Appendix E. Analysis of Calibration Target Errors

Groundwater models are calibrated by adjusting model parameters until the simulations match observed heads and fluxes. Although an ideal model would match all observations exactly, each observation (or calibration target) has associated errors. Consequently, a calibrated model should simulate values that, on average, are centered on the calibration target value and within the range of the associated errors of the calibration target (Anderson and Woessner, 2002).

This appendix assesses the ranges of errors associated with the calibration targets, examining the greatest degree of agreement (i.e., smallest errors) that might be expected for model-simulated versus observed values. Although the regional model also includes the uppermost aquifers, the local model has greater detail and resolution for these aquifers than does the regional model. Thus, the error analyses focus on shallow targets for the local model and deep targets for the regional model.

E.1. Errors Associated with Local Head Targets

Anderson and Woessner (2002) noted that calibration targets for head in a groundwater model have several sources of error, including unmodeled temporal and spatial variability, measurement errors, etc. Under the assumption that these errors are independent, the total error variance associated with a calibration target for head is the sum of the variances of the independent errors (Larsen and Marx, 1986):

$$\sigma_H^2 = \sigma_t^2 + \sigma_m^2 + \sigma_s^2 + \sigma_K^2 + \sigma_i^2 + \sigma_n^2 + \sigma_c^2$$

where

σ^2_H = the error variance associated with a calibration target for head
σ^2_t = error variance for unmodeled temporal variability
σ^2_m = variance of measurement errors
σ^2_s = variance of scale-up errors due to vertical averaging over long piezometer intervals
σ^2_K = variance of scale-up errors arising from unmodeled heterogeneity
σ^2_i = variance of interpolation errors
σ^2_n = variance of numerical errors within the solution convergence tolerance
$\sigma^2 c$ = error variance attributable to the effects of salinity.

In this study, the error variances generally are estimated using field observations. For the case of measurement and interpolation errors, only an estimate of the maximum absolute error is available. In these cases, this study will assume the errors to be normally distributed, and estimate the error variance as:

$$\sigma^2 = \left[\frac{\text{maximum absolute error}}{3}\right]^2$$

Some studies have assumed that the maximum absolute error represents two standard errors, i.e., a 95 percent confidence interval (Hill and Tiedeman, 2007). In this study, the sets of observations are generally large (100’s), which suggests that the observed maximum errors represent a higher level of confidence. This analysis therefore assumes that the range of the observations corresponds to a 99.7 percent confidence interval (equivalent to plus or minus three times the standard normal error).

Hill and Tiedeman (2007) note that calibration targets should be temporally consistent with the model simulations, that is, calibration targets should use an averaging interval for observations that is similar to that of the simulation period. Unfortunately, the wells in the local domain have insufficient records to infer a long-term average value head that would be consistent with a steady-state simulation. The errors associated with the unmodeled temporal variability are inferred from the head data collected by Locke and Meyer (2005), which included an initial measurement at each well during the inventory phase followed by the final synoptic measurement. The initial measurements were scattered over the 16 months preceding the synoptic measurement, so the differences between the initial and synoptic readings are a sample of the temporal variability of the shallow aquifers in the local model. This study uses the mean squared difference between the initial and synoptic observations to estimate the error variance due to temporal variability, yielding a value of $\sigma_t^2 = 95\ \mathrm{ft}^2$. Mandle and Kontis (1992) noted that although water levels had declined a few tens of feet near pumping centers, there is no general long-term trend in the water levels of the shallow sand and gravel aquifers.

For measurement errors, Locke and Meyer (2005) found the maximum absolute error to be 2.4 ft for steel tape measurements. With substitution, the corresponding error variance is:

$$\sigma_m^2 = \left[\frac{2.4}{3}\right]^2 = 0.64\ \mathrm{ft}^2$$

The scale-up error associated with vertical averaging is assumed to be negligible relative to other sources of error. This is justified by noting that the head observations of Locke and Meyer (2005) are grouped into hydrostratigraphic units that are explicitly represented by layers in the local model, limiting the impact of vertical averaging on calibration targets for head within the local model.

The scale-up error due to unmodeled heterogeneity was identified by Gelhar (1993), who suggested that a model using a homogeneous hydraulic conductivity will underestimate the actual variability of head. For an isotropic, two-dimensional hydraulic conductivity field with a multivariate lognormal distribution, Gelhar (1993) found that the unmodeled heterogeneity of hydraulic conductivity results in head variability whose variance is given by

$$\sigma_K^2 = 0.46\, \sigma_{\ln K}^2\ J^2\, a_{\ln K}^2$$

where

$a_{\ln K}$ = range of correlation for the natural log transform of K, the hydraulic conductivity

$\sigma_{\ln K}^2$ = variance of lnK

J = gradient within Kane County, measured from maps of Locke and Meyer (2005), approximately 0.003 ft/ft.

Pumping tests in Kane County indicate $\sigma^2_{\log K}$ = 0.13 (equivalent to $\sigma^2_{\ln K}$ = 0.69) within the major sand and gravel aquifers that are represented explicitly as zones within the local model. Gelhar (1993) found that the range of correlation can be estimated as 1/10 the extent of the modeled domain, which for the local model is approximately 30 miles. This makes $a_{\ln K}$ = 30 x 5280/10 = 1.58x10^4 ft. With substitution, the variance of scale-up errors in head due to unmodeled heterogeneity is

$$\sigma^2_K = 0.46(0.69)(0.003^2)(1.58\times10^4)^2 = 7.2\times10^2\,\text{ft}^2$$

The interpolation error is the maximum difference in head from the model-simulated head at the center of the block to the corners of a grid block. That is, the maximum interpolation error will be the gradient times $d_{\max}$, the distance from the center of the node to the corner. Assuming this error is normally distributed, the interpolation error variance is:

$$\sigma^2_i = \left(\frac{J\ x\ d_{\max}}{3}\right)^2$$

where

J = gradient (approximately 0.003 ft/ft) within Kane County, measured from the maps of Locke and Meyer (2005)

d_{max} = distance from the block center to the corner in the block

The local model has a grid spacing of 660 ft, so d_{max} = $\left[330^2 + 330^2\right]^{\frac{1}{2}} = 467\ \text{ft}$. With substitution, the error variance due to interpolation is:

$$\sigma^2_i = \left[\frac{(0.003)(467)}{3}\right]^2 = 0.22\,\text{ft}^2$$

The remaining potential sources for error are believed to be small. The density of shallow groundwater changes very little, and so the errors due to unmodeled density effects are negligible. The errors due to numerical approximation should be on the order of the convergence tolerance for the numerical solution (0.01 ft or less), and are likewise negligible.

The total error variance associated with a head calibration target in the local model is found by summing the contributions of the independent errors as:

$$\sigma^2_H = (0.95\times10^2) + 0.64 + 0 + (7.2\times10^2) + 0.22 + 0 + 0$$
$$\sigma^2_T = 8.2\times10^2\ \text{ft}^2$$

This estimate for the variance of errors associated with a calibration target for head is the expected error between the model simulated and the observed heads within the local model. That is, the standard error σ_H = 29 ft is the average error to be expected when comparing simulated heads from the local model to observed heads in the shallow aquifers.

E.2. Errors Associated with Regional Head Targets

The temporal variability of deep aquifer heads in the regional model is much less influenced than shallow aquifer heads by recharge and stream levels. Nicholas et al. (1987) noted that the temporal variability appeared to be correlated with seasonal pumping, and was as little as plus or minus 1.5 ft in locations away from pumping centers. Assuming a normal distribution, the error variance for temporal variability would then be:

$$\sigma_t^2 = \left[\frac{1.5}{3}\right]^2 = 0.25 \text{ ft}^2$$

However, temporal variability increases near pumping centers, and has been reported to have strong trends (Burch, 2002), thus unmodeled temporal variability may be much higher near pumping centers.

Observed heads in the deep aquifers generally are measured using airlines, a method that is more prone to measurement error than the steel tape method used for the shallow aquifers. Burch (2002) estimated that the maximum error of airline measurements in the deep aquifers is 10 ft. Assuming this error to be normally distributed, the variance of measurement error is estimated to be:

$$\sigma_m^2 = \left[\frac{10}{3}\right]^2 = 11 \text{ ft}^2$$

Unlike the local model, scale-up errors associated with vertical averaging in the regional model cannot be neglected because heads in the deep aquifer typically are observed in wells open to multiple hydrostratigraphic units (Burch, 2002). Nicholas et al. (1987) found that the hydraulic head in the St. Peter, Ironton-Galesville, and Elmhurst-Mt. Simon aquifers differed by approximately 60 ft (variance of 7.5×10^2 ft^2) along a borehole located away from major water supply wells. Unfortunately, the majority of head observations in the deep aquifers are near pumping centers, where model simulations suggest that pumping induces great differences in head levels between aquifers. In such areas, the variance of model-simulated heads is 3.2×10^4 ft^2 (see Section 3.2.1.2 for plots of model simulations of transient heads along observation wells in the deep aquifers). That is, the variance of scale-up errors due to vertical averaging could be $\sigma_s^2 = 7.5 \times 10^2$ ft^2 away from pumping centers and model simulations suggest that it could be $\sigma_s^2 = 3.2 \times 10^4$ ft^2 near pumping centers. This variance would be negligible for wells open to single hydrostratigraphic units that are explicitly represented as layers in the regional model.

The error due to unmodeled heterogeneity is assessed for the area of greatest resolution (northeastern Illinois), a region that is approximately 60 mi wide (3.3×10^5 ft). Gelhar (1993) suggests the correlation length will be approximately 1/10 the model scale, or 3.3×10^4 ft. No region-specific estimates are available for the variance of *Ln K*, but Gelhar (1993) suggests it is approximately 1.0 for a model of this scale, and ranges from 0.36 to 2.0. Burch (2002) gives the gradient as J = 32 ft/mi = 0.0061 ft/ft. With substitution, a conservative estimate (using the low estimate of $\sigma_{\ln K}^2$ = 0.36) for the variance of scale-up error due to unmodeled heterogeneity is:

$$\sigma_K^2 = 0.46\,(0.36)(0.006)^2\ (3.3\times10^4)^2 = 6.7\times10^3\,\text{ft}^2$$

That is, the standard error of scale-up due to unmodeled heterogeneity is approximately 8.2×10^1 ft, based on conservative estimates of heterogeneity from studies at similar sites. An estimate using the moderate estimate of heterogeneity of $\sigma^2_{\ln K} = 1$ yields:

$$\sigma_K^2 = 0.46\,(1)(0.006)^2\ (3.3 \times 10^4)^2 = 1.9 \times 10^4\,\text{ft}^2$$

The interpolation error in the regional model varies with the grid spacing; at the minimum grid spacing of 2500 ft, $d_{max} = \left[1250^2 + 1250^2\right]^{1/2} = 1.8 \times 10^3$ ft. With substitution, the error variance due to interpolation is:

$$\sigma_i^2 = \left[\frac{(0.0061)(1.8 \times 10^3)}{3}\right]^2 = 1.3 \times 10^1\ \text{ft}^2$$

The density of groundwater varies within northeastern Illinois, but Mandle and Kontis (1992) note that the effects of density do not affect groundwater flow except for deep within the Illinois and Michigan basins. For the freshwater portions of the domain emphasized in this study, the errors due to unmodeled salinity are assumed to be negligible, but this error can be large in deep, saline formations. The errors due to numerical approximation should be on the order of the convergence tolerance for the numerical solution (0.01 ft or less), and are neglected.

For wells open to single hydrostratigrapic units in the nearfield of the regional model, the total error variance associated with a head calibration target is estimated as:

$$\sigma_H^2 = (0.25) + (1.1 \times 10^1) + (0) + (6.7 \times 10^3) + (1.3 \times 10^1) + 0 + 0$$
$$\sigma_H^2 = 6.7 \times 10^3\,\text{ft}^2$$

That is, the standard error $\sigma_H = 82$ ft is the average disagreement when comparing simulated heads from the nearfield of the regional model to heads observed in wells open to single hydrostratigraphic units that are distant from pumping.

Errors associated with calibration targets for head vary widely by location and quality within the regional model. For example, wells open to multiple aquifers near pumping centers have an estimated total error variance of:

$$\sigma_H^2 = (0.25) + (1.1 \times 10^1) + (3.2 \times 10^4) + (6.7 \times 10^3) + (1.3 \times 10^1) + 0 + 0$$
$$\sigma_H^2 = 3.9 \times 10^4\,\text{ft}^2$$

That is, the standard error $\sigma_H = 200$ ft is the average disagreement when comparing simulated heads in the nearfield of the regional model to heads observed in wells open to multiple hydrostratigraphic units. In general, the error variance increases with grid spacing, model scale, heterogeneity, and density dependence. Estimating these errors goes beyond the available data, but the general principles outlined above indicate that calibration targets for head in the farfield of the model, with poorly defined heterogeneity, and deep in the Illinois Basin may have standard errors greater than 200 ft. These observations were assigned very low weights when calibrating the regional model.

E.3. Flux Target Errors

As noted in Section 2.3.2.2, calibration targets for flux were developed from stream gaging records and the ILSAM flow-accounting model for watersheds within the modeled domain. The flux targets represent the long-term average of total groundwater discharge, or base flow, to streams and drains within the watershed, The target values are estimated as the arithmetic average of Q_{80} and Q_{50} (Table E-1 and Table E-2).

Similar to the variance of head target errors, the error variance for flux targets might be decomposed into the independent contributing errors. However, unlike the head targets of this study, the flux targets are determined for a wide area and a long duration, similar to the areas and times simulated within the models. As a consequence, these flux targets are temporally and spatially consistent with the watershed-wide model fluxes at steady state, and this study has assumed the errors due to spatial variability, model resolution, and temporal variability to be negligible. The remaining identified sources of error are measurement errors, thus the total error variance associated with the calibration targets for long-term average flux are:

$$\sigma_Q^2 = \sigma_m^2$$

where

σ_Q^2 = the error variance associated with a calibration target for flux

σ_m^2 = variance of measurement (or simulation) errors for streamflow

Measurement errors for streamflow statistics vary depending on how the statistics were determined. For gaged stations such as those along the Fox River, the standard estimate of error for streamflow statistics is 10 percent or less, although this error can be larger for extreme values. For ungaged watersheds, errors for streamflow increase with the average permeability of the subsoil within the watershed and are proportional to the magnitude of streamflow. Within the Illinois Streamflow Assessment Model (ILSAM) in northeastern Illinois, two error rates have been inferred, one for watersheds with low permeability subsoils and another for high permeability subsoils (Knapp et al., 2007). The notable exception is Boone Creek, which may be receiving groundwater from outside its watershed and thus is less reliable, although this has not been quantified (Knapp, personal communication). Although errors in the statistics for gaged watersheds generally are less than those noted for ILSAM, extreme quantiles such as Q_{80} are thought to have slightly higher errors and thus this study will conservatively assume that the larger error variances inferred from ILSAM also apply to all flux targets in the model.

Table E-1. Flux Targets for Calibration of Predevelopment Steady-State Regional Model

Gage Name	***Q_{80} (ft^3/d)***	***Q_{50} (ft^3/d)***	***Q Target (ft^3/d)***	***ILSAM Error (percent)***	***Q Target Error (ft^3/d)***
Blackberry Cr near Yorkville, IL	-1,969,920	-3,542,400	-2,756,160	12	-330,739
Ferson Cr near St Charles, IL	-578,880	-1,771,200	-1,175,040	27	-317,261
Boone Cr near McHenry, IL	-578,880	-864,000	-721,440	27	-194,789
Coon Cr at Riley, IL	-829,440	-2,505,600	-1,667,520	27	-450,230
Skokie River near Highland Park, IL	-501,120	-1,166,400	-833,760	12	-100,051
Weller Cr at Des Plaines, IL	-101,952	-293,760	-197,856	12	-23,742.7
Turtle Cr at Carvers Rock Rd near Clinton, WI	-5,184,000	-7,862,400	-6,523,200	27	-1,761,264
White River near Burlington, WI	-2,246,400	-5,097,600	-3,672,000	27	-991,440

Table E-2. Flux Targets for Calibration of Local-Scale Model

Watershed	***Q_{80} (ft^3/d)***	***Q_{50} (ft^3/d)***	***Q Target (ft^3/d)***	***ILSAM Error (percent)***	***Q Target Error (ft^3/d)***
Big Rock Cr	-362,880	-2,220,480	-1,291,680	12	-155,002
Blackberry Cr	-907,200	-2,073,600	-1,490,400	12	-178,848
Coon Cr	-829,440	-2,505,600	-1,667,520	27	-450,230
Ferson Cr	-578,880	-1,771,200	-1,175,040	27	-317,261
Mill Cr	-103,680	-561,600	-332,640	12	-39,916.8
S Br Kishwaukee River	-129,600	-725,760	-427,680	12	-51,321.6
Tyler Cr	-267,840	-915,840	-591,840	12	-71,020.8
Union Ditch No 3	-362,880	-1,702,080	-1,032,480	12	-123,898

E.4. References

Anderson, M.P. and W.W. Woessner. 2002. *Applied groundwater modeling: Simulation of flow and advective transport*. Academic Press, San Diego, CA.

Burch, S.L. 2002. *A Comparison of Potentiometric Surfaces for the Cambrian-Ordovician Aquifers of Northeastern Illinois, 1995 and 2000*. Illinois State Water Survey Data/Case Study 2002-02, Champaign, IL.

Gelhar, L.W. 1993. *Stochastic subsurface hydrology*. Prentice Hall, Englewood Cliffs, NJ.

Hill, M.C. and C.R. Tiedeman. 2007. *Effective Groundwater Model Calibration*. John Wiley and Sons, Hoboken, NJ.

Knapp, H.V., A.M. Russell, J.A. Kramer, and G.P. Rogers. 2007. *Kane County Water Resources Investigations: Surface Water Accounting for Water Supply Planning in Kane County*. Illinois State Water Survey Contract Report 2007-10, Champaign, IL.

Larsen, R.J. and M.L. Marx. 1986. *An Introduction to Mathematical Statistics and Its Applications*. Prentice-Hall, Englewood Cliffs, NJ.

Locke, R.A., II and S.C. Meyer. 2005. *Kane County Water Resources Investigations: Interim Report on Shallow Aquifer Potentiometric Surface Mapping*. Illinois State Water Survey Contract Report 2005-04, Champaign, IL.

Mandle, R.J. and A.L. Kontis. 1992. *Simulation of Regional Ground-Water Flow in the Cambrian-Ordovician Aquifer System in the Northern Midwest, United States*. United States Geological Survey Professional Paper 1405-C, Washington, DC.

Nicholas, J.R., M.G. Sherrill, and H.L. Young. 1987. *Hydrogeology of the Cambrian-Ordovician Aquifer System at a Test Well in Northeastern Illinois*. United States Geological Survey Water-Resources Investigations Open-File Report 84-4165, Urbana, IL.