Chapter 4. USER GUIDE TO RESAN2-2k

RESAN2-2k performs two functions. The first function is to test residuals for indications of non-normality. This purpose is shared by earlier codes like RESAN (Cooley and Naff, 1990) and RESAN-2000 (Hill and others, 2000), but the analysis is somewhat different and more precise when using RESAN2-2k. The second function is to compute measures that can be used to detect model and system types of intrinsic nonlinearity.

Testing Weighted Residuals for Indications of Non-Normality

Like earlier versions of the residuals analysis program, RESAN2-2k generates sets of synthetic weighted residuals having the distribution given by equation 30 or 31. A normal probability plot of the actual weighted residuals distribution together with the synthetic sets will give a preliminary idea as to whether or not the weighted residuals appear to have the specified normal distribution, as described by Cooley and Naff (1990, p. 168-171). If a large number of synthetic sets of weighted residuals are generated, then the mean values, $\bar{\boldsymbol{d}}$ (equation 32), and standard deviation, $\boldsymbol{v}$ (equation 34), over all realizations can be computed for the synthetic weighted residuals. The necessary number to obtain stable results is usually from several hundred to several thousand. A normal probability plot of $\bar{\boldsymbol{d}} \pm 2\boldsymbol{v}$ gives a band within which the weighted residuals might usually be expected to lie. Figure 1 shows an example of a normal probability plot with weighted residuals plotted together with the corresponding band given by $\bar{\boldsymbol{d}} \pm 2\boldsymbol{v}$ computed from 1,000 realizations. The results used to produce this kind of plot are all written by RESAN2-2k to a file with extension _md.

RESAN2-2k computes a measure of the correlation, c_d (equation 38), between weighted residuals and the means of the synthetic residuals, and it computes the probability P that c_d or a smaller value can be obtained. The larger the probability P, the larger is the probability that the weighted residuals have the theoretical distribution given by equation 30 or 31. RESAN2-2k also computes and prints 99, 95, and 90 percent confidence limits for the correlation. These measures are written by RESAN2-2k to the file with extension #nr, where they are listed the following way:

```
CORRELATION (CED) --------------- = 0.98984
PROBABILITY OF CORRELATION (PROB) = 0.71300
99% CONFIDENCE LIMIT (CL99) ----- = 0.99624
95% CONFIDENCE LIMIT (CL95) ----- = 0.99423
90% CONFIDENCE LIMIT (CL90) ----- = 0.99304
```

CED corresponds to c_d, PROB corresponds to P, CL99 corresponds to the 99 percent confidence limit, CL95 corresponds to the 95 percent confidence limit, and CL90 corresponds to the 90 percent confidence limit.

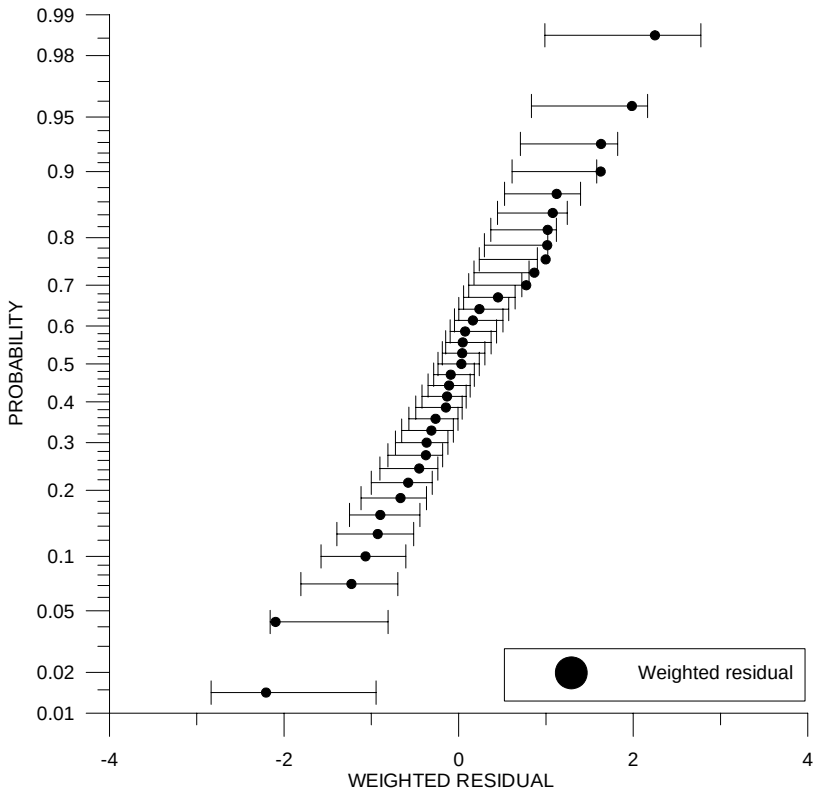

Figure 1. Ordered weighted residuals plotted with $\bar{d} \pm 2v$ confidence band (approximate 95 percent confidence intervals).

Testing for Intrinsic Nonlinearity

As described in chapter 2, RESAN2-2k computes three measures that can be used to detect model intrinsic nonlinearity: the mean weighted residual, the slope of weighted residuals in relation to weighted model functions, and $(\boldsymbol{Y} - \boldsymbol{f}(\gamma\hat{\boldsymbol{\theta}}))'\boldsymbol{\omega}^{1/2}\boldsymbol{R}\boldsymbol{\omega}^{1/2}(\boldsymbol{Y} - \boldsymbol{f}(\gamma\hat{\boldsymbol{\theta}}))$. If the first two measures are small (close to zero), and if the third measure is much smaller than $S(\hat{\boldsymbol{\theta}})$ (the weighted sum of squared residuals computed using the optimum parameter set, $\hat{\boldsymbol{\theta}}$), then the model intrinsic nonlinearity probably is small. In the RESAN2-2k output file with extension #nr, the measures are listed as

```
MEAN WEIGHTED RESIDUAL (EM) --------- = 0.99903E-01
SLOPE OF WEIGHTED RESIDUAL PLOT (SLP) = 0.56149E-05
INTRINSIC NONLINEARITY MEASURE (QINT) = 0.56149E-05 (SHOULD BE<<36.506)
```

EM and SLP are both close to 0, and QINT is much smaller than 36.506, which is the weighted sum of squared residuals at the optimum in this example. The three measures thus all indicate that model intrinsic nonlinearity is small.

Input Instructions

RESAN2-2k needs two input files that can be generated by MODFLOW-2000, the _rs file and the _ws file (Hill and others, 2000, tables 5 and 8). The first file is the input file also used by the RESAN-2000 program of Hill and others (2000). The content of the second file is weighted simulated functions and weighted residuals. To generate the files MODFLOW-2000 must be run in the Sensitivty Analysis or Parameter-Estimation mode, and OUTNAM in the Observation Process input file must be specified as a string other than "NONE." It is important that the two files be generated in two different steps using MODFLOW-2000:

1. The _rs file must be generated using a parameter set other than the optimal, $\hat{\boldsymbol{\theta}}$.
2. The _ws file must be generated using the optimal parameter set, $\hat{\boldsymbol{\theta}}$.

Before running RESAN2-2k the following changes must be made to the _rs file. First, NSETS in item 1 defines the number of sets of synthetic weighted residuals generated to test for normality of the weighted residuals. NSETS must be set to a value of several hundred or a few thousand. Second, values of additional input variables used by RESAN2-2k (but not by RESAN-2000) can be added to item 1 as described in the 'Content of the _rs Input File Needed by RESAN2-2k' section if special features of the program are wanted. Third, depending on the value of two of the variables added to item 1, it may be necessary to append two new arrays, ET and V, to the file. The content of the input files is described in the 'Content of the _rs Input File Needed by RESAN2-2k' and the 'Content of the _ws Input File Needed by RESAN2-2k' sections.

Output

The RESAN2-2k program produces two files with extensions #nr and _md. The #nr file contains input values and summarizes the results from testing for nonlinearity and from testing weighted residuals for indications of non-normality. The summary of results is described in the 'Testing Weighted Residuals for Indications of Non-Normality' and the 'Testing for Intrinsic Nonlinearity' sections. The _md file contains results that can be used to produce a probability plot of ordered weighted residuals together with the ordered means of simulated weighted residuals and a confidence band (for example, fig. 1). The content of the columns in the file is: (1) ordered weighted residuals; (2) ordered means of simulated weighted residuals; (3) standard deviations of simulated weighted residuals; (4) two times the standard deviations of simulated weighted residuals; (5) frequency/probability; probability plot position; (6) observation name; and (7) plot symbol identifier.

Content of the _rs Input File Needed by RESAN2-2k

RESAN2-2k reads the following from the _rs file. Explanations of the variables follow this list.

1. `NPAR NOBS NH NQT MPR IPR NSETS NRAN VAR STDV IPCF IPRN IERT ICOV`
   (format: 7I5, I10, F25.0, F10.0, 4I5)
2. `PARNAM(NPAR)` (format: 6(A10,1X))
3. `COV(NPAR,NPAR)` (format: 16F25.0)
4. `WT(NH)` (format: 16F15.0)
5. `WQ(NQT,NQT)` (format: 16F15.0)
6. `X(NPAR,NOBS)` (format: 16F15.0)
7. `PRM(NP,I), WP(I), I=1,MPR` (format: 8F15.0)
8. `NIPR(IPR)` (format: 16I5)
9. `WTPS(IPR,IPR)` (format: 8F15.0)
   If IERT is greater than zero, read item 10.
10. `ET(NOBS)` (free format)
    If ICOV is greater than zero, read item 11.
11. `V(NOBS,NOBS)` (free format)

The variables have the following explanations.

NPAR---is the number of estimated parameters.

NOBS---is the number of model function observations used for estimation. NOBS must thus be equal to NH+NQT.

NH-------is the number of head observations.

NQT-----is the number of observations other than heads.

MPR-----is the number of prior information equations

IPR------is the number of prior information observations with a full weight matrix.

NSETS—is the number of sets of synthetic weighted residuals generated to test for normality of the weighted residuals. NSETS must be increased from 4 (the default value set by MODFLOW-2000) to a value of several hundred or a few thousand.

NRAN---is the seed for the random number generator.

VAR-----is the calculated error variance, $\hat{s}^2 = S(\hat{\boldsymbol{\theta}})/(n-p)$.

STDV---is the theoretical error variance, $b\sigma_\varepsilon^2$, used to generate synthetic weighted residuals distributed as given by equation 30. If a value of 0 is specified, RESAN2-2k automatically sets it equal to $\hat{s}^2$, the calculated error variance from the least squares estimation, before the synthetic weighted residuals are generated.

IPCF-----is a flag. If IPCF is greater than zero, the program prints the ***I–R*** matrix.

IPRN----is a printing flag. Weighted simulated residuals are printed when IPRN is greater than zero.

IERT----is a flag. If IERT is greater than zero, the program reads the ET array described below.

ICOV----is a flag. If ICOV is greater than zero, the program reads the V array described below.

PARNAM—is the parameter name list.

COV-----is the parameter covariance matrix.

WT------are the square roots of weights for heads.

WQ------is the square root of the full weight matrix for observations other than heads.

X---------is the sensitivity matrix for all parameters and observations. (For log-transformed parameters the sensitivities are with respect to natural log-transformed parameters, not with respect to $\log_{10}$-transformed parameters.)

PRM----are coefficients for the prior information equations.

WP------are weights for prior information equations.

NIPR----is the number list for parameters with prior information observations that have a full weight matrix.

WTPS---is the square root of the full weight matrix for correlated prior information. (For log-transformed parameters the weights are for natural log-transformed parameters, not for $\log_{10}$-transformed parameters.)

ET-------are the expected values of the true errors for the model function observations, $E\left(\boldsymbol{Y}_m - \boldsymbol{f}_m\left(\boldsymbol{\gamma\theta}_*\right)\right)$, which are used to generate the synthetic, weighted, true error vectors used to test weighted residuals for normality. If IERT=0, then ET is not read but is assumed to be all zeros.

V------ is the matrix $\boldsymbol{\Omega}_m = E\left(\boldsymbol{Y}_m - \boldsymbol{f}_m\left(\boldsymbol{\gamma\theta}_*\right)\right)\left(\boldsymbol{Y}_m - \boldsymbol{f}_m\left(\boldsymbol{\gamma\theta}_*\right)\right)' \Big/ \sigma_\varepsilon^2$ for model function observations that is used to generate the synthetic, weighted, true error vectors from equation 30. If ICOV=0, then V is not read and equation 31 is used instead of equation 30.

Content of the _ws Input File Needed By RESAN2-2k

RESAN2-2k reads the following items from the _ws file NTOT times, where NTOT= NOBS+MPR+IPR. That is, the items are read once for each observation that was used to estimate the model parameters.

1. `F E ISYM DID` (format: 2(G15.7,1X),I5,2X,A)

The variables have the following explanations.

F---------is the weighted simulated equivalent to the observation.

E---------is the corresponding weighted residual.

ISYM---is a plot symbol identifier (an integer value).

DID-----is the observation name.