. 1, p. 149-161.

Donaldson, J.R. and Schnable, R.B., 1987, Computational experience with confidence regions and confidence intervals for nonlinear least squares: Technometrics, vol. 29, no. 1, p. 67-87.

Draper, N.R. and Smith, Harry, 1981, Applied regression analysis: (2nd ed.): New York, Wiley, 709 p.

Graybill, F.A., 1976, Theory and application of the linear model: North Scituate, Mass., Duxbury Press, 704 p.

Hazen, A., 1914, Storage to be provided in the impounding reservoirs for municipal water supply: Transactions of the American Society of Civil Engineers, v. 77, p. 1547-1550.

Helsel, D.R. and Hirsch, R.M., 1992, Statistical methods in water resources: New York, Elsevier, 522 p.

Hill, M.C., 1989, Analysis of accuracy of approximate, simultaneous, nonlinear confidence intervals on hydraulic heads in analytical and numerical test cases: Water Resources Research, vol. 25, no. 2, p. 177-190.

Hill, M.C., 1992, A computer program (MODFLOWP) for estimating parameters of a transient, three-dimensional ground-water flow model using nonlinear regression: U.S. Geological Survey Open-File Report 91-484, 358 p.

Looney, S. W. and Gulledge, T. R., 1985, Probability plotting positions and goodness of fit for the normal distribution: The Statistician, v. 34, p. 297-303.

McDonald, M.G. and Harbaugh, A.W., 1988, A modular three-dimensional finite-difference ground-water flow model: U.S. Geological Survey Techniques of Water-Resources Investigations, book 6, chap. A1, 548 p.

Miller, R.G. Jr., 1981, Simultaneous statistical inference: second edition, New York, Springer-Verlag, 299.

Turner, P.J., 1992, ACD/gr User's manual -- Graphics for exploratory data analysis: Center for Coastal and Land-Margin Research, Oregon Graduate Institute of Science and Technology, 19600 NW von Neumann Dr., Beaverton, Oregon, 97006-1999, Software Documentation Series, SDS3, 91-3.

Vecchia, A.V. and Cooley, R.L., 1987, Simultaneous confidence and prediction intervals for nonlinear regression models with application to a groundwater flow model: Water Resources Research, vol. 23, no. 7, p. 1237-1250.

Yeh, W.W-G., 1986, Review of parameter identification procedures in groundwater hydrology--the inverse problem: Water Resources Research, v. 22, no. 2, p. 95-108.

☆ U.S. GOVERNMENT PRINTING OFFICE: 1994—573-191 / 80019