Appendix A – Supplemental Text

1. Calibration of Reference Model

The geologic framework on which the model geometry was based was from a Delaware Geological Survey (DGS) report (Andres and Klingbeil, 2006). This framework describes thickness for each of those three aquifers and a depth to the St. Mary's Formation, which has low permeability and acts as a regional confining unit (e.g. Achmad and Wilson, 2012; Andres and Klingbeil, 2006; Hodges, 1984). In initial calibration efforts, hydraulic conductivity (K) values were adjusted individually within three geologic units. Attempts to calibrate with an automated calibration package (Hill et al., 2000) by adjusting K values within each layer to match well head values did not produce K values that were reasonable considering field measurements. Similarly, manual attempts to adjust K values of each layer also failed to improve a match to hydraulic head and streamflow data relative to a homogeneous model when applying reasonable K values. Therefore, in the interest of parsimony, a homogeneous model in which layering was represented as vertical anisotropy in K was chosen as the Reference model.

A homogeneous and anisotropic representation of aquifer K achieved a good match between simulated and measured hydraulic head and streamflow data (Fig. A1). An added advantage of the homogeneous representation is that it allows examination and a more generic understanding of how hydraulic conductivity affects flux and travel times to stream and bay boundaries without added complications that would arise from a more heterogeneous model.

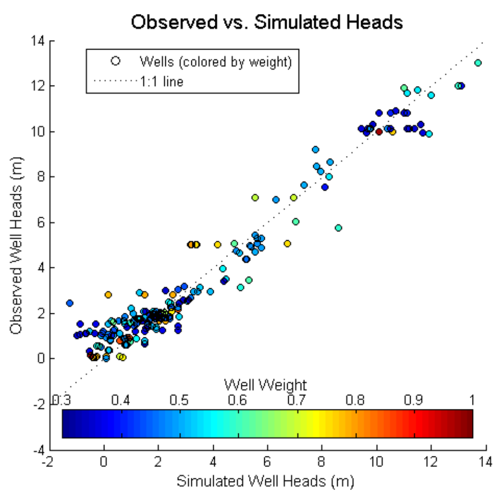

Figure A1: Simulated vs. observed heads in the Reference model. The plot shows reasonable matching throughout the range of heads measured within the model domain.

Figures A2 and A3 show that the parameters selected for the reference model (Parameter multiplier=1) produce a good match between simulated and measured heads and stream discharge values. Figure A3A shows that the calibration is sensitive to both Kh and recharge, and that equal changes to Kh and recharge result in very similar calibration values—recharge:Kh=1:1 line is shown as a dashed red line in Fig. A3A. Matching between measured and simulated stream discharge is sensitive to recharge but much less sensitive to Kh (also shown in Fig. A2). Thus, similar matching between simulated and observed stream discharge values will be met with changes to Kh when recharge is held constant—this is shown as a dashed vertical line in

Fig. A3B. The Reference model parameters were selected at the intersection of these two lines. Whereas calibrating to either well heads or stream discharge could yield a less-unique fit, calibrating to both wells and stream discharge allowed identification of unique Kh and recharge values that represent the system well.

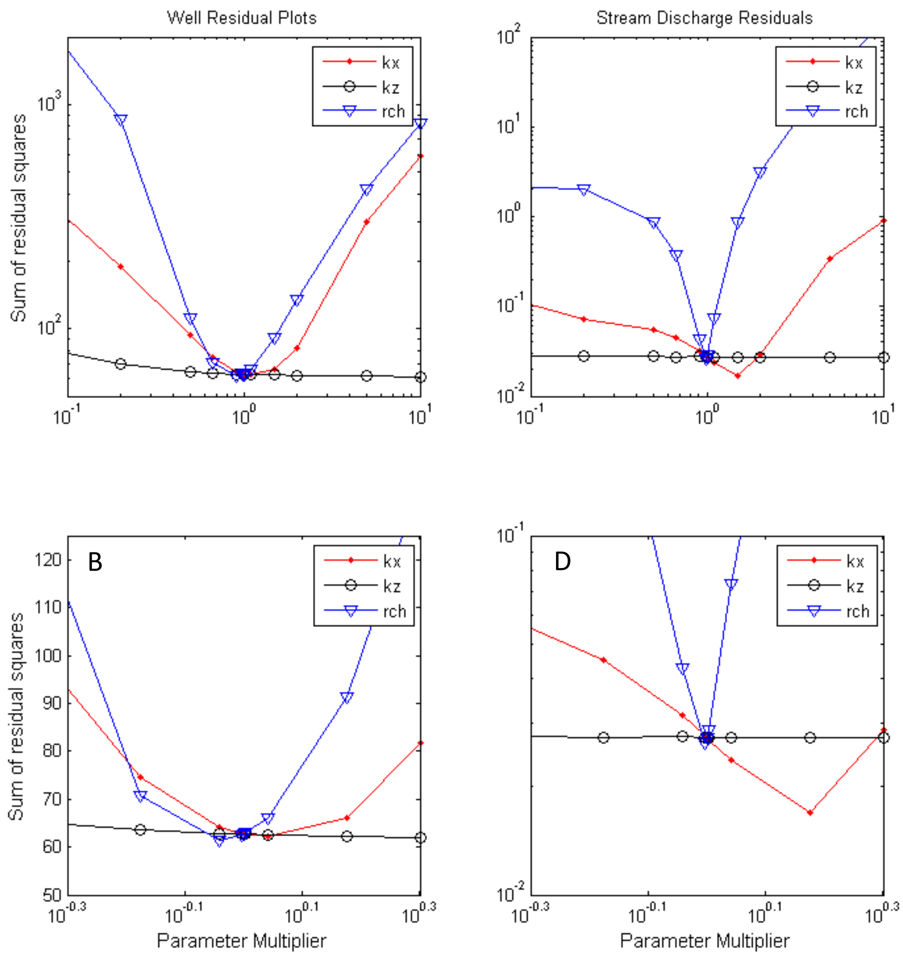

Figure A2: Sum of squared residuals for well hydraulic head (A & B) and stream baseflow (C & D) for all model runs presented in the sensitivity analysis (Fig. 4 in paper). Bottom plots show close-up view of top plots. Note lack of sensitivity to Kv.

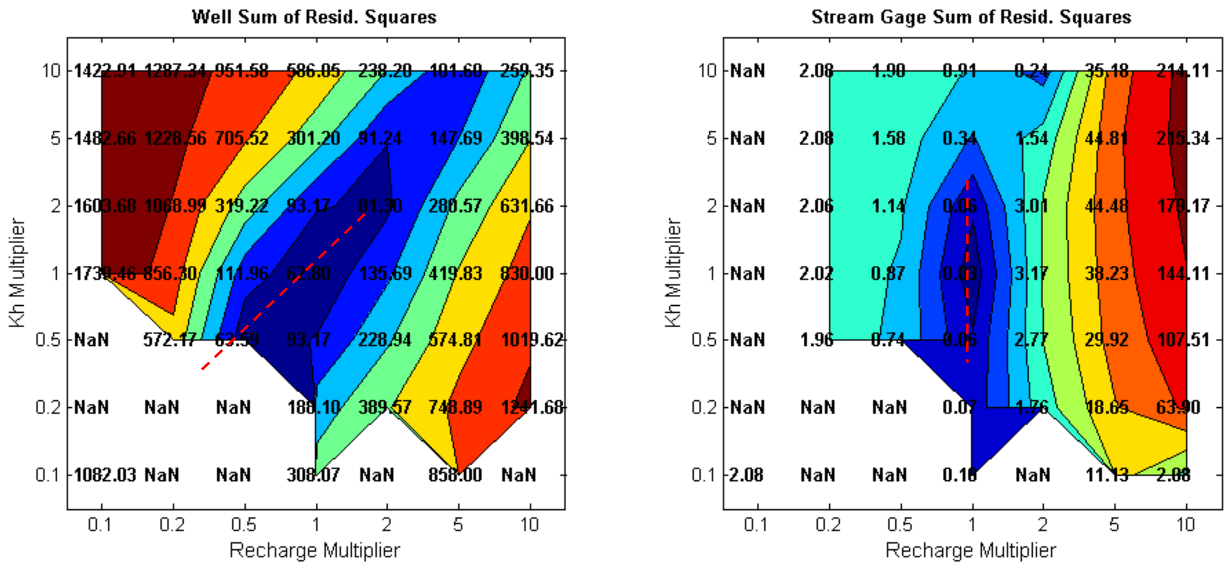

Figure A3: Contour plots of the sum of squared residuals for wells (left) and stream baseflow (right) for an array of 49 models spanning 7 values of Kh and recharge. Values for Reference model at center. Dashed red lines show regions of best fit for well (A) and stream (B) calibration targets. The parameters selected for our Reference model were at the intersection of those two lines where simulated well and stream data best matched modeled results for both calibration targets simultaneously. NaN represents models that failed to converge or in which modeled streamflow was zero. Note that Fig. A2 shows higher resolution of the calibration results about the model center.

2. *Particle Tracking*

2.1. Calculating flux-and-velocity-weighted mean groundwater transit time

Flux-and-velocity-weighted mean groundwater transit time, the average age of discharging groundwater (Maloszewski and Zuber, 1982; McGuire and McDonnell, 2006), was calculated for each model run from the ages of particles (MODPATH; Pollock, 2012) reverse-tracked from each cell with groundwater discharge to a stream or bay (a *discharge cell*). This was calculated in two steps:

First, each discharge cell was assigned a velocity-weighted transit time (TT_v ; eq A1) from the mean age (age_i = time between recharge and discharge) of all particles discharging to that cell, weighted by their velocities ($weight_i$ = particle pathlength ÷ particle age).

$$TT_v = \frac{\sum_{i=1}^{n} age_i * weight_i}{\sum_{i=1}^{n} weight_i} \qquad \text{(A1)}$$

Second, domain-wide flux-and-velocity-weighted transit times ($TT_{f\&v}$; eq A2) were calculated for each simulation from the velocity-weighted transit time of each discharge cell (age_j) weighted by the modeled flux through that cell ($flux_j$).

$$TT_{f\&v} = \frac{\sum_{j=1}^{n} age_j * flux_j}{\sum_{j=1}^{n} flux_j} \qquad \text{(A2)}$$

2.2. Sensitivity of flux-and-velocity-weighted mean transit time to the number of particles released per cell

The number of particles reverse-tracked from each cell (Figure A4) affects the calculated flux-and-velocity-weighted mean transit time. Domain-wide transit time decreased somewhat as additional particles were reverse-tracked from each cell, and the age distribution stabilized when at least 125 particles were assigned per cell. This is a much larger number of particles than is typically used in similar studies. However, figure A4 clearly indicates the importance of including a relatively large number of particles when calculating transit times in this manner.

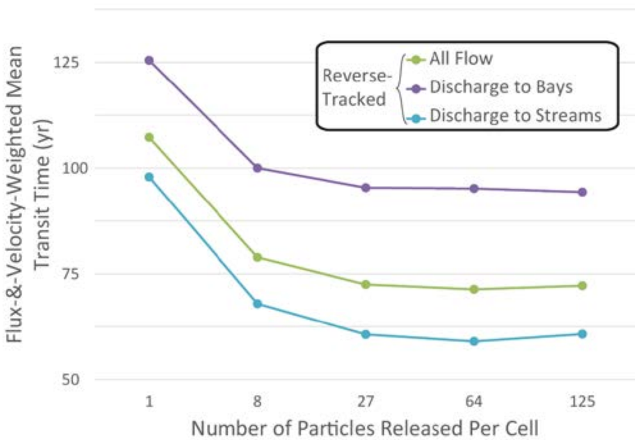

Figure A4. Sensitivity of flux-and-velocity-weighted mean transit time to the number of particles reverse-tracked from each discharge cell.