

## Chapter 6. USER GUIDE TO CORFAC-2k

The purpose of CORFAC-2k is to compute correction factors for confidence regions, confidence intervals, and prediction intervals. In the development of the procedures implemented in CORFAC-2k, Cooley (2004, section 5) assumed that the model intrinsic nonlinearity and model combined intrinsic nonlinearity are negligible. As described in previous chapters, the validity of this assumption can be investigated using BEALE2-2k and RESAN2-2k, and by analyzing weighted residuals computed by the UNC Process.

The correction factors are to be used to calculate the critical value ( $D_{1-\alpha}^2$ , equation 86) needed by UNC for the calculation of confidence and prediction intervals, and by BEALE2-2k for computation of nonlinearity measures. CORFAC-2k is used first to compute correction factors, and then BEALE2-2k, RESAN2-2k, and weighted residuals computed by UNC are used to analyze for model intrinsic nonlinearity and model combined intrinsic nonlinearity. If this indicates that model intrinsic nonlinearity and/or model combined intrinsic nonlinearity are not small then the correction factors computed by CORFAC-2k, and thus the confidence or prediction intervals computed by UNC, may be inaccurate.

When the true error covariance structure,  $\Omega$ , is known,  $c_r$  is computed from equation 66,  $c_c$  is computed from equation 68, and  $c_p$  is computed from equation 70. In the output file these correction factors are called CR, CC, and CP, respectively.

When  $\Omega$  is unknown, the approximate bound for  $c_r$  is computed from equation 82, the approximate bound for  $c_c$  is computed from equation 83, and the approximate bound for  $c_p$  is computed from equation 84. In the output file these approximations are called CRB, CCB, and CPB, respectively. CRB, CCB and CPB also are written in the output file when  $\Omega$  is known.

### Input Instructions

The correction factors are calculated by CORFAC-2k using an input file with the extension `_cf` that can be generated by MODFLOW-2000 using the UNC Process. To generate the file, MODFLOW-2000 must be run in the Uncertainty-Process mode (with `IACT=3`), and `OUTNAM` in the Observation Process input file must be specified as a string other than "NONE." Before running MODFLOW-2000, the final calibrated parameter values must be substituted from the `_b` file into the Sensitivity Process input file. The content of the input file is described in detail in the 'Content of the `_cf` Input File Needed By CORFAC-2k' section.

If the effective correlation  $c_e$  and/or  $\Omega$  is known, it is necessary to manually make the following changes to the `_cf` file produced by MODFLOW-2000 before running CORFAC-2k:

1. Change the value for the effective correlation, EC, from -1 (which is the value always written by MODFLOW-2000) to a value in the range from 0.0 to 1.0. If the value for EC is specified to be less than 0.0 or larger than 1.0, then CORFAC-2k automatically changes the value to 0.8.

2. Set ITRN to a value larger than 0 if  $\Omega$  is known, and append V, VP and CP to the file. VP and CP are only needed if correction factors are to be computed for prediction intervals.

## Output From CORFAC-2k

The CORFAC-2k program produces an output file with the extension #cf. When  $\Omega$  is known, the output for a confidence region has the form

```
CORRECTION FACTOR FOR N-P (A) ---- = 6.0528
SCALING FACTOR (B) ----- = 1.0000
EFFECTIVE CORRELATION (C) ----- = 0.77891
(N-P)/(N-A*P) VALUE (FAC) ----- = 4.1308
CORRECTION FACTOR (CR) ----- = 25.003

APPROXIMATE N-A*P VALUE (DFB) ---- = 4.4313
APPROX. (N-P)/(N-A*P) VALUE (FACB) = 4.5134
APPROXIMATE BOUND FOR CR (CRB) --- = 27.936
```

where the number of observations,  $n$ , used for parameter estimation is called N; the number of estimated parameters,  $p$ , is called P;  $a$  computed from equation 26 is called A;  $b$  computed from equation 27 is called B;  $c$  computed from equation 75 is called C;  $c_r$  computed from equation 66 is called CR;  $n - ap$  computed from equation 71 using the input value  $c_e$  (EC in input file) for  $c$  is called DFB; and  $c_b$  approximated from equation 82 is called CRB. If  $\Omega$  is unknown, then only DFB, FACB, and CRB are printed.

For each confidence interval the output is

```
CORRECTION FACTOR XI FOR C. I. ----- = 2.9713
CORRECTION FACTOR (CC) ----- = 12.274

ESTIMATED VARIANCE OF PREDICTION (VE) = 0.39250
BOUND FOR XI USING POS. VALUES (XIPB) = 4.8727
BOUND FOR XI USING NEG. VALUES (XINB) = 0.38389E-01
APPROXIMATE BOUND FOR CC (CCB) ----- = 18.497
```

where  $\xi$  computed from equation 47 is called XI;  $c_e$  computed from equation 68 is called CC; VE is computed as  $VE = \mathbf{Q}'\mathbf{Q} S(\hat{\theta})/(n - p)$ ; XIPB is the squared sum of positive values of the  $Q_{mi}$ 's used in equation 78; XINB is the squared sum of negative values of the  $Q_{mi}$ 's used in equation 78; and  $c_e$  approximated from equation 83 is called CCB. If  $\Omega$  is unknown, then only VE, XIPB, XINB, and CCB are printed.

For a prediction interval the output is

```
CORRECTION FACTOR XI FOR P. I. ----- = 0.26079
CORRECTION FACTOR (CP) ----- = 1.0773

ESTIMATED VARIANCE OF PREDICTION (VE) = 0.15901
BOUND FOR XI USING POS. VALUES (XIPB) = 0.77966
BOUND FOR XI USING NEG. VALUES (XINB) = 0.93736
APPROXIMATE BOUND FOR CP (CPB) ----- = 4.2872
```

where  $\xi_a$  computed from equation 55 is called XI,  $c_p$  computed from equation 70 is called CP; VE is again computed as  $VE = \mathbf{Q}'\mathbf{Q}S(\hat{\theta})/(n-p)$ ; XIPB is the squared sum of positive values of the  $Q_{mi}$ 's used in equation 81; XINB is the squared sum of negative values of the  $Q_{mi}$ 's used in equation 81; and  $c_p$  approximated from equation 84 is called CPB. If  $\mathbf{\Omega}$  is unknown, only VE, XIPB, XINB, and CPB are printed.

### Content of the \_cf Input File Needed By CORFAC-2k

The \_cf file contains the following items.

1. NPE NOBS NQ MPR IPR VAR EC ITRN (format: 5I10,2F14.0,I10)
2. PARNAM(NPE) (format: 6(A10,1X))
3. OBSNAM(NOBS) (format: 6(A12,1X))
4. WH(NH) (NH=NOBS-NQ) (format: 8F15.0)
5. WQ(NQ,NQ) (format: 8F15.0)
6. X(NPE,NOBS) (format: 6F13.0)
7. PRM(NPE,J), WP(J), J=1,MPR (format: 8F15.0)
8. NIPR(IPR) (format: 8I10)
9. WTP(IPR,IPR) (format: 6F13.0)
10. NOINT NHI NQI NPI ITYP IDIF (format: 6I10)
11. GNAM(NOINT) (format: 6(A12,1X))
12. WG(NOINT) (format: 8F15.0)
13. Z(NPE,NOINT) (format: 6F13.0)
- If ITRN is greater than zero, read the following item.
14. V(NOBS,NOBS) (free format)
- If ITRN is greater than zero and ITYP is greater than 1, read the following item NOINT times.
15. VP CP(NOBS) (free format)

The variables are defined as:

.

NPE-----is the number of estimated parameters.

NOBS---is the number of observations used for estimation.

NQ-----is the number observations other than hydraulic heads.

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

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

VAR-----is the calculated error variance,  $\hat{s}^2$ .

EC-----is the effective correlation,  $c_e$ .

ITRN----is a flag that when specified to be greater than zero directs CORFAC-2k to read and use V, VE, and CP (defined below).

PARNAM—are the parameter names.

OBSNAM—are the observation names.

WH-----are the square root of the weights for hydraulic heads.

WQ-----is the square root of the full weight matrix for the 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 the square root of weights for prior information equations.

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

WTP----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.)

NOINT—is the number of intervals for which a correction factor is calculated.

NHI-----is the number of intervals for hydraulic heads.

NQI-----is the number of intervals for head-dependent flows.

NPI-----is the number of intervals for parameters.

ITYP---is a flag for the type of interval: ITYP=1 is for confidence intervals; ITYP=2 is for prediction intervals.

IDIF-----is a flag. When IDIF=1 it indicates in the output that intervals pertaining to heads and head-dependent flows are for differences of heads or flows. When IDIF=0 it indicates that the intervals pertain to heads and head-dependent flows.

GNAM—is the name of each prediction for which an interval correction factor is calculated.

WG-----are the square roots of the weights for the predictions corresponding to  $\omega_p$ . Values must always be specified for WG whenever NOINT is greater than 0, but WG is only used in calculations pertaining to prediction intervals. For prediction intervals for log-transformed parameters WG must be specified for the natural log-transforms of the parameters, not for the  $\log_{10}$ -transforms.

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

V-----is equal to  $\mathbf{\Omega}_m = E(\mathbf{Y}_m - \mathbf{f}_m(\boldsymbol{\gamma}\boldsymbol{\theta}_*))(\mathbf{Y}_m - \mathbf{f}_m(\boldsymbol{\gamma}\boldsymbol{\theta}_*))' / \sigma_\varepsilon^2$ .

VP-----is equal to  $\hat{\omega}_p^{-1} = E(Y_p - g(\boldsymbol{\gamma}\boldsymbol{\theta}_*))^2 / \sigma_\varepsilon^2$ .

CP-----is equal to  $\mathbf{C} = E(\mathbf{Y}_m - \mathbf{f}_m(\boldsymbol{\gamma}\boldsymbol{\theta}_*))(\mathbf{Y}_p - g(\boldsymbol{\gamma}\boldsymbol{\theta}_*))' / \sigma_\varepsilon^2$ .