CHAPTER 4

Evaluation of NPS Return Flow to the River Using a Water Balance Model

4.1. Water Balance Model Applied to the LARV

Water Balance Model Equation.

The purpose of the water balance model is to determine the volume of unaccounted for water in each reach. We begin with a basic water balance model as describe in most hydrology texts — (Wanielista, Kersten, Eaglin, et al. 1997).

change in storage = inputs - outputs

Adding the variables, both known and unknown, present in the LARV we have the following equation:

$$\frac{\Delta S}{\Delta t} = Q_{in,US} + \sum Q_{in} + P + R + B - Q_{out,DS} - \sum Q_{out} - E - T - F \tag{1}$$

Where:

$\frac{\Delta S}{\Delta t} =$ Stored volume change between time steps.

$Q_{in,US} =$ Flow in the river entering the study reach at the upstream end.

$\sum Q_{in} =$ Flow gained by the river from tributaries and other gauged sources.

$P =$ Volume of water gained to the river due to precipitation falling directly on the river's surface.

$R =$ Volume of water gained to the river due to precipitation runoff from adjacent land.

$B =$ Volume of water gained to the river due to subsurface flow.

$Q_{out,DS} =$ Flow in the river leaving the study reach at the downstream end.

$\sum Q_{out}$ Flow lost from the river to canals and other gauged sinks.

$E =$ Volume of water lost from the river due to direct evaporation from the water's surface.

$T =$ Volume of water lost from the river due to plant transpiration.

$F =$ Volume of water lost from the river due to infiltration into the subsurface flow.

If we combine the terms that are unknown or unmeasured, we arive at the following equation:

$$\frac{\Delta S}{\Delta t} = Q_{in,US} + \sum Q_{in} + P - Q_{out,DS} - \sum Q_{out} - E + Q_{NPS} \tag{2}$$

Where:

$Q_{UNPS}=$ The sum of gains from non-point sources and losses to non-point sinks ($Q_{NPS} = R + B - T - F + Q_{U,in} - Q_{U,out}$).

There is no reasonable method for differentiating the components of Q_{UNPS}, therefore the abbreviation NPS in this thesis refers to both non-point sources and non-point sinks. Q_{UNPS}includes the non-point source gains from groundwater sources(B), non-point source losses to groudwater sinks F), transpiration losses from plants in the river channel (T), and gains from precipitation runoff from adjacent land (R). Additionally, this term includes ungauged flows leaving and entering the river. Ungauged gains to the river ($Q_{U,in}$) are suspected to be primarily in the form of irrigation drainage from adjacent farmland. Other sources could be due to errors in underestimating flows entering the river or overestimating flows leaving the river. Ungauged losses from the river ($Q_{U,out}$) are suspected to be primarily in the form of minor or unauthorized withdrawls from the river channel. Of the ungauged flows, irrigation drainage from adjacent farmlands is assumed to be the largest contributor.

The two groundwater components of $Q_{UNPS}$are suspected of being the largest components of Q_{UNPS}. Water transfer between the aquifer and river happens continually whereas $Q_{U,in}$, $Q_{U,out}$, R, and T are not continious. $Q_{U,in}$ and $Q_{U,out}$ only occur periodically when individuals activly withdraw from the river or allow irrigation runoff to return to the river. R only occurs during rain events. Within the LARV, most rainwater is captured in irrigation canals. Only precipitation falling in the riparian zone is likely to reach the river. T only occurs during growing season. This value is also only considering the transpiration happening within the river channel and does not include the riparian zone. Any losses due to transpiration in the riparian zone are first considered river losses to the aquifer (F).

Re-arrangeing equation (2) 2 to solve for the unknown values produces equation 3. Due to the nearly identical method of calculating flow (Q) and it's associated error and uncertainty, these terms were associated with each other. Likewise, the precipitation (P) and evaporation (E) terms were associated with each other.

$$Q_{UNPS} = \left(Q_{out,DS} + \sum Q_{out} - Q_{in,US} - \sum Q_{in}\right) - \frac{\Delta S}{\Delta t} - (P + E) \tag{3}$$

4.2. Stochastic and Deterministic Models

Determinsitic and stochastic models are used in both the unaccounted for water and mass balance models. Deterministic models are fully determined by the input parameters or variables. Randomness is not included. Stochastic models include one or more random parameters. Given the same input parameter values, a stochastic model will produce different results with each itteration.

There are many recoginzed methods for solving stochastic models. Solutions to these models are not definite and the term ”solve” must be taken loosely. Any solution from a stochastic model is one of a potentially infinte number of possible solutions.

The Monte Carlo (MC) simulation technique was used to obtain solutions for all stochastic models in this thesis. The MC technique is conceptually simple. The stochastic model is repetetively solved and the solution of each itteration are used to define the solution statistics of the model. The number of itterations performed is determined in one of two ways. One way is to calculate identifier statistic(s) after each run. Identifier statistic(s) are those that the modeler has determined to be of value. Usually, these statistics are monitored to identify when the change in the statistic has reached a predtermined threshold. The alternate method of determining the number of itterations to perform is more fixed. The model is run for a estimated number of itterations. A set of results for each itteration is saved. After all itterations are calculated he identifier statistic(s) are calculated for each itteration. The modeler then determines the number of itterations based on the results.

Both methods for determining the number of required itterations have their own benefits and drawbacks. The first method is more computationally intensive and is slower to solve, but the number of required itterations is never exceeded. The second method is quicker to solve, but may require more model runs if the modeler underestimated the number of required itterations. This method also has the added benefit that if, after the model is run, the modeler determines that a differnet identifier statistic is required, then that statistic can be calcualted on the existing solution set. Due to the complexity of the models in this thesis, it was determined that the second method was the most appropriate.

It was determined that for the sake of simplicity, all of the models calculated in this thesis would use the same number of itterations. The USR mass balance model is the most complex model as it has the largest number of input variables and uncertainty terms. The identifier statistics used were the mean, variance, and skewness, which are the first, second, and third moments of the probability density. These were calculated for each itteration. The threshold between the observed itteration and the previous itteration was fixed at 0.1%. The identifier statistics reached the threshold in the following order: mean, variance, and skewness. Skewness reached its breakpoint shortly before the 500th itteration. A judgement call was made to increase the factor of safety. Therefore, the number of stochastic model itterations was fixed at 5,000.

4.3. Error and Uncertainty.

Error and uncertainty are two terms that appear equivalent, but are not. Error is the difference between the true value and the measured value. This variability is due to measuring instrument inaccuracy and inprecision. No instrument is perfectly accurate or precise and this variation may not be capturable. Error of this type is usually specified by the instrument manufacturer. Error is a combination of systemic error and random error. Measurement errors due to instrument calibration shifts and other known or unknown sources is systemic error. Random error is the that which remains even after all systemic errors are accounted for. Random error varies unpredictably and is uncorrectable. Error does not significantly impact this study. All studies that use this data or similar data will be subjected to the same error. When compared to uncertainty, error is nearly insignificant.

Uncertainty of a measured value is an interval around that value such that any repetition of the measurement will produce a new result that lies within this interval. An uncertainty estimate should address error from all possible effects (both systematic and random) and, therefore, usually is the most appropriate means of expressing the accuracy of results. This is consistent with ISO guidelines . However, in the case of this thesis, most measurements only include random error and systematic error is not address. Since, for most of the data used in this thesis, we are not the data originators, there is no reliable way to account for systemic error. In most cases, the data originators have provided uncertainty ranges which include instrument measurment random error and uncertainties due to temporal variations of the measured location.

ref

Reported flow rates are an excellent example of uncertainty values. The reported value includes a range, in this case given as a percentage of the reported value. The true value lies somewhere within this range. The National Institute of Science and Technology has defined how uncertainty is to be reported. As per the Engineering Statistics Handbook (), there are four tenents for stating and combining components of uncertainty:

Ref-year only

(1) Each uncertainty component is quantified by a standard deviation.

(2) All biases are assumed to be corrected and any uncertainty is the uncertainty of the correction.

(3) Zero corrections are allowed if the bias cannot be corrected and an uncertainty is assessed.

(4) All uncertainty intervals are symmetric.

In this example, the flow rate standard deviation is stated as a percentage of the reported flow. This indicates that the uncertainty increases as the flow rate increases. Biases are assumed to be corrected since none are reported.

The true value of a parameter is not directly observable and is only valid for a specific time and location. Any repeated measurements will be subjected to temporal variablilty. Measurements taken by two identical instruments will be subjected to temporal variability. The measured value is the value that is measured by the instrument. The difference between the two values is uncertainty. The relationship between measured value, true value, and uncertainty is described in equation 4.

$$\mathrm{mv}=\mathrm{tv}\ +\varepsilon \tag{4}$$

Where:

mv = measured value

tv = true value

ε = uncertainty

If the true value were known, then it would be a simple matter of algebra to determine the error from any measured value 5. This second equation shows uncertainty as a positive value when, algabraically, it should be negative. ε remains a positive term because it is assumed to be equally distributed around zero.

$$\mathrm{tv} = \mathrm{mv} + \varepsilon \tag{5}$$

Figure 4.1 is a graphical representation of a pair of possible scenarios when considering true and measured values. In all of the graphs, the uncertainty distribution is shown as the dashed black curve. The vertical black line at zero indicates the error associated with the true value. The vertical red and blue lines indicates the error for the two scenarios. The red line error value is near the black line. This indicates that the error is small and the measured value is near the true value. The blue line error value is a significant distance from the true value, but is within the uncertainty distribution. This indicates that the error is large and the measured value is significantly different than the true value. Both scenarios are plausible as the error values are within the uncertainty distribution associated with the true value. Graphs 1 and 2 show the red and blue scenarios.

The application of equation 5 can produce undesireable results. They are shown on graphs 3 and 4 of figure 4.1. In both of the scenarios, the uncertainty distribution of the true value is applied to the measured values and is depicted as dashed red and blue curves. In the red scenario where the error is small, depicted in graph 3, the overlap of the true value uncertainty distribution and the measured value uncertainty distribution is large and are almost identical. In the blue scenario, depicted in graph 4, the overlap is small. The overlap areas in graphs 3 and 4 are shaded gray. These gray areas are representative of the probability that the true value and measured value are from the same distribution.

Ideally, the true value would be estimated based on the measured value and the relationship between the true value and the particular measured value. Since the true value

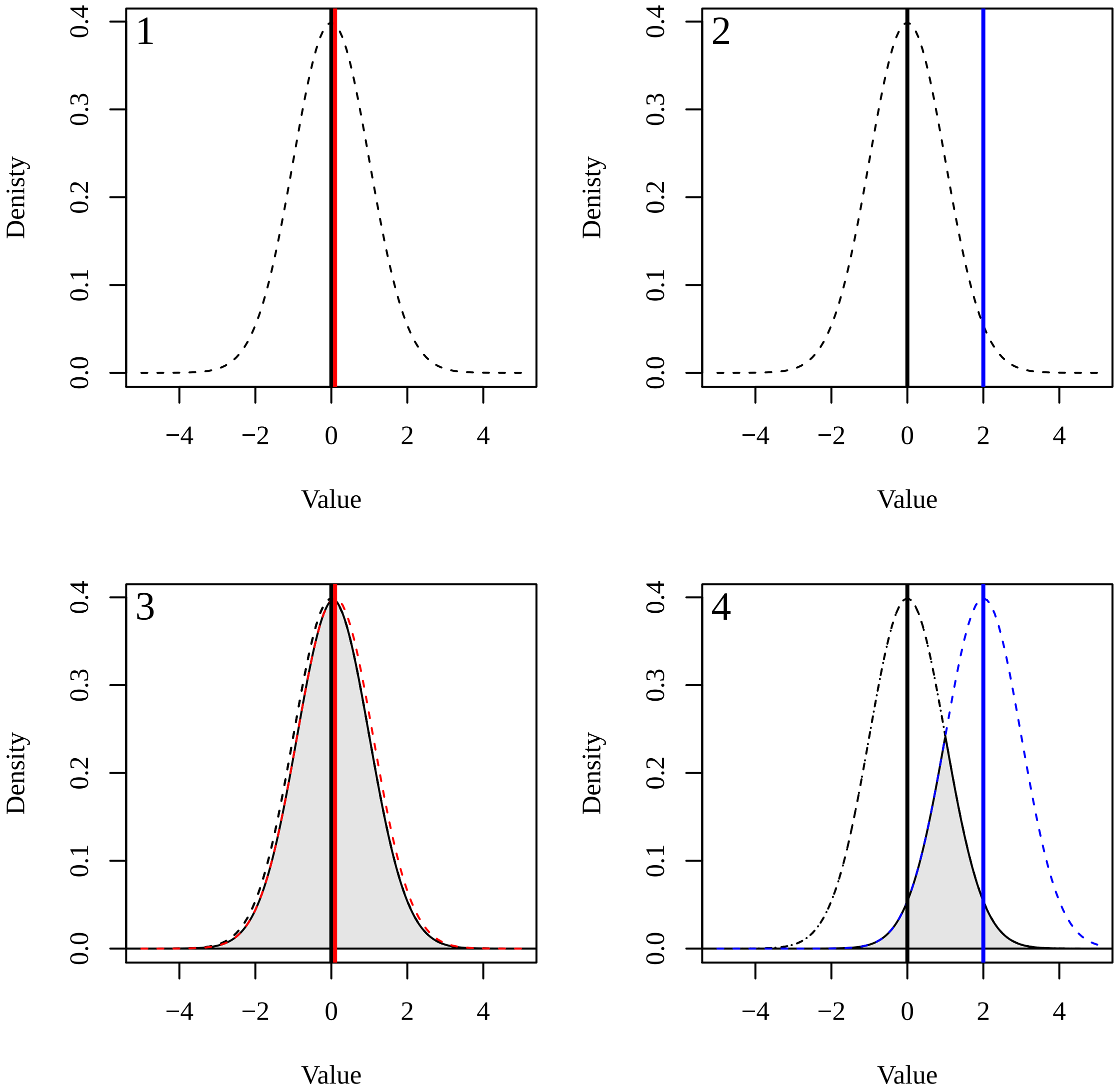

Figure 4.1. Comparison of measured value and true value. The black vertical line depicts the error associated with the true value. The black dashed line is the distribution of the uncertainty of the measured value from the true value. The red and blue vertical lines depict two measured value scenarios. The red and blue dashed curved depict the uncertainty distributions applied to the associated measured values. The grey area is representative of the probability that the true value and measured value are from the same distribution.

is not known, the only reasonable course is to estimate the true value as in equation 5. Additional uncertainty can be attributed to this approach, but the uncertainty would be variable based on the difference between the true and measured value. Since this difference is unknown, it is unreasonable to guess.

III - define univariate probability distributions as estimate discriptions of an uncertainty parameter

Defining uncertainty is quite cumbersome as each variable comes with its own distinct definition of uncertainty. The definitions usually come in the form of a parametric distribution.

Distributions require as few as one factor. When a distribution had more than one definition, only standard, accepted definition was used. This thesis does not discuss the computations of the different distributions. The primary concern regarding distributions is to determine which of the well established distributions best fit the data.

References

Wanielista, Martin, Robert Kersten, Ron Eaglin, et al. (1997). *Hydrology: Water quantity and quality control.* John Wiley and Sons.