

# **EFFECTIVE GROUNDWATER MODEL CALIBRATION: With Analysis Of Data, Sensitivities, Predictions, and Uncertainty**

By Mary C. Hill and Claire R. Tiedeman

## **Answers to Problems Posed in the Exercises**

This document contains the answers to the Problems posed in the exercises at the ends of Chapters 1 through 8. The questions posed in the Problems are not repeated here; refer to the book for these questions. Answers to Problems in the exercises of Chapter 9 are not included because these exercises are optional.

### **Exercise 3.2c: Check simulated values.**

No. The residuals in Table 3-3 suggest that there is no data input error. There are no residuals that are significantly larger in absolute magnitude than the others.

### **Exercise 3.2d: Calculate weights on hydraulic-head and flow observations**

For the head observations, first calculate the total variance of measurement error for the head observations, by adding the variance for the elevation measurement and the variance for the water-level measurement. This total variance is  $1.0025 \text{ m}^2$ , as shown in Table 3.2. Then, calculate the weight as the inverse of this variance, which equals  $0.9975 \text{ m}^{-2}$ . In the table 'DATA AT HEAD LOCATIONS' in the MODFLOW-2000 LIST output file, note that the square root of the weight ( $0.9988 \text{ m}^{-1}$ ) is printed.

For the flow observation, first calculate the standard deviation of measurement error by multiplying the coefficient of variation by the observed value:  $\sigma = 0.10 \times (-4.4 \text{ m}^3/\text{d}) = -0.44 \text{ m}^3/\text{d}$ . Then calculate the variance by squaring the standard deviation:  $\sigma^2 = 0.1936 \text{ m}^6/\text{d}^2$ . Finally, calculate the weight for the flow observation as the inverse of this variance, which equals  $5.165 \text{ d}^2/\text{m}^6$ . Again, note that the square root of the weight ( $2.27 \text{ d}/\text{m}^3$ ) is printed in the table 'DATA FOR FLOWS REPRESENTED USING THE RIVER PACKAGE' of the MODFLOW-2000 LIST output file.

### **Exercise 3.3: Evaluate model fit using starting parameter values**

*Initial model fit:*

The model fit with the starting parameter values shows that the simulated heads at almost all observation locations are larger than the observed values. This suggests bias in the fit, in that there is a systematic overprediction of the observed values. The weighted residual for the flow observation is small, indicating a good fit to the flow data.

*Comparison of residuals and weighted residuals:*

For the head observations, the residuals are essentially equal to the weighted residuals, because the square root of the weights for these observations is equal to about 1.0. For the flow observation, the weighted residual is larger than the residual, because the square root of the weight is 2.27.

**Exercise 4.1b: Use dimensionless and composite scaled sensitivities ( $dss$  and  $css$ ) to evaluate observations and defined parameters**

*Using  $dss$  to identify observations most important to estimation of parameter  $HK_1$ :*

The scaled sensitivities of Table 4.1 indicate that observations hd02.ss, hd03.ss, hd04.ss, hd05.ss, hd06.ss, hd08.ss, hd09.ss, and hd10.ss are much more important to estimation of  $HK_1$  than are observations hd01.ss, hd07.ss or flow.ss. Observation hd09.ss has the largest  $dss$  (in absolute value) and by this measure, provides the most information about parameter  $HK_1$ .

To explain the sensitivities using knowledge of the flow system dynamics, first consider how sensitivities are calculated. For perturbation sensitivities, parameter values are perturbed, the flow model is executed, and the change in dependent variable (here, hydraulic head or flow) is computed. The change in the dependent variable is then divided by the parameter perturbation. In MODFLOW-2000, sensitivities are computed by the sensitivity-equation method (Hill, 2000, p. 67-71), by which sensitivities are obtained by numerically solving the derivative of the flow equation. Regardless of which method is used to calculate the sensitivities, a good way to understand them is to think of the sensitivities as the change in hydraulic head or flow that occurs in response to a perturbation in the parameter value. This approach is used to answer the exercises related to sensitivities.

The sensitivity of the simulated river discharge to  $HK_1$  is very small, because the river is the only discharge location in the flow system, and the primary source of water to the system is specified areal recharge. Thus, regardless of the value of  $HK_1$ , simulated flow to the river will approximately equal the recharge. If recharge were the sole source of water to the flow system, then river discharge would exactly equal the areal recharge, and the sensitivity of the river discharge to  $HK_1$  would be zero. However, there is a small amount of water entering the aquifer through the head-dependent boundary with the adjoining hillside. This influx has a very small influence on the flow system, but can vary in response to changes in  $HK_1$ , which in turn causes the river discharge to have a non-zero sensitivity with respect to  $HK_1$ .

To understand the insensitivity of the simulated equivalent of observation hd01.ss to  $HK_1$ , consider its location (see Figure 2.1b). Observation hd01.ss is located in layer 1 directly below the river. All water that discharges to the river must flow through column 1 of layer 1. The river discharge is a function of  $K_{RB}$  and the difference between the river stage (which is constant) and the hydraulic head in layer 1, and approximately equals the specified areal recharge, as noted above. Thus, to maintain a similar river discharge in response to a change in  $HK_1$ , the hydraulic head in layer 1 beneath the river must remain approximately the same. The sensitivity of the simulated equivalent of observation hd01.ss to  $HK_1$  is not exactly equal to zero because the flow into the system through the hillside boundary varies slightly in response to changes in  $HK_1$ , as discussed above.

Observation hd07.ss is located in layer 2 directly below the river (see Figure 2.1b). The sensitivity of its simulated equivalent to  $HK_1$  is greater than that for observation 1 because all of the river discharge does not need to flow through layer 2. Nonetheless, the

head at the observation hd07.ss cell is strongly controlled by the overlying river, resulting in a smaller sensitivity to HK\_1 than at most other locations.

*Using css to assess likelihood of estimating all parameters:*

Parameter HK\_1 has the largest composite scaled sensitivity, and the *css* for parameters HK\_1, HK\_2, RCH\_1, and RCH\_2 are also relatively large. The *css* for K\_RB is more than two orders of magnitude smaller than that for HK\_1 and the *css* for VK\_CB is almost two orders of magnitude smaller than that for HK\_1, indicating that the regression might have difficulty estimating these two parameters.

#### **Exercise 4.1c. Evaluate parameter correlation coefficients (*pcc*) to assess parameter uniqueness.**

*Parameter pairs with pcc greater than 0.90 or 0.95:*

When the flow observation is included, the *pcc* is greater than 0.90 for parameter pairs HK\_1 & RCH\_1 (*pcc*=0.95) and HK\_2 & RCH\_2 (*pcc*=0.98).

*Likelihood of being able to independently estimate all parameters:*

The *pcc* value for pair HK\_2 – RCH\_2 suggests that it may be difficult to independently estimate these two parameters.

*Explanation of why all the parameters are extremely correlated when only hydraulic-head observations are included:*

Darcy's Law can be used to understand the extreme parameter correlation that results when only hydraulic-head observations are included. Equation 1.1 in Section 1.4.1 presents the one-dimensional form of Darcy's Law, and equation 1.2 shows Darcy's Law rearranged and solved for hydraulic head. Equation 1.2 clearly shows that the calculated head is a function of the ratio  $Q/K$ , and thus the same value of  $h$  can be computed from numerous different values of parameters  $Q$  (e.g. recharge) and  $K$ , as long as the ratio  $Q/K$  remains the same. In the weighted least squares objective function with weights set to constants, simulated values at observation locations are the only terms that are a function of the parameter values. Thus, when only hydraulic-head observations are included and all important parameters related to the flows and hydraulic conductivities of the system are being estimated by the regression, the same value of the objective function can be calculated from numerous combinations of the parameters. In other words, there is no set of unique parameter values that produce simulated values equal or close to the observed head values. This principle also applies to simulated hydraulic heads in more complex ground-water flow models, because the equations solved are based on Darcy's Law.

*Reason why the correlation coefficients calculated by UCODE\_2005 are unable to capture the extreme parameter correlation when using only hydraulic-head observations:*

The correlation coefficients calculated by the perturbation method are not as accurate as those produced by MODFLOW's sensitivity-equation method. This inaccuracy generally is most pronounced when the actual *pcc* are very close or equal to 1.0, as in this exercise. As discussed in Section 4.4.2, correlation coefficient accuracy tends to decrease as parameter sensitivity decreases, which explains why the *pcc* are least accurate for pairs involving parameter K\_RB.

**Exercise 4.1d. Use contour maps of one-percent sensitivities for the steady-state flow system.**

*Reasons why sensitivities for HK\_1 and HK\_2 are negative and why sensitivities for HK\_1 are larger (in absolute value) than those for HK\_2:*

Figure 4.4a shows that the one-percent scaled sensitivities of steady-state hydraulic heads in layer 1 to HK\_1 are near zero at the river, for reasons given in the answer to Exercise 4.1b. Away from the river, the sensitivity is negative, and increases in absolute value quickly, and then more slowly, with distance from the river. The negative values indicate that in response to an increase in HK\_1, heads will decline. The increase in absolute value with distance from the river indicates that the amount of decline increases with distance from the river, resulting in a flattening of the lateral hydraulic gradient through layer 1.

The underlying physics that result in the negative HK\_1 sensitivities can be explained by considering the flux through any cross section parallel to the river in layer 1. The flux  $Q_{A,lay1}$  through the area  $A_{lay1} = \Delta x \times \Delta y$  can be calculated using Darcy's Law as  $Q_{A,lay1} = -HK_1 \times A_{lay1} \times ((h_2 - h_1) / \Delta x)$ . All water in the system flows toward the river, so  $Q_{A,lay1}$  also equals the flux that has entered the system upgradient of  $A_{lay1}$  (from recharge and inflow at the hillside), minus the flux through layer 2. As noted in the answer to Exercise 4.1b, almost all flux in the system originates as constant recharge. Therefore, in response to an increase in HK\_1, the flux  $Q_{A,lay1}$  will change only if the flux through layer 2 changes. Assuming that the change in layer 2 flux is small,  $(h_2 - h_1) / \Delta x$  must decrease by the same amount that HK\_1 increases, for  $Q_{A,lay1}$  to remain constant. Because the head at the river is essentially invariant (see answer to Exercise 4.1b), this decrease in  $(h_2 - h_1) / \Delta x$  is achieved by decreasing  $h_1$  more than  $h_2$ . These decreases in heads in response to an increase in HK\_1 cause the negative sensitivities.

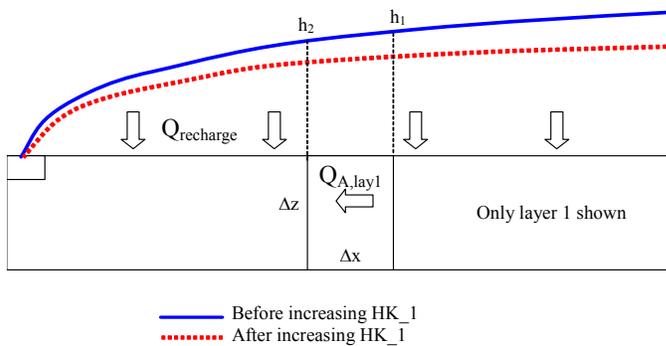


Figure 1. Schematic cross section through layer 1 of the steady-state model, showing the hydraulic gradient before and after increasing parameter HK\_1.

The one-percent sensitivities of hydraulic heads in layer 2 to HK\_1 are very similar to the sensitivities in layer 1 to HK\_1 (Figure 4.4a). This is because the vertical hydraulic conductivity of the confining bed controls flow between the two layers, and this parameter value stays the same when HK\_1 is increased. Therefore, heads in layer 2 must change by a similar amount as in layer 1, to produce the same fluxes between model layers and through layer 2.

The explanation for the negative one-percent sensitivities of hydraulic heads in layer 2 to HK\_2 (fig. 4.4b) is similar to that for the sensitivities of heads in layer 1 to HK\_1. Consider the flux through the area A within a cross section parallel to the river in layer 2. This flux is  $Q_{A,lay2} = -HK_2 \times A_{lay2} \times (\partial h / \partial x)$ , and equals the cumulative flux that has entered layer 2 through the confining layer upgradient of  $A_{lay2}$ , plus the influx from the hillside to layer 2, which is very small. If HK\_2 is increased, then the hydraulic gradient in layer 2 must decrease for  $Q_{A,lay2}$  to remain the same, assuming that the vertical fluxes do not change substantially. This decrease in hydraulic gradient is achieved by greater head declines with distance from the river, as explained above.

The one-percent scaled sensitivities for HK\_1 are larger than those for HK\_2 because the starting value of HK\_1 is larger than that of HK\_2. Again consider Darcy's Law for flux through cross sections in layers 1 and 2. In layer 1,  $Q_{A,lay1} = -HK_1 \times A_{lay1} \times (\partial h / \partial x)$ ; increasing HK\_1 by one-percent of its value means that  $\partial h / \partial x$  needs to decrease by this same amount to maintain similar horizontal flows through layer 1. In layer 2,  $Q_{A,lay2} = -HK_2 \times A_{lay2} \times (\partial h / \partial x)$ ; increasing HK\_2 by one-percent of its value means that  $\partial h / \partial x$  needs to decrease by this amount to maintain similar horizontal flows through layer 2. Because HK\_1 is greater than HK\_2, the required decrease in gradient in layer 1 is larger than that in layer 2.

*Reason why sensitivities for RCH\_1 and RCH\_2 are positive and why those for RCH\_1 only vary over the left half of the system, whereas those for RCH\_2 vary over the entire domain:*

To understand the one-percent sensitivities of hydraulic head in layers 1 and 2 to the recharge parameters (Figures 4.4e,f), note that an increase in recharge will cause a corresponding increase in the flux through the system toward the river. To convey this increased flux, the hydraulic gradient towards the river must become steeper. This requires hydraulic head to increase more and more with distance from the river, which explains the positive sensitivity values.

An increase in parameter RCH\_1, which applies to the left half of the aquifer, causes the hydraulic gradient to become steeper only over the left half of the aquifer. In the right half of the aquifer, the hydraulic gradient is unchanged because flow through the system has not increased. In contrast, although RCH\_2 applies only to the right half of the aquifer, an increase in this parameter affects fluxes throughout the aquifer, because the additional recharge must flow through the left half of the aquifer to reach the river.

*Explanation of VK\_CB sensitivities:*

The one-percent sensitivities of hydraulic head in layers 1 and 2 to VK\_CB (Figure 4.4d) exhibit very different patterns. In layer 1, the sensitivity is near zero at the river to maintain the unchanged discharge to the river. To the right of the river, the sensitivities are negative, and increase in absolute value with distance from the river, indicating that the lateral hydraulic gradient through layer 1 flattens in response to an increase in VK\_CB. This is because if VK\_CB is increased, more of the recharge will flow into layer 2 of the model, and, consequently, less will be conveyed to the river through layer 1. In layer 2, the one-percent sensitivity of heads to VK\_CB is negative at the river, and becomes smaller in absolute value with distance from the river. This pattern indicates that

the hydraulic gradient through layer 2 becomes steeper in response to an increase in  $VK\_CB$ , as needed to convey the increased flow through layer 2.

*Reason why sensitivities for  $K\_RB$  are constant throughout the system:*

The one-percent sensitivities of hydraulic heads to  $K\_RB$  (fig. 4.4c) are the same at all model nodes. Discharge to the river through riverbed area  $A_{RB}$  is  $Q_{RB} = -K\_RB \times A_{RB} \times ((h_{aq} - H_{riv}) / \Delta z)$ , where  $h_{aq}$  is the simulated head in the river cells, and  $H_{riv}$  is the river stage.  $Q_{RB}$  also equals the sum of specified areal recharge and the much smaller flux from the hillside. In response to an increase in  $K\_RB$ , simulated head in the river cells decreases so that  $(h_{aq} - H_{riv}) / \Delta z$  decreases and  $Q_{RB}$  remains approximately the same. All other fluxes within the aquifer must also remain the same because recharge is not changed. This is achieved by maintaining the same hydraulic gradients throughout the system, which in turn is achieved by changing head at all model nodes by the same amount as at the river nodes. This is depicted schematically in Figure 2.

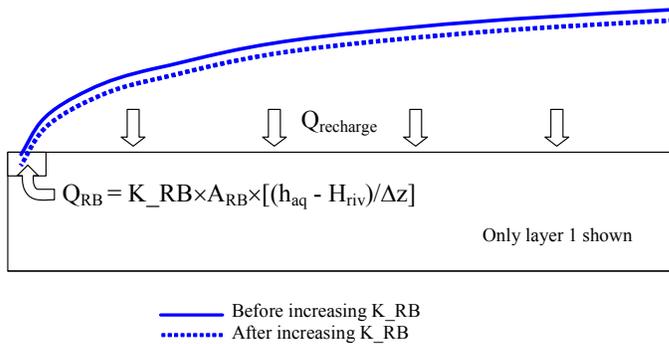


Figure 2. Schematic cross section through layer 1 of the steady-state model, showing the hydraulic gradient before and after increasing parameter  $K\_RB$ .

#### **Exercise 4.1e. Evaluate leverage statistics.**

*Parameters with largest dimensionless scaled sensitivities:*

For observations flow01.ss and hd01.ss, the  $dss$  are small for all parameters. For observation hd07.ss, the  $dss$  for parameter  $VK\_CB$  is large, but the  $dss$  for all other parameters are small. For observation hd09.ss, the  $dss$  for parameters  $HK\_1$ ,  $HK\_2$ ,  $RCH\_1$ , and  $RCH\_2$  are large.

*Evaluation of whether high leverage observations are dominated by sensitivity or correlation considerations:*

For flow01.ss and hd01.ss, the  $dss$  values clearly indicate that the high leverage values are not dominated by sensitivity considerations. These observations have high leverage values because they prevent large correlations. The role of flow01.ss in preventing large correlations is explained in the answer to Exercise 4.1c. The role of hd01.ss in preventing large correlations is illustrated in Table 1 below. Comparison of this table with Table 4.2 shows that when this observation is omitted, the absolute values of all correlations between all parameters except  $VK\_CB$  increase substantially.

Table 1. Parameter correlation coefficient (*pcc*) matrix calculated with 9 hydraulic-head observations (hd01.ss is omitted) and 1 flow observation.

|       | HK_1 | K_RB         | VK_CB | HK_2         | RCH_1        | RCH_2        |
|-------|------|--------------|-------|--------------|--------------|--------------|
| HK_1  | 1.00 | <b>-0.96</b> | 0.50  | <b>-0.97</b> | <b>0.98</b>  | -0.94        |
| K_RB  |      | 1.00         | -0.67 | 0.94         | -0.89        | 0.90         |
| VK_CB |      |              | 1.00  | -0.40        | 0.35         | -0.33        |
| HK_2  |      | symmetric    |       | 1.00         | <b>-0.96</b> | <b>0.99</b>  |
| RCH_1 |      |              |       |              | 1.00         | <b>-0.95</b> |
| RCH_2 |      |              |       |              |              | 1.00         |

Observation hd01.ss prevents large correlations because it constrains the value of simulated head in the river cells, which in turn constrains the value of parameter  $K_{RB}$ . The flux out the river is:  $Q_{RB} = -K_{RB} \times A_{RB} \times ((h_{aq} - H_{riv}) / \Delta z)$ , where  $h_{aq}$  is the simulated head in the river cells, and  $H_{riv}$  is the river stage (see answer to Exercise 4.1d). Without observation hd01.ss to constrain the value of  $h_{aq}$ , multiple combinations of  $h_{aq}$  and  $K_{RB}$  will produce the same value of  $Q_{RB}$ .

Observation hd07.ss is located beneath the river in layer 2 (Figure 2.1b), and has large leverage because it helps constrain fluxes in the cells beneath the river. The flux into column 1 of layer 1 is the sum of (A) the applied recharge in column 1, (B) the horizontal inflow from column 2, and (C) the vertical inflow from layer 2 (see Figure 4.5).

Observation hd07.ss partly constrains the value of component (C), because together with observation hd01.ss, it constrains the vertical gradient from layer 1 to layer 2. This constraint on component (C) in turn constrains component (B), because component (A) is fixed. If this observation were absent, there would be less constraint on how much flux is apportioned between components (2) and (3). Interestingly, observation hd07.ss alone does not appear to prevent large correlations; removing it from the calculation does not significantly increase any of the *pcc* values. This is discussed further in the answer to Exercise 7.1b.

Observation hd09.ss has large leverage because it has dimensionless scaled sensitivities for parameters HK\_1, HK\_2, RCH\_1, and RCH\_2 that are equal to or larger than those of any other observations (Table 4.1), owing to its distant location from the river.

### Exercise 5.1a. Assess relation of objective-function surfaces to parameter correlation coefficients

*Using Darcy's Law to explain complete correlation of parameters when only hydraulic-head observations are used:*

This explanation is given in the answer to Exercise 4.1c. Note also that for the two-parameter problem, the simulated heads are a function of the exact ratio of  $Rch_{Mult}/K_{mult}$  ( $Q/K$  in equation 1.2), as indicated by objective-function contours that have a slope of 1.0 (Fig. 5.4a). In general, for more complex groundwater models with several hydraulic conductivity and recharge parameters, simulated heads are a function of a linear combination of these parameters.

*Adding a single flow observation:*

Adding one flow observation prevents the problem from being completely correlated because in this case, the objective function depends on the simulated equivalents of both the head and the flow observations. Although simulated heads are still a function of the ratio  $RchMult/KMult$ , simulated discharge to the river is a function only of  $RchMult$ . Thus, if  $RchMult$  and  $KMult$  vary but their ratio remains constant, the simulated river discharge will vary and different objective-function values will be calculated. In fact, the value of  $RchMult$  completely determines this simulated value, because areal recharge is the major influx to the flow system, and the river is the sole discharge boundary. Thus, in the regression, the estimate of  $RchMult$  is completely dependent on the value of observed flow to the river, which means that any error in the flow measurement is directly propagated to the  $RchMult$  estimate. Because of this, several flow measurements should ideally be included in regressions of ground-water flow models.

*Objective function contours parallel to one axis:*

If the objective function contours were all parallel to one axis, the problem would be insensitivity. For example, contours parallel to the  $RCH\_MULT$  axis in Figure 5.4 would indicate that the objective function is dependent only on parameter  $RCH\_MULT$ , and completely independent of, and insensitive to, the value of  $K\_MULT$ .

### **Exercise 5.1b. Examine the performance of the modified Gauss-Newton method**

*Convergence and regression performance for the different runs:*

In Run 1, the starting parameter values are near those that minimize the true objective function surface, and the regression converges even though *max-allowed-change* is very large. This is because in the vicinity of these starting parameter values, the linearized objective function resembles the true objective function.

In Run 2, the starting parameter values are far from those that minimize the true objective function surface and *max-allowed-change* is very large. The parameters determined by the regression do not smoothly progress to the true minimum and convergence fails. Performing nonlinear regression to determine parameter values that minimize the true objective function involves a series of iterations in which the values are determined for a linearized form of the objective function. The values that minimize this function can be determined in a single iteration. Although this function resembles the true objective function surface in the vicinity of the parameter values around which it is linearized, further away it can deviate substantially from the true function. Thus, when the parameters around which the function is linearized are very different from those that minimize the true objective function and *max-allowed-change* is very large, the parameter values calculated by the regression might not progress closer to the minimum of the true objective function.

In Run 3, *max-allowed-change* is decreased to 0.5, so the allowed change in parameter values from one iteration to the next is a maximum of 50 percent. Even with starting values that are still very different from those that minimize the true objective function, the regression is well-behaved because the parameter changes are damped. At each iteration, the regression is allowed to proceed only part way toward the minimum of the linearized surface, and thus proceeds gradually toward the minimum of the true surface in a few iterations.

Finally, in Run 4, with small *max-allowed-change* and starting parameter values that are in an area of the objective function surface that slopes steeply toward the true minimum, the regression is very well-behaved and proceeds quickly to the true minimum.

*Estimated parameter values, uniqueness, and regression performance for the different observation types and weightings:*

For the first set of runs, with only hydraulic heads, the regression ‘converges’ to different sets of ‘optimal’ parameter estimates for different starting parameter values because there is no unique minimum in the objective-function surface (and thus no set of optimal parameter estimates). Instead, there is a trough in the surface, the bottom of which is defined by a line that represents the nonunique minima of the surface. An infinite number of combinations of RchMult and Kmult result in an objective function value on this line. Run 1 ‘converges’ quickly because the starting parameter values are close to the trough that defines the minima of the true surface in Figure 5.4a, even though *max-allowed-change* is large. In Run 2, the parameter values in iterations 2 through 10 are outside the boundaries of Figure 5.4a. With a very large value of *max-allowed-change*, the regression moves to the minimum of the objective function surface that is linearized about  $K\_Mult = 9.0$  and  $Rch\_Mult = 0.2$ . This minimum is extremely far from these starting parameter values, and the changed parameter values are highly unreasonable. Once the regression has moved into this region, the minima of subsequent objective function surfaces linearized about the unreasonable parameter values remain far away. In Run 3, the regression behavior is better because *max-allowed-change* is small. However, the ‘optimal’ parameter estimates are different from those in Run 1 because the parameter values progressed to a different part of the trough than in Run 1. In Run 4, the starting values are far from the true values but in an area of the objective function surface with a much steeper gradient than in Run 3, and the regression converges to parameter estimates in the part of the trough that is downgradient (in objective function space) from the starting values.

The result that the regression iterations converged to different parameter values in Run 3 than in Run 1 and Run 4 indicates that the regression cannot estimate a unique set of optimal parameters for this problem. This illustrates the importance of performing regression runs with different starting parameter values if high correlations are suspected, to test whether the parameter estimates are unique. This is especially important when using perturbation sensitivities, because, as discussed in Chapter 4.4.2, the less accurate perturbation sensitivities computed can produce inaccurate parameter correlation coefficients that do not reveal the actual high correlations. In this case, the only reliable way to detect problems with parameter correlation is to start the regression with different initial parameter values, and check that it converges back to the same optimal parameter estimates each time.

For the second set of runs, with hydraulic head observations and flow observations weighted with a coefficient of variation of 10 percent, the objective function has a unique minimum and Runs 1, 3, and 4 all converge to the same set of optimal parameter estimates. In Run 2, in which the starting parameter values are in a very flat area of the objective function surface, the regression behavior is similar to that for the case with only hydraulic-head observations. Even for parameter estimation problems that have a unique minimum and for which a small value of *max-allowed-change* is used, the regression can

be ill-behaved when the starting values are in a flat area very far from the optimal estimates.

For the third set of runs, with hydraulic head observations and flow observations weighted with a coefficient of variation of 1 percent, the regression behavior of all runs is essentially the same as for the second set of runs. In this problem, the objective function surface has a much better defined minimum than that for the problem with a flow observation of 10 percent (Figure 5-4), but away from this minimum, the surface is still very flat in the lower right corner. Thus, in Run 2, the parameter values are still well outside the bounds of the figure, and the regression does not converge.

### Exercise 5.1c. Derive the Gauss-Newton normal equations (optional)

The following derivation is modified from that given by Draper and Smith (1998, p. 135).

Substituting equation 5-11 into equation 3-2 produces the following:

$$S(\underline{b}) = [\underline{y} - \underline{y}'(\underline{b}_r) - \underline{X}_r(\underline{b} - \underline{b}_r)]^T \underline{\omega} [\underline{y} - \underline{y}'(\underline{b}_r) - \underline{X}_r(\underline{b} - \underline{b}_r)]$$

Let the expression  $\underline{y} - \underline{y}'(\underline{b}_r) = \underline{y}_r$ , for simplification of terminology throughout the derivation :

$$S(\underline{b}) = [\underline{y}_r - \underline{X}_r(\underline{b} - \underline{b}_r)]^T \underline{\omega} [\underline{y}_r - \underline{X}_r(\underline{b} - \underline{b}_r)]$$

To minimize  $S(\underline{b})$ , take its derivative with respect to  $\underline{b}$ , and set this derivative equal to 0:

$$\frac{\partial}{\partial \underline{b}} [(\underline{y}_r - \underline{X}_r(\underline{b} - \underline{b}_r))^T \underline{\omega} (\underline{y}_r - \underline{X}_r(\underline{b} - \underline{b}_r))] = 0$$

Expanding the expression in brackets, and evoking the matrix multiplication rule  $(AB)^T = B^T A^T$ :

$$\frac{\partial}{\partial \underline{b}} [(\underline{y}_r^T - (\underline{b} - \underline{b}_r)^T \underline{X}_r^T) (\underline{\omega} \underline{y}_r - \underline{\omega} \underline{X}_r(\underline{b} - \underline{b}_r))] = 0$$

And:

$$\frac{\partial}{\partial \underline{b}} [\underline{y}_r^T \underline{\omega} \underline{y}_r - \underline{y}_r^T \underline{\omega} \underline{X}_r(\underline{b} - \underline{b}_r) - (\underline{b} - \underline{b}_r)^T \underline{X}_r^T \underline{\omega} \underline{y}_r + (\underline{b} - \underline{b}_r)^T \underline{X}_r^T \underline{\omega} \underline{X}_r(\underline{b} - \underline{b}_r)] = 0$$

Because the term  $(\underline{b} - \underline{b}_r)^T \underline{X}_r^T \underline{\omega} \underline{y}_r$  is a scalar, it equals its transpose, which is  $\underline{y}_r^T \underline{\omega} \underline{X}_r(\underline{b} - \underline{b}_r)$ . Also,  $\underline{\omega} = \underline{\omega}^T$  because the weight matrix is symmetric. Thus, the middle 2 terms of the above equation are equal :

$$\frac{\partial}{\partial \underline{b}} [\underline{y}_r^T \underline{\omega} \underline{y}_r - 2(\underline{b} - \underline{b}_r)^T \underline{X}_r^T \underline{\omega} \underline{y}_r + (\underline{b} - \underline{b}_r)^T \underline{X}_r^T \underline{\omega} \underline{X}_r(\underline{b} - \underline{b}_r)] = 0$$

Furthermore,  $\frac{\partial}{\partial \underline{b}} [\underline{y}_r^T \underline{\omega} \underline{y}_r] = 0$  because  $\underline{y}_r$  and  $\underline{\omega}$  are not functions of  $\underline{b}$ , so the derivative reduces to:

$$\frac{\partial}{\partial \underline{b}} [-2(\underline{b} - \underline{b}_r)^T \underline{X}_r^T \underline{\omega} \underline{y}_r + (\underline{b} - \underline{b}_r)^T \underline{X}_r^T \underline{\omega} \underline{X}_r(\underline{b} - \underline{b}_r)] = 0$$

Because  $\frac{\partial}{\partial \underline{b}} \underline{b}_r = 0$ , this reduces to:

$$\frac{\partial}{\partial \underline{b}} [-2 \underline{b}^T \underline{X}_r^T \underline{\omega} \underline{y}_r + (\underline{b} - \underline{b}_r)^T \underline{X}_r^T \underline{\omega} \underline{X}_r (\underline{b} - \underline{b}_r)] = 0$$

Taking the derivative of each term produces:

$$-2 \underline{X}_r^T \underline{\omega} \underline{y}_r + 2 \underline{X}_r^T \underline{\omega} \underline{X}_r (\underline{b} - \underline{b}_r) = 0$$

This yields:

$$\underline{X}_r^T \underline{\omega} \underline{X}_r (\underline{b} - \underline{b}_r) = \underline{X}_r^T \underline{\omega} \underline{y}_r$$

Substituting the expressions  $\underline{y}_r = \underline{y} - \underline{y}'(\underline{b}_r)$  and  $\underline{d}_r = \underline{b} - \underline{b}_r$  produces equation 5-10 :

$$(\underline{X}_r^T \underline{\omega} \underline{X}_r) \underline{d}_r = \underline{X}_r^T \underline{\omega} (\underline{y} - \underline{y}'(\underline{b}_r)) \quad 5-10$$

### Exercise 5.2b. First attempts at estimating parameters by nonlinear regression.

*Examine the changing values of the parameters and max-calculated-change:*

Throughout the 10 iterations, the absolute value of *max-calculated-change* remains very large (Fig. 5.5) compared to the TOL convergence criteria of 0.01, instead of gradually declining toward TOL, as should occur in a well-posed problem. In addition, *max-calculated-change* oscillates between large positive values and large (absolute) negative values. These behaviors suggest that the regression is not well-posed, and that simply increasing the number of allowable iterations will not help the regression converge. The parameters associated with *max-calculated-change* for most of the iterations are HK\_2 and VK\_CB.

The changing parameter values in Table 5.4 show that the regression is trying to significantly decrease the values of these two parameters. These decreases are limited by the user-specified value of *max-allowed-change* equal to 2.0, which causes damping of parameter changes so that the maximum parameter change is  $\pm 200$  percent, and by MODFLOW-2000 prohibiting negative hydraulic conductivity values. (In MODFLOW-2000, if the damped parameter change results in a negative value of a parameter that cannot physically be negative, the parameter is assigned a value two orders of magnitude lower than its starting value. In UCODE\_2005, a similar effect can be obtained using the Constrain keyword in the Parameter\_Values input block.) Overall, the behavior of *max-calculated-change* and of the changing parameter values suggests that there may not be enough information provided by the observations to estimate HK\_2 and (or) VK\_CB. The sum of squared weighted residuals decreases each iteration (Fig. 5.5), but this does not necessarily indicate that the regression is behaving well. For diagnosing problems with the regression, the values of *max-calculated-change* and the parameters are generally more useful than are the sum of squared weighted residuals.

*Consider the composite scaled sensitivities:*

When calculated at the starting parameter values, the *css* for K\_RB and VK\_CB are less than two orders of magnitude smaller than that for HK\_1, which has the largest *css*. This

suggests that the regression may have difficulty estimating  $K_{RB}$  and  $VK_{CB}$ , as discussed in Exercise 4.1b. The  $css$  calculated after iteration 2 for parameter  $HK_2$  is much smaller, and for parameter  $K_{RB}$  is much larger, than the  $css$  calculated at the starting parameter values (Table 5.5). (The changing values of the  $css$  illustrate the model nonlinearity with respect to the hydraulic conductivity parameters; the values of  $HK_2$ ,  $VK_{CB}$ , and  $K_{RB}$  calculated by the regression after iteration 2 are much smaller than their starting values and consequently the sensitivities differ.) The changed  $css$  for  $HK_2$  and  $K_{RB}$  are the reason that  $HK_2$  instead of  $K_{RB}$  is associated with the largest value of *max-calculated-change* for several of the iterations. However, the  $css$  calculated at the more reasonable starting values should be used to guide decisions about how to improve the behavior of the regression. Thus, attention should be focused on parameters  $VK_{CB}$  and  $K_{RB}$ , rather than on  $HK_2$ .

### **Exercise 5.2c. Assign prior information on parameters.**

*Compare the regression performance with the results of exercise 5.2b:*

The regression performance of exercise 5.2c has improved significantly compared to that of exercise 5.2b. Table 5.6 shows that the absolute values of *max-calculated-change* decrease throughout the regression iterations, reaching the convergence criterion in five iterations. *Max-calculated-change* is less than *max-allowed-change* in all iterations. Note that the damping parameter is less than 1.0 in iteration 3. This occurs because *max-calculated-change* oscillates from a negative value in iteration 2 to a positive value in iteration 3; see Appendix B for details about the calculation of the damping parameter in this situation. In this regression run, the parameter values are not changed by orders of magnitude, as occurred in exercise 5.2b, but rather change gradually and smoothly approach the optimal values. The sum of squared weighted residuals at the estimated parameter values is much smaller than in the final iteration of the run for exercise 5.2b.

*Reason that the two parameters with prior information have estimates that are nearly identical to the respective prior value:*

The estimated values of the parameters with prior information ( $VK_{CB}$  and  $K_{RB}$ ) are nearly identical to the prior values because the observations provide very little information towards estimating these two parameters. In other words, the simulated equivalents of the observations are relatively insensitive to these parameters. There are no changed values of these parameters that significantly improve the fit to the observations and reduce the terms of the objective function related to the observations. Thus, the regression estimates values of these parameters that minimize the term of the objective function related to the prior information, which occurs when the parameter values are very close to the prior values. The parameter estimates are not exactly equal to the prior values because the observation data provide a small amount of information about the parameters – the values of observation sensitivities for  $VK_{CB}$  and  $K_{RB}$  are not exactly zero.

*Conclusion about whether the prior information is actually regularization:*

The prior information for this problem should be regarded as regularization. A coefficient of variation of 0.3 is used to calculate the weight for each parameter, which assumes that the limits of the 95-percent linear individual confidence interval on each prior value are

plus and minus about 60 percent of the value (two standard deviations, where the standard deviation is 30 percent of the prior value). Because of issues such as lack of data, heterogeneity, and differences between the scale of a measurement and the scale of a hydraulic conductivity zone in a model, prior hydraulic conductivity estimates are generally considered known only to within plus or minus an order of magnitude (e.g., plus or minus 100 percent).

### **Exercise 5.2d: Parameter estimates and objective-function values**

The estimated values differ from the true values because the observations include measurement error, as described in Exercise 3.2d. The regression produces the best possible fit to these error-laden observations, not to the true values of hydraulic head and flow. When the estimated parameter values are used to calculate the fit to the true values without error, the fit is worse.

### **Exercise 6.1a: Examine objective-function values**

*Verification of the maximum-likelihood objective function value:*

Equation 3.3 for the maximum-likelihood objective function is:

$$S'(b) = (ND+NPR) \ln 2\pi - \ln |\underline{w}| + \underline{e}^T \underline{w} \underline{e} \quad 3.3$$

ND is the number of observations, which equals 11, and NPR is the number of prior information equations, which equals 2.

To calculate the  $\ln |\underline{w}|$  term, note that the determinant of a diagonal matrix is the product of the diagonal elements. The weight for each of the head observations is 0.9975 and the weight for the flow observation is 5.165 (we omit the units of the weights here, because the natural log, which is dimensionless, appears in the maximum-likelihood objective function).

The weight for the prior on K\_RB is:  $1.0/(0.3 \times (1.2 \times 10^{-3})^2) = 7.716 \times 10^6$ .

The weight for the prior on VK\_CB is:  $1.0/(0.3 \times (1.0 \times 10^{-7})^2) = 1.111 \times 10^{15}$ .

Thus,  $\ln |\underline{w}| = \ln (0.9975^{10} \times 5.165 \times [7.716 \times 10^6] \times [1.111 \times 10^{15}]) = 52.12$ .

The  $\underline{e}^T \underline{w} \underline{e}$  term is the least squares objective function value, which equals 10.56.

Substituting these values into equation 3-3:

$$S'(b) = (13)\ln 2\pi - 52.12 + 10.56 = -17.67.$$

The value of the maximum likelihood objective function can be found near the bottom of the MODFLOW-2000 GLOBAL output file and the UCODE\_2005 main output file.

*Explanation for why the objective-function values may not be the best indicators of model fit:*

The objective function values do not account for the drawbacks of adding additional defined parameters to the model. Objective function values will decrease (or possibly remain the same) as more parameters are added. However, adding parameters has the drawback of decreasing the confidence in the estimated parameter values, because the information provided by the calibration data is spread over more parameter values.

**Exercise 6.1b: Demonstrate the circumstance in which the expected value of both the calculated error variance and the standard error is 1.0 (optional)**

Step (a) can be accomplished using many programs. For example, with Microsoft Excel uniformly distributed random numbers with a mean of zero can be generated using the function “rand(a,b)-(b-a)/2”, where a and b are the lower and upper limits of the uniform distribution, respectively. The variance of the distribution equals (b-a)<sup>2</sup>/12. Normally distributed random numbers can be generated using the function “norminv(rand(),0,σ)”, where σ is the standard deviation.

The calculation in steps (2) through (4) can be expressed as:

$$s^2 = \frac{\sum_{i=1}^n \left[ \frac{a_i^2}{\sigma^2} \right]}{n}$$

where  $a_i$  are the random numbers generated in step (1) and  $\sigma^2$  is the variance of the distribution.

Generating 10, 100, 1,000, and 10,000 random numbers from a normal distribution with a mean of 0.0 and a variance of 1.0 produced the following values of  $s^2$ .

| n     | $s^2$ | $\frac{n}{\chi_U^2}$ | $\frac{n}{\chi_L^2}$ |
|-------|-------|----------------------|----------------------|
| 10    | 1.717 | 0.49                 | 3.08                 |
| 100   | 1.086 | 0.77                 | 1.35                 |
| 1000  | 0.945 | 0.92                 | 1.09                 |
| 10000 | 0.982 | 0.97                 | 1.03                 |

Different values result from different sets of generated numbers, but two characteristics are consistent: (1)  $s^2$  tends to be close to 1.0, and (2)  $s^2$  tends to approach 1.0 as n increases. The expected variation in  $s^2$  as n increases can be calculated using equation 6.2. For the population variance of 1.0 expected for the weighted residuals, the upper and lower 95-percent confidence interval limits on  $s^2$  are shown in the third and fourth columns of the table above.

For step (6) of this exercise, the calculation can be expressed as:

$$s^2 = \frac{\sum_{i=1}^n \left[ \frac{a_i^2}{\sigma_a^2} \right] + \sum_{i=1}^n \left[ \frac{b_i^2}{\sigma_b^2} \right]}{2n}$$

where  $a_i$  and  $b_i$  are the random numbers in the two sets,  $\sigma_a^2$  and  $\sigma_b^2$  are the variances of the two distributions, and n is the number of generated values in each set. Generating 10, 100, 1,000, and 10,000 random numbers from normal distributions with a mean of 0.0 and variances of 1.0 and 4.0 produced the following values of  $s^2$ .

| n of samples from each distribution | $s^2$ |
|-------------------------------------|-------|
| 10                                  | 1.611 |
| 100                                 | 1.011 |
| 1000                                | 0.979 |
| 10000                               | 0.988 |

**Exercise 6.1c: Evaluate calculated error variance, standard error, and fitted error statistics**

*Comparing  $s^2$  to the expected value of 1.0:*

The calculated error variance  $s^2$  for the regression run of exercise 5.2c is 1.51. This value is fairly close to 1.0, but that is a subjective determination. The confidence interval on this value provides a more objective determination. Applying equation 6.2 using the results from the regression run of Exercise 5.2c,  $n = 11+2-6 = 7$ , and the chi-square values are given in the exercise. The calculated 95-percent confidence interval is 0.66; 6.21. This interval contains the value 1.0, indicating that if the weighted residuals are random, then the value 1.51 does not significantly deviate from 1.0, and the model fit is consistent with the statistics used to calculate the weights. The randomness of the weighted residuals is assessed in Exercise 6.2.

*Calculation of the fitted standard deviation for heads:*

The standard deviation of measurement error for the head observations is  $1.0025^{1/2} \text{ m} = 1.0013 \text{ m}$ , so the fitted standard deviation for heads equals  $1.0013 \text{ m} \times 1.23 = 1.23 \text{ m}$ . This value is very small compared to the total head loss over the system (between the hillside and the river) of about 75 m, indicating that the model provides an excellent overall fit to the head observations.

**Exercise 6.1d: Examine the AIC, AIC<sub>c</sub>, and BIC statistics**

*Verify the values of AIC and BIC, and calculate AIC<sub>c</sub>:*

From equation 6.3b, the AIC statistic is  $AIC(\underline{b}') = S'(\underline{b}') + NP \times 2$ . From Exercise 6.1a,  $S'(\underline{b}') = -17.67$ , and  $NP=6$ . Thus, the value of the AIC statistic equals  $-17.67 + 12 = -5.67$ .

From equation 6.4, the BIC statistic is  $BIC(\underline{b}') = S'(\underline{b}') + NP \times \ln(ND+NPR)$   
 $= -17.67 + 6 \times \ln(11+2) = -2.28$ .

From equation 6.3b, the AIC<sub>c</sub> statistic is:

$$AIC(\underline{b}') = S'(\underline{b}') + NP \times 2 + \frac{(2 \times NP \times (NP + 1))}{(NOBS + MPR - NP - 1)}$$

$$= -17.67 + (6 \times 2) + (2 \times 6 \times (6 + 1))/(11 + 2 - 6 - 1) = 8.33.$$

MODFLOW-2000 calculates the AIC and BIC statistics, whereas UCODE\_2005 calculates AIC, AIC<sub>c</sub>, and BIC.

For this model, AIC<sub>c</sub> should be used instead of AIC, because  $NOBS/NP = 11/6 = 1.8$  is less than 40.

*Improvement in model fit needed to reduce values of AIC, AIC<sub>c</sub>, and BIC:*

To determine the required improvement in model fit, first write the statistics with the maximum likelihood function ( $S'(\underline{b}')$ ) decomposed into its components:

$$AIC = (ND+NPR) \ln 2\pi - \ln |\underline{\omega}| + S(b) + NP \times 2$$

$$AIC_c = (ND+NPR) \ln 2\pi - \ln |\underline{\omega}| + S(b) + NP \times 2 + \frac{(2 \times NP \times (NP + 1))}{(NOBS + MPR - NP - 1)}$$

$$BIC = (ND+NPR) \ln 2\pi - \ln |\underline{\omega}| + S(b) + NP \times \ln(ND+NPR)$$

Then, substitute values for all terms except  $NP$  and the least squares objective function (here the notation  $S(b)$  is used for the least squares objective function, replacing the  $\underline{e}^T \underline{\omega} \underline{e}$  terminology that is used in equation 3.3). See the answer to exercise 6.1a for the calculation of terms of  $S'(\underline{b}')$ .

$$AIC = -28.3 + S(b) + 2NP$$

$$AIC_c = -28.3 + S(b) + 2NP + (2NP \times (NP+1))/(12-NP)$$

$$BIC = -28.23 + S(b) + 2.56NP$$

For the AIC statistic to decrease when parameters are added,  $S(b)$  must decrease by an amount greater than 2.0 times the number of parameters added.

The necessary decrease in  $S(b)$  for the  $AIC_c$  statistic to decrease when parameters are added is a nonlinear function of the number of parameters added. Substituting  $NP=1, 2,$  and  $3$  into the above equation for  $AIC_c$  indicates that  $S(b)$  must decrease by more than 2.4, 5.2, and 8.7 for  $AIC_c$  to decrease when the number of added parameters is 1, 2, and 3, respectively.

For the BIC statistic to decrease when parameters are added,  $S(b)$  must decrease by an amount greater than 2.56 times the number of parameters added.

**Exercise 6.2a. Graph of weighted residuals versus weighted simulated values and the minimum, maximum, and average weighted residuals.**

*Randomness of weighted residuals:*

The weighted residuals appear evenly distributed about zero. The weighted residuals for the flow observations and prior information are clustered near a weighted residual value of zero, which is curious, but is not an indication of nonrandomness. This behavior is explained in the answer for Exercise 6.2e, below.

The weighted residuals for the flow and prior also are located in a different position on the horizontal axis from the weighted head residuals. This occurs because of the different weighted simulated values for the different data types, and is not problematic.

Note that assessing randomness in this plot is made more difficult by the small sample size. Small samples that are actually drawn from random distributions can appear non-random simply because of the small number of data points.

*Maximum, minimum, and average weighted residuals:*

The minimum weighted residual is -2.05, the maximum weighted residual is 1.59, and the average weighted residual is 0.114. The minimum and maximum are not significantly larger in absolute value than other weighted residuals, indicating that the fit to these observations is not significantly worse than for other observations. If some of the smallest or largest weighted residuals are significantly larger in absolute value than others, this information can help to quickly identify parts of the model that might suffer from major construction problems, or observations that have been misinterpreted. The average weighted residual is small, but this is not necessarily an indication that the overall fit is good. In nonlinear regression, the average residual will usually be close to zero for the optimal parameter values.

**Exercise 6.2b. Graphs of observations versus simulated values. Examine the correlation coefficient  $R$ .**

*Utility of the three different graphs shown in Figure 6-7:*

Figure 6.7a is a more useful graph for diagnosing problems with the regression than is Figure 6.7b or 6.7c. In Figures 6.7b and c, the wide range in the magnitudes of weighted or unweighted observed and simulated values obscures the details of problems with model fit.

*Evaluation of  $R$ :*

The value of  $R$  indicates an excellent match between the trends in the weighted simulated and weighted observed values. However,  $R$  is not a useful diagnostic statistic. Because of the wide range in magnitudes noted above, the plotted values will commonly lie close to a 45-degree straight line on graphs like those in Figures 6.7b and c, and the value of  $R$  will be very close to 1.0, even if there are problems with the model fit.

**Exercise 6.2c. Graphs of weighted residuals against independent variables. Evaluate runs statistic.**

*Evaluation of spatial randomness of weighted residuals:*

Because of the small number of hydraulic-head residuals for this problem, it is very difficult to assess whether the weighted residuals are randomly distributed on the maps of the model domain. In this situation, it is more useful to assess whether any of the largest (in absolute value) weighted residuals are indicative of a significant problem with the model construction or with the calibration observations.

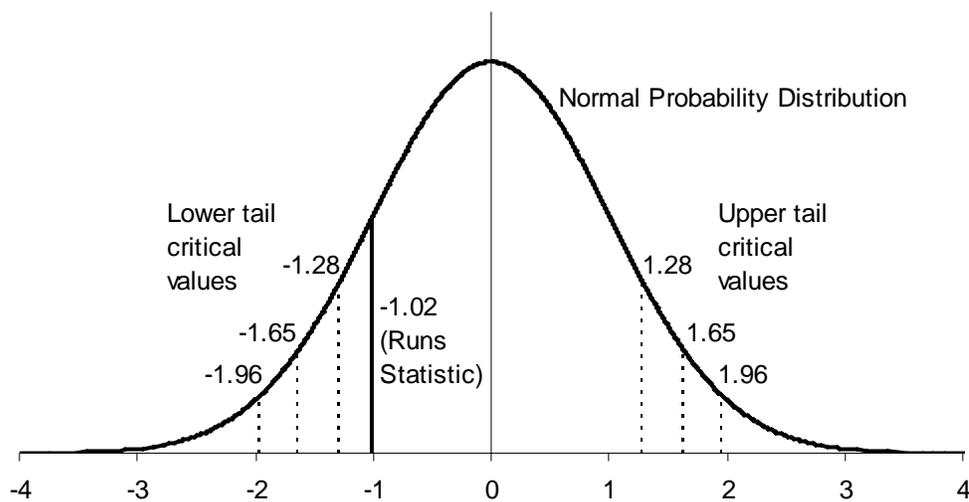
*Physical reasons for the three large weighted residuals in model layer 1:*

Two of the largest weighted residuals in model layer 1 are in column 4. Recall that in this flow system, simulated hydraulic heads are the same in any model column. Thus, it is impossible for the regression to closely match both of these observed values, which differ because they were generated by adding random noise to the head values calculated for the true system (see Exercise 3.2d). The result that the weighted residual for hd02.ss in row 4, column 4 is positive and that for hd04.ss in row 13, column 4 is negative indicates that the regression estimated parameters that produce a simulated value in between the two observed values. The sum of the squared weighted residuals (terms of the objective function) for these two head observations will be smallest when the simulated value in

column 4 is roughly in between the two observed values than if the simulated value matches one observed value very well and the other very poorly.

The third large weighted residual in layer 1 is associated with observation hd01.ss. This observation has small dimensionless scaled sensitivities for all model parameters, as shown in Table 4.1 and explained in the answer to Exercise 4.1b, and the simulated equivalent is largely controlled by the value of river stage. The observed value differs from the river stage because of the noise used to generate this value. Thus, the residual for this observation is essentially fixed at a value equal to the observed head minus the river stage, and it is impossible for the regression to estimate parameter values that would reduce this residual.

*Value of runs statistic and critical values graphed on a normal probability distribution:*



*Conclusions about randomness of weighted residuals with respect to their order:*

Because the negative runs statistic is to the right of the lower tail critical values, we conclude that we can not reject the hypothesis that that the residuals are random with respect to the order in which they are listed in the observation data files. In effect, this is an indication that the residuals are random with respect to this order; however, because the runs statistic is evaluated in the context of hypothesis testing, it is necessary to state the conclusion as not rejecting the hypothesis, rather than accepting the hypothesis.

**Exercise 6.2d. Evaluate normal probability graphs and the correlation coefficient  $R_N^2$ .**

*Evaluation of the normality of weighted residuals:*

The residuals on the normal probability graph in Figure 6.11 do not lie in a straight line, because of the cluster of weighted residuals near zero that form a kink in the plot. These are the weighted residuals for the flow observation and the prior information. This full set of weighted residuals does not appear to be normally distributed. Based on this normal probability plot alone, it is not known whether the non-normality reflects correlation in the residuals caused by the fitting of the regression or other problems that might indicate

an inadequate model. The resolution to this issue is explained below in the answer to the Exercise 6.2e.

$R_N^2$ , the correlation between ordered weighted residuals and normal order statistics, is 0.941 for the weighted head and flow residuals ( $ND=11$ ) and is 0.926 for the weighted head, flow, and prior residuals ( $ND+NPR=13$ ). Table D.3 provides critical values of  $R_N^2$  above which the hypothesis cannot be rejected that the weighted residuals are independent and normally distributed. However, this table does not include critical values for the case in which  $ND$  or  $ND+NPR$  is less than 35, as indicated in the output file shown in Figure 6.12. With  $ND$  or  $ND+NPR$  equal to 35, the critical value of  $R_N^2$  at a significance level of 0.05 is 0.943; the critical values decrease as  $ND$  or  $ND+NPR$  decreases.

It is likely that 0.941 is larger than the critical value for  $ND=11$ , and that the hypothesis cannot be rejected that the head and flow weighted residuals are independent and normally distributed.

The fit to the prior information is different, statistically, than the fit to the heads and flows, because the weighted residuals for both prior information values are very small. The smaller  $R_N^2$  value of 0.926 for the set of residuals that includes the prior information indicates that the addition of these two weighted residuals produces a sample that is less normal and (or) independent. Because it is not known whether 0.926 would be larger than the critical value for  $ND+NPR=13$ , we cannot make a conclusion about the independence and normality of the full set of weighted residuals using the  $R_N^2$  statistic.

*Comparison of generated sets of residuals to critical values:*

For the ten sets of 13 random numbers we generated, the  $R_N^2$  values are 0.986, 0.937, 0.990, 0.983, 0.979, 0.990, 0.964, 0.954, 0.944, 0.977. These values can be obtained by using a spreadsheet to explicitly calculate the terms of equation 6.18, or by using a built-in spreadsheet function for calculating the correlation between two ordered sets of numbers. Using the critical value for a set of 35 numbers at 5-percent significance level (0.943), it is expected that  $R_N^2$  will be less than the critical value 5 percent of the time, or for 0.5 sets. For our generated sets,  $R_N^2$  is less than the critical value for one of the ten sets, which is roughly consistent with what is expected.

**Exercise 6.2e. Determine acceptable deviations from random, independent, and normal weighted residuals.**

*Comparison of weighted residuals with generated independent and correlated random numbers:*

The distribution of weighted residuals versus weighted simulated values (Figure 6.7a) is much more similar to the graphs of correlated random numbers (Figure 6.13b) than to the graphs of independent random numbers (Figure 6.13a). This is because in Figure 6.7a and

in each of the graphs of Figure 6.13b, the values for flows and the prior are very close to zero.

Similarly, the trends in the normal probability graph of weighted residuals (Figure 6.11) match those of the generated correlated random numbers (Figure 6.14b) more closely than those of the independent random numbers (Figure 6.14a). In the plots of correlated random numbers, there is a distinct kink at a value of about zero, which is consistent with the kink in the plot of weighted residuals. This kink occurs primarily because the weighted residuals for the flow observation and the prior estimates are close to zero.

*Reasons for the apparent lack of randomness and normality:*

In this simple flow model, we can explain why the weighted flow and prior residuals are each close to zero. The flow is the only observation preventing complete correlation, so it is matched closely in the regression. The prior estimates are on two parameters for which the observations provide little information, and so the regression matches the prior estimates closely. Additional explanation of this is given in the answer to Exercise 5.2c. Thus, certain characteristics of the flow system influence why these three weighted residuals are clustered near zero. These characteristics are represented in the sensitivities from which the expected correlations of weighted residuals are calculated, and, therefore, the plots of correlated random numbers resemble the weighted residuals more closely than plots of independent random numbers.

The conclusions from this exercise indicate that the non-normal appearance of the residuals in Figure 6.11 can be explained by the fitting of the regression. Therefore, it is not necessary to search for problems with the model construction that might be causing this behavior of the weighted residuals.

### **Exercise 7.1a. Evaluate composite scaled sensitivities.**

*Differences between the initial and final css values:*

For nonlinear models, sensitivities of simulated values with respect to parameter values will differ when calculated at different parameter values. Therefore, the sensitivity components ( $\partial y'_k / \partial b_j$ ) of the *css* calculation are different at the final parameter estimates compared to their values at the initial parameter values. The *css* calculation also involves scaling by the parameter values  $b_j$ , and the final parameter estimates are different from the initial parameter values.

*Effect of nonlinearity and scaling:*

If the model is too nonlinear, and sensitivities are vastly different when calculated at different parameter values, then the *css* will not be very useful for guiding which parameters to estimate and which to specify. In this case, gradient methods might not be successful for performing parameter estimation and global optimization methods, discussed briefly in Section 5.2, might need to be used. If parameter values change significantly over the course of a regression run, then the scaling by  $b_j$  can cause large differences in the *css* calculated at initial and final parameter values. This can potentially cause the *css* to underestimate or overestimate the information that the observations provide about a particular parameter. The modeler needs to consider this possibility when choosing parameters to include or omit from the regression.

*Considering attempting regression without prior information:*

The *css* in Figure 7.5b do not suggest that regression should be attempted without prior on K\_RB or VK\_CB. The final *css* values for these two parameters are still very small compared to those for the other four parameters.

*Explanation for small flow and prior weighted residuals:*

The small *css* for K\_RB and VK\_CB show that the head and flow observations provide little information about these two parameters. In the regression, there is therefore no reason for the estimates of K\_RB and VK\_CB to be different from their prior values. This is explained in more detail in the answer to Exercise 5.2c.

### **Exercise 7.1b. Evaluate leverage statistics**

Comparison of the leverage statistics shown in Exercise 4.1e with those in Table 7.2 indicates that the statistic for hd01.ss has drastically changed, from 0.99 to 0.00. This change is caused by the addition of prior information on parameter K\_RB. With no prior information, observation hd01.ss is important to constraining the value of simulated head in the river cells, which in turn constrains the value of parameter K\_RB (see answer to Exercise 4.1e), and prevents high correlations between several parameters. With prior information specified on K\_RB, observation hd01.ss is no longer important to constraining its value and preventing correlations, and so it has no leverage.

Comparison of the leverage statistics shown in Exercise 4.1e with those in Table 7.2 also indicates that the statistics for hd07.ss and hd09.ss remain large.

Observation hd07.ss has high leverage because it is important to preventing high correlations among parameters HK\_2, RCH\_1, and RCH\_2, as shown by comparing Tables 7.4 and 7.6. The correlations calculated without observation hd07.ss for parameter pairs HK\_2 & RCH\_1, HK\_2 & RCH\_2, and RCH\_1 & RCH\_2 are all greater in absolute value than 0.95, indicating that the regression might have difficulty converging to unique parameter estimates if observation hd07.ss were omitted. Its omission also causes correlations for pairs HK\_1 & RCH\_1 and HK\_1 & RCH\_2 to significantly increase in absolute value.

Observation hd07.ss is located beneath the river in layer 2 (Figure 2.1b), and is important to reducing parameter correlations because of its role in constraining fluxes into the river cells in column 1, as discussed in the answer to Exercise 4.1d. Note that in the calculations for Exercise 4.1d, omitting observation hd07.ss did not produce significant increases in any *pcc*. This is probably because observation hd01.ss was more important than hd07.ss to constraining the fluxes beneath the river, as indicated by the large increases in *pcc* calculated with its omission. In this exercise, hd01.ss no longer plays an important role in constraining these fluxes, because of the prior information on K\_RB. Therefore, hd07.ss plays a larger role, particularly in constraining the vertical flux from layer 2 to layer 1 beneath the river.

As discussed in the answer to Exercise 4.1e, observation hd09.ss has large leverage because it has dimensionless scaled sensitivities for parameters HK\_1, HK\_2, RCH\_1, and RCH\_2 that are equal to or larger than those of any other observations (Table 7.5), owing to its distant location from the river. As discussed in the answer to Exercise 4.1c,

the flow observation has high leverage because it is the only observation that prevents complete correlation of the model parameters.

Finally, the prior information, which was absent in the calculations for Exercise 4.1e, has very high leverage because it constrains the values of the insensitive parameters  $K_{RB}$  and  $VK_{CB}$ .

### **Exercise 7.1c. Evaluate the importance using influence statistics**

*Explanation of influential observations:*

Observations  $hd07.ss$ ,  $hd09.ss$ ,  $flow.ss$ ,  $K_{RB}$  prior and  $VK_{CB}$  prior each have a Cook's D value that is larger than the critical value of 0.308. All of these observations except for  $hd09.ss$  have DFBETAS values for at least four of the parameters that are larger than the critical value of 0.555.

For this model the observations with large influence are identical to those with large leverage, and the discussion in the answer to Exercise 7.1b that explains the high leverage values in the context of sensitivities and preventing high correlation also explains the high influence statistics.

This exercise and Exercise 7.1b illustrate that using the dimensionless and composite scaled sensitivities together with parameter correlations does not always reveal the observations most important to estimating the parameters. Observations with very small  $dss$  can have large values of leverage, DFBETAS and Cook's D if they are important to reducing parameter correlations.

### **Exercise 7.1d. Evaluate the uniqueness of the parameter estimates using correlation coefficients**

*Most highly correlated parameter pairs:*

Parameter pairs  $HK_2$  &  $RCH_2$  ( $pcc = 0.91$ ) and  $HK_2$  &  $RCH_1$  ( $pcc = -0.85$ ) are most highly correlated. Hydraulic conductivity and recharge parameters typically have the largest correlations in ground-water problems, because of their relation in Darcy's Law (see answer to Exercise 4.1c). However, these high correlations of 0.91 and -0.85 are not so large to indicate potential problems with nonuniqueness of the estimated parameters.

*Differences between correlations calculated at starting and final parameter values:*

The parameter correlations calculated at the final parameter values are different from those calculated at the starting values. This is expected, because the correlations are a function of the sensitivities, which for nonlinear models are different when calculated for different parameter values.

*Differences in parameter correlation coefficients calculated using perturbation and sensitivity-equation sensitivities:*

The two sets of parameter correlation coefficients are nearly identical. Differences result from the less accurate sensitivities calculated by UCODE\_2005, and tend to be largest for coefficients related to insensitive parameters. The greatest differences occur for

parameter K\_RB, which is the parameter with the smallest sensitivities, as shown in Figure 7.5b (note, however, that these differences are quite small).

### **Exercise 7.1e. Detecting non-unique parameter estimates**

Part (1): Perform regression with flow observation omitted

*Regression run results:*

This regression run is ill-behaved. Throughout the 10 iterations, the value of *max-calculated-change* does not gradually approach the TOL value of 0.01.

*Reason for high correlations:*

The explanation for the correlations from this run is given in the answer to Exercise 4.1c.

Part (2): Test model nonuniqueness

Starting the regression with the parameter values in sets 1 and 2 results in the same parameter estimates as were obtained when using the original starting parameter values. This is consistent with the conclusion from Exercise 7.1d that the highest parameter correlation (0.92) indicates parameter uniqueness is not likely to be problematic.

The strength of this approach is that it uses regression runs to more definitively test the conclusions drawn from the parameter correlation coefficients. One weakness is that often it is not possible to test convergence for a wide range of different starting parameter values because as these values diverge from the original starting values, the regression becomes less well-posed, and may not converge.

### **Exercise 7.1f. Evaluate the precision of the estimates using standard deviations, linear confidence intervals, and coefficients of variation**

*Largest confidence intervals and coefficients of variation:*

Parameter HK\_2 has the largest confidence interval; its size in terms of percent of estimated value is significantly larger than that for any of the other parameters. This parameter also has the largest coefficient of variation. The order of the relative sizes of the linear confidence intervals will always match the order of the relative sizes of the coefficients of variation. This is because the linear confidence interval is a linear function of the coefficient of variation. The confidence interval equals the estimated parameter value plus and minus a statistic from the student *t*-distribution times the standard deviation of the parameter, and the standard deviation equals the coefficient of variation times the parameter estimate.

*Relative uncertainty among the parameters:*

The relative sizes of the intervals show that HK\_1 is the most precisely estimated parameter, and HK\_2 is the least precisely estimated. For the remaining four parameters, RCH\_1 and RCH\_2 are estimated with slightly more precision than are K\_RB and VK\_CB.

*Percent of confidence intervals that include their true values:*

83 percent (five out of six) of the confidence intervals include their true values. This percentage is significantly smaller than 95 percent. The confidence interval for VK\_CB

does not include its true value. Prior information was imposed for this parameter, and as discussed in the answer to Exercise 5.2c, the weighting used indicates that this prior should be considered regularization because this weighting reflects a smaller degree of uncertainty than is realistic. This causes the confidence interval calculated for VK\_CB to underestimate the actual degree of uncertainty, and thus decreases the likelihood that the interval contains the true parameter value.

**Exercise 7.1g. Compare estimated parameter values with reasonable ranges.**

For parameters HK\_1, K\_RB, VK\_CB, RCH\_1, and RCH\_2, the estimate and a large portion of the confidence interval lie within the reasonable range, indicating that the parameter estimates are reasonable. For parameter HK\_2, the parameter estimate is near the lower limit of the reasonable range, there is a very high degree of uncertainty, and a large portion of the confidence interval lies outside the range. This result is similar to that described in Section 7.6 for parameter C. With this degree of uncertainty, it is inconclusive whether the parameter estimate is reasonable. Additional data about model features related to the HK\_2 parameter and (or) additional observations are needed to resolve this question. The situation for HK\_2 illustrates the importance of including confidence intervals in the analysis of parameter reasonableness. If the estimate of HK\_2 were considered without its confidence interval, we would probably conclude that the estimate is reasonable, and would neglect to pursue important additional information that could help produce a more certain parameter estimate.

**Exercise 7.1h. Evaluate the precision of the estimates using nonlinear confidence intervals**

*Comparison of individual linear and nonlinear intervals:*

The linear 95-percent confidence intervals are always symmetric about the parameter estimate (for all parameters, the estimated value is shown as a value of 100 on the vertical axis in Figure 7.7), because they are calculated as the estimated value plus and minus a statistic times the parameter standard deviation. The nonlinear interval for HK\_1 is approximately symmetric about the estimate, but for all other parameters, the nonlinear intervals are asymmetric about the parameter estimate. Nonlinear intervals can be asymmetric because the limits are obtained through simulations, rather than by an arithmetic expression that guarantees symmetry. The linear and nonlinear intervals are roughly equal in size for all parameters except HK\_2, for which the linear interval is larger.

*Percent of confidence intervals that include their true values:*

All of the nonlinear 95-percent confidence intervals contain the true parameter value. In contrast, one individual linear interval (for VK\_CB) does not contain its true value. This result is consistent with the fact that these intervals are more accurate than the linear intervals.

*Relative uncertainty among the parameters:*

The relative sizes of the nonlinear intervals are the same as for the linear intervals – HK\_2 is the least certain parameter and HK\_1 is the most certain parameter.

**Exercise 7.2. Consider all the different correlation coefficients presented**

The equations for the three correlation coefficients are as follows:

$$R = \frac{\sum_{i=1}^{ND} (\omega_i^{1/2} y_i - m_y) (\omega_i^{1/2} y'_i - m_{y'})}{\left[ \sum_{i=1}^{ND} (\omega_i^{1/2} y_i - m_y)^2 \right]^{1/2} \times \left[ \sum_{i=1}^{ND} (\omega_i^{1/2} y'_i - m_{y'})^2 \right]^{1/2}} \quad (6.11a)$$

$$R_N^2 = \frac{[(\underline{e}_0 - \underline{m})^T \tau]^2}{[(\underline{e}_0 - \underline{m})^T (\underline{e}_0 - \underline{m})] (\tau^T \tau)} \quad (6.18)$$

$$pcc(j, k) = \frac{cov(j, k)}{var(j)^{1/2} var(k)^{1/2}} \quad (7.5)$$

The equations are similar in that the numerator is a covariance (or a covariance squared for  $R_N^2$ ) and the denominator is the product of the standard deviations (variances for  $R_N^2$ ) of the two variables involved in the covariance. They are also similar because absolute values of 1.0 result when the variables involved are coordinated with each other in some way. The coordination may be that the variables are equal, but other types of coordination produce the same result. For example, equation 6.11a can be used to show that R values approach 1.0 as simulated values approach being expressed as linear functions of the observed values. For example, if each simulated value equals its associated observed value plus 10, R equals 1.0.

The equations are different in the variables involved and whether coordination between the variables is advantageous to model calibration.

**Exercise 7.3a: Test for linearity using the modified Beale's measure**

*Analysis of modified Beale's measure:*

No. In contrast, the modified Beale's measure indicates that the model is highly nonlinear, so the linear confidence intervals are not accurate.

*Use of nonlinear instead of linear intervals in Exercise 7.1g:*

Yes. In Exercise 7.1g it was concluded that it is inconclusive whether parameter HK\_2 is reasonable, because although its estimate lies in the reasonable range of values, its linear confidence interval extends well outside of the reasonable range. In contrast, the nonlinear interval for HK\_2 shown in Figure 7.8 more closely coincides with the reasonable range of values, and it can be concluded that the parameter estimate is reasonable.

*Change in the modified Beale's measure when the prior weights are changed:*

The modified Beale's measure significantly increases when the weights on the two prior values are decreased (the coefficient of variations are increased). This larger value of the measure more realistically reflects the model nonlinearity.

### **Exercise 7.3b: Test for linearity using total and intrinsic model nonlinearity**

*Total model nonlinearity:*

No. The total model nonlinearity statistic is orders of magnitude larger than the value of 0.09 below which it can be concluded that the model is effectively linear. This result is consistent with the analysis of the modified Beale's measure, by which the model was also considered highly nonlinear.

*Intrinsic model nonlinearity:*

No. The model has a relatively small degree of intrinsic model nonlinearity. The value of 0.142 is just slightly larger than the value of 0.09 below which it can be concluded that the model is intrinsically linear.

*Effect of decreasing the weights on the prior information:*

The total model nonlinearity increases substantially when the prior weights are decreased, whereas the intrinsic model nonlinearity stays about the same. The measures calculated with the decreased weights are more realistic than those calculated with the weighting that is considered regularization.

### **Exercise 8.1a: Predict advective transport**

The advective path from the landfill goes to the supply well in the lower aquifer, and takes  $0.457 \times 10^{10}$  seconds, or 145 years, to reach the well.

### **Exercise 8.1b: Determine the parameters that are important to the predictions using prediction scaled sensitivities and parameter correlation coefficients**

*Answer to Question 2:*

The *pss* in Figure 8.8 show that overall, parameters RCH\_2 and POR\_1&2 are most important to all components of advective transport at all times. The *css* in this figure show that the head and flow observations provide substantial information about RCH\_2, but provide no information about POR\_1&2, because this parameter is not applicable to the flow model calibration. The conclusion from analyses of the *pss* and *css* is that additional information needs to be collected about POR\_1&2. However, this parameter has a relatively large prior weight that reflects independent field data and indicates the parameter has small uncertainty. The weighting does not factor into the *pss* and *css* analyses, but does factor into the *ppr* calculations, as discussed in the answer to Exercise 8.1c.

The *pcc* in Tables 8.4 and 8.5 indicate that the estimates of all parameters except K\_RB might be nonunique when the head and flow data alone are considered, but that the advective transport predictions require unique estimates of all parameters. This indicates that the prior information is essential for producing unique parameter estimates (compare Tables 7.6 and 8.4).

*Pss close to zero for K\_RB, VK\_CB, and RCH\_1:*

The *pss* for K\_RB are nearly zero because the value of this hydraulic conductivity parameter is relatively large compared to the other hydraulic conductivity parameters in the flow system, and thus the sensitivity of simulated values to K\_RB is small. The *pss* for RCH\_1 are close to zero because this recharge parameter applies to, and only affects fluxes in, the left half of the aquifer, whereas the advective transport path lies almost exclusively in the right half of the aquifer. The *pss* for VK\_CB are small because the advective transport path lies mostly in layer 1. The particle enters the confining unit at a time of 100 years, which is why the sensitivities at this time are larger.

*Pss for POR\_CB:*

The *pss* for POR\_CB are zero for all predictions at 10 and 50 years because the particle has not entered the confining unit at these times. The *pss* is non-zero only for prediction A100z, because the particle enters the confining unit at 100 years, and parameter POR\_CB affects the particle transport time in the vertical direction only.

*Effective porosity parameter to include in additional analyses:*

The *pss* suggest that POR\_CB has a very small effect on advective transport, and that it is reasonable to omit this parameter from further analyses of prediction sensitivity and uncertainty. In contrast, the advective transport predictions are highly sensitive to POR\_1&2, and this parameter should be included in the additional analyses.

### **Exercise 8.1c: Determine the parameters that are important to the predictions using the parameter-prediction statistic**

*Comparison of pss and ppr results:*

There are significant differences between the *pss* results in Figure 8.8 and the *ppr* results in Figure 8.9a,b. The *ppr* results show that (a) parameters VK\_CB and RCH\_1 are very important to the predictions, whereas the *pss* results show that the predictions are insensitive to these two parameters. In addition, the *ppr* results show that POR\_1&2 is not important to the predictions, but the *pss* show that the predictions are very sensitive to this parameter.

According to the *ppr* results, all parameters except K\_RB and POR\_1&2 would be beneficial to further investigate for improving the predictions.

*Explanation of different rankings:*

The different rankings by the *ppr* and *pss* statistics can be explained by considering that the *ppr* statistic is a function of parameter uncertainty and correlation as well as of prediction sensitivity. The absolute values of the parameter correlation coefficients (*pcc*) for the calibrated model with prior information omitted (Table 8.4) are at least 0.97 for parameter pairs composed of two parameters from the set VK\_CB, HK\_2, RCH\_1, and RCH\_2 and are 0.90 for pairs involving HK\_1 and one of these parameters. In contrast, the absolute value of all *pcc* involving K\_RB are no greater than 0.40. Also, because POR\_1&2 is not applicable in the flow model calibration, its correlation with all other parameters is 0.0.

The large  $pcc$  for parameters in the set VK\_CB, HK\_2, RCH\_1, and RCH\_2 strongly affect the  $ppr$  statistics. When potential new information is added on any one parameter in the set, the uncertainty of that parameter is reduced, and the uncertainty of all other parameters in the set also is substantially reduced. In contrast, adding the same level of new information on K\_RB does not substantially reduce the uncertainty of any other parameters, and adding new information on POR\_1&2 has no effect on the uncertainty of other parameters. This is the reason that parameters VK\_CB, HK\_2, RCH\_1, and RCH\_2 have relatively large values of the PPR statistic, despite the insensitivity of the predictions to some of these parameters.

The reason that prediction A100z has a large  $pss$  for POR\_1&2 but a small  $ppr$  value is related to parameter uncertainty. Although the flow model observations provide no information about porosity, a large prior weight is used in the  $ppr$  calculations, which causes the standard deviation of POR\_1&2 to be small. Decreasing this standard deviation by 10 percent produces a small absolute decrease, which translates to a small decrease in the uncertainty of A100z.

*Standard deviation decreases:*

Figure 8.9c shows that the absolute decreases in prediction standard deviations are smaller for prediction A100z than for predictions A100x and A100y. However, the decreases for A100x, A100y, and A100z are similar percentages of their respective transport distances of 5695 m, 3864 m, and 53.8 m. These results suggest that future data collection efforts can be guided by the  $ppr$  results.

*Parameter pairs most beneficial to simultaneously investigate:*

Figure 8.9d shows that it would be most beneficial to collect additional data on any parameter pair that does not include K\_RB or POR\_1&2.

### **Exercise 8.1d: Determine the importance of existing observations using the observation-prediction statistic**

*Identification of observations that rank as most important to predictions:*

Observations hd01.ss and flow01.ss rank as most important to the predictions by the  $opr$  statistic.

*Explanation for observation rankings by  $opr$  statistic:*

The  $dss$  in Table 7.5 and the  $pss$  in Figure 8.8 do not explain the importance of observations hd01.ss and flow01.ss. The  $dss$  indicate that the simulated values corresponding to these two observations are relatively insensitive to all the model parameters. The  $pss$  show that the predictions in at least one direction at 100 years are sensitive to parameters HK\_1, HK\_2, and RCH\_1, but because of the small  $dss$  values for hd01.ss and flow01.ss, this does not help to explain the  $opr$  results.

Evaluating the effect of omitting observations on the parameter correlation coefficients ( $pcc$ ) does help explain the  $opr$  results. The largest  $opr$  values are associated with the flow observation. When the flow observation is omitted, all  $pcc$  equal 1.0, because with only head observations, all model parameters are perfectly correlated (see answer to Exercises 4.1c). Thus, removing this observation causes the maximum possible increases

in the  $pcc$ , as summarized in Table 8.6. This table also explains the large  $opr$  statistic for observation  $hd01.ss$ ; its omission also causes large increases in  $pcc$ . The flow system dynamics that cause  $hd01.ss$  to strongly affect parameter correlations is explained in the answers to Exercises 4.1e and 7.2b.

*Evaluation of standard deviation increases:*

Figure 8.10b shows that when either  $hd01.ss$  or  $flow01.ss$  is omitted, the increases in the standard deviations for prediction A100z are substantially smaller than those for predictions A100x and A100y. However, as was the case for Exercise 8.1c, the uncertainty increases for A100x, A100y, and A100z are each very large compared to their respective transport distances of 5695 m, 3864 m, and 53.8 m. Thus, the absolute changes in prediction uncertainty do not alter conclusions about observation importance made using the  $opr$  statistic results.

**Exercise 8.1e: Assess the likely importance of potential new observations to the predictions using dimensionless and composite scaled sensitivities and parameter correlation coefficients**

Figure 8.11 shows that the potential head observation collected under pumping conditions provides substantial information about  $HK\_1$ ,  $HK\_2$ ,  $RCH\_1$ , and  $RCH\_2$ , relative to the information that the existing observations provide. In addition, comparison of Tables 8.4 and 8.7b shows that the presence of this observation reduces extremely high parameter correlations associated with parameter  $VK\_CB$ . This is because it would be collected under pumping stresses, and thus contributes information about vertical flows that is not provided by any of the existing observations. The potential flow observation provides only a small amount of new information relative to the existing observations. In addition, comparison of Tables 8.4 and 8.7a shows that the potential flow observation alone provides very little information towards reducing parameter correlations.

The advective transport predictions are most sensitive to  $RCH\_2$  and  $POR\_1\&2$  and are moderately sensitive to  $HK\_1$  and  $HK\_2$ . Thus, according to this analysis using  $dss$ ,  $css$ ,  $pss$ , and  $pcc$ , the improved information about  $HK\_1$ ,  $HK\_2$ , and  $RCH\_2$  that the potential new head observation provides would warrant collecting this additional data.

Note that whereas the existing flow observation is essential for preventing complete correlation of all parameters, the potential new flow observation plays essentially no role in reducing correlations. This is because even though it is collected under a different flow regime, its role is somewhat redundant – complete correlation has already been prevented by the presence of the existing flow observation.

**Exercise 8.1f: Assess the likely importance of potential new observations to the predictions using the observation-prediction ( $opr$ ) statistic**

The  $opr$  statistics in Figure 8.12 show that collection of the potential head observation is likely to contribute information important to most of the advective transport predictions, but that the collection of the potential flow observation is not likely to contribute important information. This conclusion is consistent with that drawn in Exercise 8.1e.

The potential head observation has larger values of the  $opr$  statistic than does the flow observation primarily because of its role in reducing parameter correlations and because

it is sensitive to parameters to which the predictions are sensitive, as discussed in the answer to Exercise 8.1e. The potential new flow observation is relatively unimportant because it provides little sensitivity information and does not substantially reduce any correlations.

*Best locations within model domain for collecting new head data:*

The *opr* statistics in Figure 8.13 show that the best location for collecting new head data is in the roughly circular area around and to the right of the well.

*Using reductions in parameter correlation coefficients to help explain results:*

As shown by comparing Figures 8.13 and 8.14, the area with the largest *opr* statistics is roughly coincident with the area in which a potential new head observation would most reduce any *pcc* value that is greater than 0.90 in the calculation without potential new observations. This shows that in this model, the role of potential new head observations in reducing high parameter correlations is the main factor that produces large *opr* statistics. New head observations under pumping conditions most reduce the correlations for parameter VK\_CB, because they provide information about vertical flows, and thus about the vertical hydraulic conductivity of the confining bed; this information is not provided by any of the existing observations in the model without pumping.

### **Exercise 8.2a: Calculate linear confidence intervals on the components of advective transport.**

*Why linear simultaneous intervals are larger than the linear individual intervals:*

As discussed in Section 8.4.1, simultaneous intervals account for uncertainty in more than one predicted quantity, whereas individual intervals account for uncertainty in only one quantity. Because of this, simultaneous intervals are always at least as large as individual intervals.

*Which type of interval is preferred representation of uncertainty:*

The simultaneous intervals are preferred, because they simultaneously account for uncertainty in all nine of the advective transport predictions.

*Answer to Question 5:*

The linear simultaneous intervals in Figure 8.15b show that in the x and y directions, the predicted advective transport path is moderately uncertain at 50 years, and is highly uncertain at 100 years. The intervals in the z direction shown in Figure 8.16 are moderately uncertain at 10 years and highly uncertain at 50 and 100 years.

At 50 years, the confidence interval in the x direction is about twice the length of the particle transport distance, and in the y direction is about four times the transport distance. However, the total area covered by the confidence intervals in the x and y directions is relatively small and is confined to the upper right portion of the aquifer. In the z direction, the confidence interval extends over about 75 percent of the aquifer thickness.

At 100 years, the area covered by the confidence intervals in the x and y directions extends over the entire model domain, and the interval in the z direction extends over the entire model thickness. Note that the captions for Figures 8.15b and 8.16 indicate that

these intervals also extend outside of the model domain; this is because the calculation of linear intervals does not take into account the physics of the flow system, and thus can include unrealistic values.

**Exercise 8.2b: Calculate nonlinear confidence intervals on the components of advective transport.**

The nonlinear simultaneous intervals in Figures 8.15d and 8.16 show that in the x and y directions, predicted advective transport is highly uncertain at 50 and 100 years, and in the z direction is highly uncertain at all times. In the x and y directions, the uncertainty represented by the nonlinear intervals is greater than that of the linear intervals for transport at 10 and 50 years. In the z direction, the nonlinear intervals are larger at 10 years, about the same at 50 years, and substantially smaller at 100 years.

The nonlinear intervals are preferred over the linear intervals because they are more accurate. They are calculated using model simulations, so that they include only realistic values. The particle projection method used in the ADV package can produce unrealistic values, but the modeler can filter out these unrealistic results when constructing the intervals, as is done in Figure 8.15d using dashed lines. The more accurate nature of the intervals is clearly shown by the result for the y and z components of transport at 100 years. In the simulations used to calculate the intervals, the effect of the pumping well prevents particles from migrating in the y direction into the lower half (in plan view) of the aquifer, whereas the linear intervals cannot account for this constraint (compare Figures 8.15b and d). The simulations also keep the vertical location of the particle within the top and bottom elevations of the aquifer, whereas the linear intervals extend well outside of these elevations (Figure 8.16).