ADDITIONS TO THE UCODE MANUAL APPLICABLE TO UCODE 3.0

Until the next of the formal release of the UCODE manual, a working copy of the manual can be maintained by inserting the following sections in the original copy of the manual. A hard copy can be ordered by contacting igwmc@mines.edu.

THIS NEW DISCUSSION OF WEIGHTING (with particularly useful application to concentration observations) IN UCODE SHOULD FOLLOW TABLE 1, ON PAGE 12 OF THE ORIGINAL MANUAL

Weighting Observations

Weights are applied to the residuals to reflect their relative certainty, and thus importance to obtaining a close fit between the simulated value and the field observation. Generally, the inverse of the variance of the measurement is used to weight squared-residuals. This can be determined for practical applications by converting approximate knowledge of the accuracy of a measurement into more exact terms through a statement such as: "I am 95% confident that this measurement is within 1 meter." From a table describing the normal distribution, one determines that 95% confidence is associated with a standard deviation of 1.96. Consequently, 1 meter represents 1.96 standard deviations. It follows that one standard deviation is 1/1.96 meters, or 0.51 meters. The variance is the square of the standard deviation, or 0.26 meters squared. Finally the weight applied to the squared residual is the inverse of the variance, or 3.84 square-meters^{-1}. If an observation is the result of the difference between two measurements (e.g. seepage to a stream is the difference between the measured stream flow at two locations), the variance is calculated for each measurement and summed to determine the variance on the difference.

Errors associated with some types of measurements (e.g. hydraulic head in a ground-water system) are independent of the magnitude of the observation. While for some types of measurements, the error is related to the magnitude of the true value in the field setting (e.g. concentration in a ground-water system). Consequently, it is appropriate to allow weights to be proportional the value. Such an approach facilitates attainment of equally desirable fit to both large and small values. The true concentrations are not known, so either the observed or the simulated values need to be used to determine the weight.

Researchers including Barlebo and others, 1998; Sonnenborg and others, 1996; Gailey and others, 1991; and Wagner and Gorelick, 1987, experimented with using the simulated concentration to determine the weight. These researchers used the same

relative weight for any given type of observation. In UCODE the user defines a dimensionless factor, cv_j, that reflects the confidence associated with the simulated value, $\hat{c}_j$, thus the weight is:

$$w_{c_j} = \frac{1}{\left(cv_j \hat{c}_j\right)^2}$$
(3)

Keidser and Rosbjerg (1991) and Van Rooy and others (1989) used the observed values, $\tilde{c}_j$, to calculate weights. Keidser and Rosbjerg (1991) added a constant, η, to prevent large differences between simulated and observed values at low concentrations from dominating the regression, thus the weight is:

$$w_{c_j} = \frac{1}{\left(cv_j \tilde{c}_j + \eta\right)^2}$$
(4)

The estimated parameter values are sensitive to the value of the constant (Van Rooy and others, 1989). If the constant is too small then large deviations at the periphery of the plume will dominate. Van Rooy and others, suggest setting η to 10-percent of the source strength.

        Use of observed concentration to determine the weight can result in biased estimated parameter values (Anderman and Hill, 1999). Calculation of weights using observed concentration produced biased parameter estimates (average parameter estimates were significantly, as much as 35 percent, different than the true values). Calculation of weights using simulated concentrations produced unbiased parameter estimates. Anderman and Hill (1999) demonstrated that this deviation is due to the fact that the weighted residuals based on observed concentrations are not normally distributed, even when the true errors are normal, while the weighted residuals based on the simulated concentrations are normally distributed. They further indicated that using simulated values to calculate the weights was problematic when parameter values are not close to the optimal parameter values because the simulated values and, thus, the weights, are unrealistic. They improved the regression by using observed concentrations to calculate weights for early parameter iterations and using simulated concentrations when estimated parameter values changed less than 10 percent between parameter-estimation

iterations. UCODE makes this approach available and uses the scaling parameter described below to calculate the weights. This is described in more detail in the section entitled: Universal Input File—*filename.uni.*

Gailey and others (1991) introduced a scaling parameter to the weight equations to ensure that the hydraulic-head and concentration residuals form a single population with a uniform variance and that both types of observations contribute equally to the calibration. It is initially set to 1.0 and is subsequently calculated at each parameter-estimation iteration as,

$$\lambda_r = \frac{\dfrac{S_h(\underline{b}_r)}{nh}}{\dfrac{S_c(\underline{b}_r)}{nc}} \tag{5}$$

which is, the ratio of the average weighted squared residual for hydraulic heads and the average weighted squared residual for concentrations. λ_r replaces the numerator in the equations 3 and 4. UCODE allows the user to choose whether to use this alternative weighting scheme and to define: the individual weighting factors for each observation; which observations receive the alternative weighting, which observations are included in the numerator of the scaling parameter, the constant η, and the threshold level for using the simulated rather than the observed value for calculating the weight.

**THIS NEW DISCUSSION OF A THE UCODE RUN COMMAND, REPLACES
THE PREVIOUS DISCUSSION ON PAGES 23 & 24 OF THE ORIGINAL
MANUAL**

Run Command

The run command needs to be executed in the directory containing the UCODE
input files. An example run command is:

perl   ucode *fn [optional open window flag]*

where: "perl" is the command necessary to invoke the Perl program on the computer
where the code is installed; and "ucode" is the pathname for the UCODE Perl script;
filename is a file name prefix that needs to be replaced using any name that conforms
with file name restrictions of the operating system being used and an open window flag
of 1 keeps the DOS window open to allow viewing of error messages until the user
presses the "enter" key, while a 0 or absence of a flag causes the window to close upon
termination. The only exception is that spaces are not allowed in *fn*, even on operating
systems that allow spaces in file names.

THIS NEW DISCUSSION OF A THE UCODE TOL PARAMETER, REPLACES THE PREVIOUS DISCUSSION ON PAGE 26 OF THE ORIGINAL MANUAL

TOL [*optional UCODE version 3.0 and higher* **MARQUARDT_DIRECTION FACTOR**

INCREMENT]

When parameter values change less than the fractional amount define by the value of TOL between regression iterations, parameter estimation converges (0.01 is recommended). The Marquardt parameter is utilized during the regression when difficult mathematical conditions arise, typically resulting from a poorly posed problem. The MARQUARDT_DIRECTION, and FACTOR and INCREMENT are optional input items and can be controlled by the user in UCODE version 3.0 and higher. The direction is the angle (in degrees) between the down gradient direction and the direction of the parameter update vector. If this angle approaches 90°, the regression will not make progress towards the minimum sum-of-weighted-squared residuals. The default is 85°. The Marquardt parameter, u_r, starts at zero and for iterations in which the condition described by Cooley and Naff (1990, p.71-72) is met using $u_r = 0$, u_r is increased as:

$$u^{new}_r = FACTOR\ u^{old}_r + INCREMENT$$

If FACTOR and INCREMENT are not specified, the defaults are 1.5 and 0.001 respectively.

AN OPTIONAL INPUT ITEM CAN BE INCLUDED BEFORE THE LIST OF OBSERVATIONS IN THE *.UNI FILE, THIS SHOULD BE PLACED BETWEEN THE EXPLANATION OF THE "NUMBER-RESIDUAL-SETS" AND "LIST OF OBSERVATIONS" ON PAGE 28 OF THE ORIGINAL MANUAL

OPTIONAL FLAG - If the user has an extensive list of observations, the time required to compare each observation name to every previously entered name can become excessive. To avoid this check include an additional line before the observation list with the following flag beginning in column 1, otherwise omit this line entirely:

DO_NOT_CHECK_OBSERVATION_NAMES

*DO NOT CHECK OBSERVATION NAMES OPTION _*UCODE version 3 and higher OPTIONAL LINE OF INPUT_: **-** If the user has an extensive list of observations, the time required to compare each observation name to every previously entered name can become excessive. To avoid this check include the following:

DO_NOT_CHECK_OBSERVATION_NAMES

on a line by itself immediately before the first observation in the uni file.

**THIS NEW DISCUSSION OF A THE STAT-FLAG FOR OBSERVATIONS IN
THE *.UNI FILE, REPLACES THE PREVIOUS DISCUSSION ON PAGE 29 OF
THE ORIGINAL MANUAL**

STAT-FLAG - A flag indicating whether STATISTIC is a variance,
standard deviation, or coefficient of variation.

STAT-FLAG	STATISTIC
-1	cv_j of alternate weighting scheme, described under options below, UCODE version 3 and higher
0	variance
1	standard deviation
2	coefficient of variation

**If any observations are assigned a negative stat-flag, then the
ALTERNATE WEIGHTING OPTION must be used and the
CONSTANT and THRESHOLD must be specified at the end of the
list of observations** as discussed in the section *Universal Input File
UCODE Version 3.0 Additions—filename.uni* below.

OBSERVATIONS CAN NOW BE ASSOCIATED WITH X,Y,Z COORDINATES AS DESCRIBED IN THIS DISCUSSION WHICH SHOULD APPEAR AFTER THE PLOT-SYMBOL DISCUSSION ON PAGE 29 OF THE ORIGINAL MANUAL

If the user wishes to relate the observations to x, y, z, locations and/or to points in time then *fn.xyz* file is created by the user with the format:

observation-ID x-coordinate y-coordinate t-coordinate t-coordinate

The observations may be listed in any order, but their names must exactly match the names in the *fn.uni* file and all observations listed in *fn.uni* must be included in *fn.xyz*. If the file *fn.xyz* exists in the directory where UCODE is launched, UCODE will place "x, y, z, t" at the end of the line in each plot file for which a link to an x,y,z,t is appropriate. In files where data related to prior information is included, the x, y, z, t values will be printed as 0, 0, 0, 0.

THIS NEW DISCUSSION OF THE END COMMAND IN THE *.UNI FILE, REPLACES THE PREVIOUS DISCUSSION ON PAGE 29 OF THE ORIGINAL MANUAL

END - Indicates the end of the data input for this file**.** UCODE is not case-sensitive to the word "END". **If the modeler chooses to use one of UCODE's optional features, then END will not appear at this point in the uni file**, but will instead appear at the end of the file after all the optional commands. The description of flags for optional features follows. The flags are not case-sensitive. Because some computers are sensitive to eth presence of a carriage return at the end of the final line of input, it is good practice to include a blank line at the end of all UCODE input files.

THERE ARE MANY NEW, OPTIONAL, INPUT ITEMS FOR UCODE3.0 THAT CAN BE INCLUDED BEFORE THE END COMMAND OF THE *.UNI FILE, THE FOLLOWING PAGES SHOULD BE PLACED AFTER THE DISCUSSION OF THE END COMMAND ON PAGE 29 OF THE ORIGINAL MANUAL

Universal Input File UCODE Version 3.0 Additions—filename.uni

New items related to UCODE version 3.0 follow. UCODE version 3.0 is completely back compatible with previous UCODE input files.

ALTERNATE WEIGHTING OPTION – Unique weighting allows the user to vary the weight throughout the regression for some of the observations. It was designed particularly with the idea of adjusting the weight of concentration observations in a ground-water model, but may be applicable to situations in other disciplines. If the user specifies unique weighting by a negative stat-flag for any observation, the first line following the last observation in the *filename.uni* file needs to contain the flag UNIQUE_WEIGHT (not case-sensitive) followed by the values for CONSTANT and THRESHOLD:

UNIQUE_WEIGHT CONSTANT THRESHOLD

where, the <u>non-zero, positive</u> constant, η, is used to calculate weights using <u>observed</u> values for observations with negative stat-flag values during <u>early parameter iterations</u> as follows:

$$w_{\widetilde{c}_j} = \frac{\lambda_{r-1}}{\left(cv_j\widetilde{c}_j + \eta\right)^2}$$

(6)

where:

cv_j is a dimensionless user-defined weight factor, calculated from the "STAT" value

$\widetilde{c}_j$ is the jth observed concentration

η is a <u>non-zero, positive</u> **CONSTANT**. Keidser and Rosbjerg (1991) suggested adding the constant to prevent large relative differences at low concentrations from dominating the regression. Van Rooy and others (1989) found that the estimated parameter values are sensitive to the value of the constant that is chosen. A small value for the constant caused large residuals at the periphery of the

11

plume to dominate the regression. They suggested using a value equal to 10-percent of the source strength for η.

During <u>later parameter iterations,</u> when the estimated parameter values are changing by less than the **THRESHOLD** value (0.1 is a reasonable choice when working with a tolerance of 0.01), weights are calculated using <u>simulated</u> values for observations with negative stat-flag values as follows:

$$w_{\tilde{c}_j} = \frac{\lambda_{r-1}}{\left(cv_j \hat{c}_j\right)^2} \tag{7}$$

or, if the simulated value is less than or equal to zero, then:

$$w_{\tilde{c}_j} = \frac{\lambda_{r-1}}{\left(\eta\right)^2} \tag{8}$$

where:

$\hat{c}_j$ is the jth simulated concentration

λ_{r-1} is the scaling parameter initially set to 1, then calculated from

iteration r-1:

$$\lambda_{r-1} = \frac{\dfrac{S_h\left(\underline{b}_{r-1}\right)}{nh}}{\dfrac{S_c\left(\underline{b}_{r-1}\right)}{nc}} \tag{9}$$

where:

$S_h\left(\underline{b}_r\right)$ is $\displaystyle\sum_{i=1}^{nh} w_{h_i}\left(h_i - \hat{h}_i\left(\underline{b}_r\right)\right)^2$,

w_{h_i} is the weight of the ith observation that is not receiving alternate weighting, i.e. those with stat-flags equal to or greater than 0 (in ground-water applications this is likely a hydraulic head),

h_i is the ith observed value not receiving alternate weighting, i.e. stat-flags equal to or greater than 0 (in ground-water applications this is likely a hydraulic head),

$\hat{h}_i\left(\underline{b}_r\right)$ is the ith simulated value, from iteration r, not receiving alternate weighting, i.e. stat-flags equal to or greater than 0 (in ground-water applications this is likely a hydraulic head),

nh   is the number of values not receiving alternate weighting, i.e. those with stat-flags equal to or greater than 0 (in ground-water applications this is likely the number of hydraulic heads),

$$S_c(\underline{b}_r) \text{ is } \sum_{j=1}^{nc} w_{c_j}\left(c_j - \hat{c}_j(\underline{b}_r)\right)^2,$$

w_{c_j} is the weight of the jth observation to receive alternate weighting, i.e. those with stat-flags equal to -1 (in ground-water applications this is likely a concentration),

c_j is the jth observed value to receive alternate weighting, i.e. stat-flags equal to -1 (in ground-water applications this is likely a concentration),

$\hat{c}_i(\underline{b}_r)$ is the jth simulated value, from iteration r, to receive alternate weighting, i.e. stat-flags equal to -1 (in ground-water applications this is likely to be a concentration), and

nc   is the number of values to receive alternate weighting, i.e. those with stat-flags equal to -1 (in ground-water applications this is likely the number of concentrations).

To prevent difficulty converging the weights are held constant at the last calculated weight once the parameter values change by an amount less than the "final threshold" which is calculated as:

$$\text{final threshold} = 0.1\,(THRESHOLD - TOL) + TOL$$
(10)

ALTERNATE PRINT FILES OPTION – If the user would like to have the extensive lists of observations, residuals, and sensitivities printed to alternate files so that it is easier to read the primary output files, the following flag is specified:

ALTERNATE_PRINT (not case-sensitive)

When this specification is made the alternate output for phases 1 through 3 is written to *fn._alt*, alternate output for phase 33 is written to *fn._3a*, for phase 44 is written to *fn._4a*, and for 45 is written to *fn._5a*. If it is desired that the intermediate information be printed to these alternate files, the intermediate printing option will need to be turned on. Otherwise the alternate files will hold this information for the initial and the final parameter values.

PRINT GRAPH FILE HEADER OPTION - If the user would like to have labels printed as the first line in the output files that are used for graphing, the following flag is specified:

HEADER_PRINT (not case-sensitive)

When this specification is made a header will appear as the first line of the following files: *fn._cp, fn._c4, fn._c5, fn._nm, fn._os, fn._r, fn._rd, fn._rg, fn._s1, fn._s2, fn._s3, fn._sp, fn._w, fn._ws, fn._ww.*

LINEAR EXTRACTION OPTION - If the user would like to extract each observation in its order of occurrence in the *filename.uni* file from an output file that includes only the simulated equivalents, one on each line, in the proper order, the following flag is specified:

LINEAR_EXRTACT (not case-sensitive)

In this case the extract file contains only the codes <*filename* and END, and may appear as follows:
> <head_flow.out
> <concentration.out
> END

Blank lines and comments should not occur in either the *filename.ext* file or the application output files that contain the values to be extracted.

GROUP STATISTICS REPORTING OPTION - If the user would like to calculate statistics by observation type then plot-flags need to define the groups of interest in the observation list of the *filename.uni* file and specify the following flag after the observation list and before the END flag in the *filename.uni* file.

GROUP_STATS (not case-sensitive)

In this case, composite scaled sensitivities will be printed for each group, for each iteration, and the final residual statistics (minimum, average, maximum, #positive, #negative, number of runs, and the sum of squared weighted residuals) will be printed for each group for the optimal values.

RESTART OPTION - If the user would like to restart the regression program at a previous iteration with different regression controls or omitting insensitive parameters without repeating the time-consuming executions of the forward model, the restart option can be selected by specifying:

RESTART_SAVE (not case-sensitive)

W hen this specification is made scratch files are written to a subdirectory within the directory where UCODE was executed and labeled with the iteration number. Each directory will contain the following files:

> *filename.uni*
> *u9_zscr.pre*
> *u9_zscr.2*
> *u9_zscr.3*
> *u9_zscr.5*
> *u9_zscr.6*
> *u9_zscr.8*
> *u9_zscr.12*
> *u9_zscr.13*
> *and if central differences were utilized:*
> *u9_zscr.1*
> *u9_zscr.11*

If the user decides to restart the regression from an intermediate stage they need to:

- copy all of these files from the selected iteration to the parent directory

- edit the *filename.uni* file to:

  - specify any desired alternative regression controls, such as varying MAX-CHANGE or NOPT etc.   NOTE: the observations cannot be changed at this time because only the extracted values are saved and these are not available for other observations
  - specify **RESTART_RUN** (not case-sensitive) in the position where **RESTART_SAVE** had been

- edit the *u9_zscr.pre* file by simply removing the asterisk in front of the parameter number that should be fixed at the value it had at that point in the regression.

The *u9_zscr.pre* file contains information such as:

```
5 6
F yes
<bcf.tpl
>bcf.sub
/!k1,,,,,,,,,,! *0  3.0e-5  3.0e-3  0.03  %14.7e  1  1
/!kv,,,,,,,,,,! *1  2.0e-7  1.0e-6  0.03  %14.7e  1  1
/!k2m,,,,,,,,,,! !2  4.0e-6  4.0e-4  0.03  %14.7e  1  0
<riv.tpl
>riv.sub
/!krb,,,,,,,,,,! *3  1.2e-4  1.2e-2  0.03  %14.7e  1  1
<rcharr.tpl
>rcharr.sub
/!rch1,,,,,,,,,,! *4  0.5e-8  4e-8  0.1  %15.7e  0  1
/!rch2,,,,,,,,,,! *5  0.5e-8  4e-8  0.1  %15.7e  0  1
END
```

Notice that the initial parameter values have been replaced with either a *
followed by the parameter number, if the parameter was being estimated
or, an ! followed by the parameter number, if it was not estimated. The
only option you have is to remove the * indicating that you would like to
proceed from that point in the regression with the parameter fixed at the
value at the start of the iteration. The parameter values are listed in
u9_zsrc.8 in order of occurrence in the *filename.pre* file with all of the
parameters to be estimated list first and the parameters that have a "do not
estimate flag second.

For example, the first 5 values in the following list are k1, kv, krb, rch1,
rch2, and the last value is k2m which is not estimated.

```
 0.4031901E-03
 0.1817285E-06
 0.1134773E-02
 0.1050016E-07
 0.1449825E-07
 0.4000000E-04
```

PARALLEL PROCESSING OPTION – To use the parallel option, **the companion Master-Tasker code is needed and must be operational on the master and slave computers in order to use the parallel option in UCODE.** If the parallel option is invoked, rather than indicating END at this location in the *filename.uni* file, a PARALLEL flag is input followed by the information required to undertake the parallel solution as delineated below. The Master-Tasker code manages a "master" computer and "slave" computers that have been setup to accept parallel executions of the application codes. All batch files that are to be executed on slave machines must include a final line as follows:

<div align="center">echo process_complete>%1</div>

PARALLEL (not case-sensitive) - Indicates the parallel option of UCODE will be invoked.

If "PARALLEL" is specified, each the following items must be specified in the following order, starting a new line for each item delineated in bold below:

PATH_to_MASTER - path to the directory where the subdirectories to hold the information for the parallel executions will be created, including a slash at the end. (e.g. D:\master\)

PATH_to_DYNAMIC_FILES - path to the directory where files that change throughout the parallel execution are stored the application codes, batch files and input files are held including a slash at the end. Everything in this directory will be copied to the slave machines. These files need to be all that is needed to conduct a forward model run when the proper command is given, except the input files that are created by UCODE substitution of values into template files.

PATH_to_STATIC FILES - path to the directory where files that will not change throughout the parallel execution are located. Everything in this directory will be copied to the slave machines only once at the beginning of the parallel operation. The setup of the application must reference these files with the correct relative directory references.

REPEAT_TIME - frequency with which to check "slave" machines where one forward execution of the application model(s) is performed to obtain information for calculation of sensitivities.

REPEAT_TIME_UNIT - time unit for REPEAT_time (s=second, m = minutes, h = hours)

DEFAULT_SPACE - minimum amount of space required to hold the application software, data files, and output

SPACE_UNIT - unit for default_SPACE (M = megabytes, G = gigabytes)

DEFAULT_SPEED - speed assumed for slave computer to perform one forward execution of the application model(s) in the default_time (see below)

SPEED_UNIT - unit for default_SPEED (Mhz = megahertz)

DEFALT_TIME - estimated time required to perform one forward execution of the application model(s) on a machine with the default speed

TIME_UNIT - time unit for default_TIME (s=seconds, m = minutes, h = hours)

DEFAULT_LAUNCH - command to launch one forward execution of the application model(s) when executed within the directory indicated as PATH_to_APPLICATION_FILES

TIME_OUT - time limit for waiting for SHORT TERM responses (e.g. acknowledgement that requests have been made or status has been updated) from the Master (secs) before terminating

   ABORT_CONDITION

-   condition to leave slave machines in after the program terminates unnaturally

   specify, either:

   clean

   leave

AFTER ALL OPTIONAL FEATURES HAVE BEEN SPECIFIED, the *fn.uni* file ends with:

END (not case-sensitive) - Indicates the end of the data input for this file.

ADDITIONAL REFERENCES FOR UCODE3.0 ON PAGE 61 & 62 OF THE ORIGINAL MANUAL

Anderman, E. R. and Hill, M. C., 1999, a New Multistage Groundwater Transport Inverse Method: Presentation, Evaluation, and Implications, *Water Res. Research*, 35(4), p. 1053-1063.

Barlebo, H. C., Hill, M. C., Rosbjerg, D., and Jensen, K.H., 1998, Concentration Data and Dimensionality in Groundwater Models: Evaluation Using Inverse Modeling, *Nordic Hydrology*, v. 29, p. 149-178.

.

.

Gailey, R. M., Gorelick, S. M., and Crowe, A. S., 1991, Coupled Process Parameter Estimation and Prediction Uncertainty Using Hydraulic Head and Concentration Data, *Adv. in Water Res.*, 14(5), p. 301-314.

.

.

Keidser, A. and Rosbjerg, D., 1991, A Comparison of Four Inverse Approaches to Groundwater Flow and Transport Parameter Identification, *Water Res. Research*, 27(9), p. 2219-2232.

.

.

Sonnenborg, T. O., Engesgaard, P., and Rosbjerg, D., 1996, Contaminant Transport at a Waste Residue Deposit, 1. Inverse Flow and Nonreactive Transport Modeling, *Water Res. Research*, 32(4) p. 925-938.

Van Rooy, D., Keidser, A., and Rosbjerg, D., 1989, Inverse Modelling of Flow and Transport, in Groundwater Contamination (Proceedings of the Symposium held during the Third IAHS Scientific Assembly, Baltimore, MD, May 1989), Linda Abriola ed., IAHS Publ. no. 185, p. 11-23.

Wagner, B. J. and Gorelick, S. M., 1987, Optimal Groundwater Quality Management Under Parameter Uncertainty, Water Res. Research, 23(7), p. 1162-1174.