AN ANALYTICAL APPROACH FOR EVALUATING THE INFUENCE OF UNCERTAINTY IN UNIMPAIRED FLOW DEVELOPMENT

Yi Zhang[1] and Owen McKeon[2]

AUTHORS: [1]Basin Analysis Program, Environmental Protection Division, Georgia Department of Natural Resources, 2 Martin Luther King, Jr. Drive, SE, Suite 1058 East Floyd Tower, Atlanta, Georgia 30334; [2]ARCADIS U.S., Inc. 2849 Paces Ferry Road, Suite 400, Atlanta, GA 30339

REFERENCE: *Proceedings of the 2011 Georgia Water Resources Conference*, held April 11–13, 2011, at the University of Georgia.

Abstract. With certain increasing demands on Georgia's waters, the General Assembly recognized the need for a proactive and comprehensive plan to meet future challenges. Thus the Assembly authorized the Georgia Comprehensive State-wide Water Management Plan, Surface Water Availability Assessment. One of the important products developed from this assessment was Unimpaired Flows (UIF), which are the basis for all current and future water resources assessments in Georgia. However, an important question in regard to UIF is their reliability and uncertainty. In this article, an analytical approach was proposed and used to analyze and quantify uncertainty of all variables and factors that affect UIF development. Several typical statistical distributions, such as normal and uniform distributions, were investigated. Confidence intervals and statistics of UIF and resultant 7Q10 flows based on these different distributions were calculated. The results of this analytical approach are compared with a numerical solution to the problem in order to verify, validate, and assess the efficiency of the proposed approach.

INTRODUCTION

The history of demand and competition for Georgia's water resources is long and involves themes not only within Georgia herself but also with her neighboring states. In 2007, a historic drought in Georgia and the Southeast as a whole caused the triggering of the most stringent conservation operation of Corps' reservoirs. Consequently, high tension was also trigged within Georgia and with Alabama and Florida. With certain increasing demands on Georgia's waters, the General Assembly recognized the need for a proactive and comprehensive plan to meet future challenges. Thus the Assembly authorized the Georgia Comprehensive State-wide Water Management Plan in January 2008. The Surface Water Availability Assessment is an important part of this plan. The objective of this assessment is the quantifying of surface water availability within Georgia given current and future consumptive uses.

One of the most important steps for this assessment is the calculation of "unimpaired flows" (UIF). Unimpaired flows are defined for this study as historically observed flows with human influences removed [1]. UIF are used because they provide a consistent and unbiased basis for hydrologic and statistical analysis of surface water availability and unbiased assessment of impacts of water use on water availability within affected river basins.

From 2009, the Georgia Environmental Protection Division (GA EPD) has worked with ARCADIS Inc. and successfully developed UIF for all major river basins in Georgia for the period of 1939 to 2007. The UIF developed are now being used for Consumptive Use Assessment (CUA) under current and future water uses.

However, questions often faced are: "How accurate are the developed UIF and related 7Q10 flows"? What are the uncertainties involved in calculation of UIF and 7Q10 flows? What are the possible ranges and confidence intervals of UIF and 7Q10 flows after considering these uncertainties? How does the uncertainty in UIF and 7Q10 flows affect the consumptive use assessment? Answering such questions is not an easy task due to the complexity of the problem. This paper is aimed at developing a statistical analytical approach to answer these questions.

STATEMENT OF THE PROBLEM

Unimpaired flows are defined for this study as historically observed flows with human influences removed. Though simple in definition, the actual derivation of UIFs involves many steps. Each step introduces its own element of uncertainty. The problem faced is how to determine the total resulting uncertainty. The uncertainty of UIF calculations is the cumulative effect of the uncertainties of these steps.

Human influences considered in unimpaired flow derivation are: 1) flow regulation by reservoirs, 2) net reservoir evaporation, 3) water withdrawals, and 4) wastewater returns. Water withdrawals and returns are associated with municipal, industrial, thermal power, and agricultural water uses. The long term observed streamflow information to be utilized as a basis of unimpaired flows is provided by both active and inactive United States Geological Survey (USGS) gage stations. Therefore, in the technical analyses of water availability, these station locations are where UIFs are computed and thus are termed basic nodes. The unimpaired flow calculation process generally involves the following steps:

1. Collect and inventory streamflow data, reservoir data, and water use data. Identify data gaps and any need for data filling or time series data extension.

2. Compute reservoir inflows, holdouts, and releases from storage in river reaches upstream of nodes.

3. Compute net surface evaporation (evaporation less precipitation) from reservoirs in river reaches upstream of nodes.

4. Fill observed streamflow and reservoir inflow time series to produce time series data coverage for the entire period of record.

5. Compute local incremental flows by routing upstream observed flows and subtracting from downstream observed flows. These observed streamflow time series should have had any existing gaps filled prior to this routing. Negative locals may result depending on multi-hourly variability within daily average flow readings, reach lengths, and routing methods employed.

6. Compute local incremental unimpaired flows by adjusting local incremental flows to remove the following effects of reservoirs and human uses of water:

    a. Holdouts and releases from reservoir storage

    b. Net surface evaporation from reservoirs

    c. Net diversions (withdrawals less returns) in river reaches between nodes by municipal, industrial, thermal power, and agricultural water users; both direct (surface) water withdrawals and indirect (groundwater) pumping and resulting depletion of surface waters are accounted for in this step of the analysis

In each step of the above calculation process, uncertainty occurs due to various factors. For example, measurement errors of observed flow and water use data, data filling error and flow routing error etc. are sources of error. USGS gage data used for historical observed flows contain uncertainty from measurement error to various degrees. The magnitude of this measurement error needs to be quantified. The range of uncertainty in USGS field observation needs to be determined. Another example are gaps in observed gage flows if the observed flows in the referenced gages at nearby locations are used to fill the gaps by regression. In this case, uncertainty of observed flows in the referenced gage and regression approach need also be evaluated. Water use data and its range of uncertainty need to be determined. Moreover, since flow routing methods were used to route the upstream observed flow to a downstream node in UIFs development, uncertainty occurs during the routing process. This uncertainty also needs to be quantified. If there is a reservoir in the reach upstream of the basic node being computed, the magnitude of uncertainty in reservoir regulation and net evaporation from its free water surface should be evaluated. The uncertainty of UIFs calculation is the cumulative effect of uncertainties of all its components.

In order to reduce the complexity of the problem, a relatively simple case was considered, one without reservoir regulation.

METHODOLOGY

An approach to a solution is found by staring from the definition of UIFs. From this definition the uncertainties that affect UIF calculation can be identified. Furthermore, methods are found to quantify these uncertainties and determine the upper and lower bounds of variable uncertainty. With these methods, the accumulative effect of these uncertainties can be evaluated, and in turn, the possible range of UIFs and their confidence interval can be determined. Finally, based the range of developed UIFs, the range and confidence interval for 7Q10 flows can be determined.

In addition, an important aspect of this paper's method of solution is the use of an approximate analytical approach rather than a numerical approach such as Monte Carlo simulation.

UIF Calculation Formula. Based on above definition, the local UIF between two basic nodes can be expressed by the following general equation:

$$Q_{LUIF} = Q_{Obs} - R\left(Q_{Obs}^{up}\right) + \Delta S_{HO} + WU + E_{Net} \tag{1}$$

Where, Q_{LUIF} is the local incremental flow at the basic node being computed. Q_{Obs} is the observed flow at the basic node being computed. Q_{Obs}^{up} is the observed flow at the node upstream of the basic node being computed. ΔS_{HO} is the reservoir holdout in the reach upstream of the basic node being computed. WU is net water uses in the reach upstream of the basic node being computed. E_{Net} is the net evaporation of reservoir. Rt( ) is the flow routing function that routes flow at upstream node to downstream node, i.e., the basic node being computed.

The cumulative unimpaired flow at the basic node being computed can be expressed as:

$$Q_{CUIF} = Rt(Q^{up}_{CUIF}) + Q_{LUIF} \tag{2}$$

Where, Q_{CUIF} the cumulative UIF at the basic node being computed and Q_{CUIF}^{up} is the cumulative UIF at the node upstream of the basic node being computed.

If there is no reservoir in the reach, Equation (1) can be simplified as:

$$Q_{LUIF} = Q_{Obs} - Rt(Q^{up}_{Obs}) + WU \tag{3}$$

Substituting equation (3) into equation (2) gives:

$$Q_{CUIF} = Rt(Q^{up}_{CUIF}) + Q_{Obs} - Rt(Q^{up}_{Obs}) + WU \tag{4}$$

At a node where there are no human influences on water use as defined above (e.g., the headwater node of a basin), Equation (3) can be further simplified as:

$$Q_{CUIF} = Q_{LUIF} = Q_{Obs} + WU \tag{5}$$

Quantify Uncertainties of Variables and Components

As mentioned earlier, a simple case which does not include reservoirs is considered in this paper (Equations 3 and 4). Equation (3) shows that local unimpaired flow at a basic node is determined by (1) observed flow at the node – both actual and filled, (2) net water uses in the reach upstream of the node, and (3) routed flow from the upstream node. Uncertainty occurs for each variable or component.

(1a – actual observed flows) For observed flows, USGS gage data were used. However, uncertainty may be introduced because of measurement error by human error or by equipment error. USGS has field measurement rating for its data error. They rate measurement data by "excellent", "good", "fair" and "poor". "Excellent" means approximately 95% of the data is within 5% of actual flow. "Good" means approximately 95% of the data is within 10% of actual flow. "Fair" means approximately 95% of the data is within 15% of actual flow. Records which do not meet these criteria are rated "Poor". The "Poor" designation was not assigned a percentage by USGS, but was assumed to be within 20% of the actual flow for this paper. From these field measurements, the error range for observed flow for each day can be determined.

(1b – filled flows) Some observed flow records had missing data, or gaps. These gaps at the target gages were filled by using regression methods on observed flows at reference gages. To determine the range of uncertainty for such filled data, additional regressions were performed between the upper and lower bounds (UB, LB) of the dependent and independent data (d, i) used to develop the regression equations for filling gaps. There were four such regressions: LBi to LBd (i.e., the lower bound "LB" of the independent data "i" to the lower bound of the dependent data "d"), LBi to UBd, UBi to LBd, UBi to UBd. From these four regressions, two possible values for filled data upper bounds and two possible values for filled data lower bounds were found. The maximum upper bound and the minimum lower bound were selected from these four values to form the upper and lower bounds of the filled data at the target gage.

(2 – water use) For net water use, which is defined as the difference between withdrawals minus returns, lacked measurement error information. Therefore the assumption water made that net water use data has a 10% error range, i.e., ±5% of measured values.

(3 – routed upstream flow) For routed flows, uncertainty results from inaccuracies in the routing parameters and method and from inaccuracies in the observed flows being routed. Since determining the inaccuracies in the routing parameters and method is difficult, only the uncertainty resulting from measurement error in the observed flows was considered. Thus the uncertainty of routed flows was determined by routing the upper bound and lower bound of the upstream observed flows to obtain the upper and lower bound of the routed flows. An approach to evaluate uncertainty caused by the routing method itself is currently being investigating.

Quantify Uncertainty of UIF. After the uncertainty of each variable or component was quantified, the uncertainty of UIFs is the cumulative effect of the uncertainties of those variables and components. To quantify this UIF uncertainty, we need to make some assumptions for the component variables. We assume that the observed flow, net water use and routed flow are independent random variables and are either normally distributed or uniformly distributed within their upper bound and lower bound. Since the confidence interval for the uniform distribution assumption is larger than that for the normal distribution assumption, the uniform distribution assumption is more conservative., Therefore the assumption of a uniform distribution of component variables was chosen for this study.

The question to consider now is, what is the distribution of Q_{LUIF} and Q_{CUIF} in equations (3) and (4)? The Central Limit Theorem [2] can help to answer this question. The generalized Central Limit Theorem tell us that if S_n is the sum of n independently and identically distributed random variables X_i each having a mean, μ_i, and variance, σ_i^2, then in the limit as n approaches infinity, the distribution of S_n approaches a normal distribution with mean $E(S_n) = \sum \mu_i$ and variance $Var(S_n) = \sum \sigma^2$. One condition for this generalized Central Limit Theorem is that X_i has a negligible effect on the distribution of S_n. This general theorem is very useful for it states that if a hydrologic random variable is the sum of n independent effects and n is relatively large, the distribution of the variable will be approximately normal.

To apply this theorem to the above question, if one or two variables are dominating, the UIF should follow the uniform distribution, otherwise, it should be approaching normal distribution. As we mentioned earlier, since the confidence interval for the uniform distribution assumption is larger than the one under the normal distribution assumption, the result is more conservative if UIF are likewise assumed to be uniformly distributed. Hence, the confidence interval of UIF can be determined based on the uniform distribution. Finally, if the confidence interval of UIFs and their upper limit and lower limit are known, the upper bound and lower bound of 7Q10 flows can be calculated.

CASES STUDIES

Two case studies are presented to illustrate the approach.

Case Study 1. This case is at the basic node Claxton, which is a headwater node on the Ogeechee River . Since Claxton is a headwater node, there is no flow routed to it from upstream., Therefore, the cumulative UIF is the same as local UIF. This node has a complete observed flow record, therefore, no filling is required. The only uncertainties in UIF development are measurement errors from the observed flow and water use data at the Claxton basic node.

Figure 1 shows the developed UIF, the 95% confidence interval lower bound, and the 95% confidence interval upper bound for 2007. The figure shows that the developed UIF is within the 95% confidence interval. The average UIF flow at Claxton is 466 cfs and the average confidence interval range is 29.6 cfs. The ratio of the average confidence interval range and the average UIF is 6%, which is very small. Thus the error range of the developed UIF is 6% with 95% confidence.

Table 1 shows the monthly 7Q10s calculated from the developed UIFs and their respective 95% confidence intervals. Column 2 to 5 represents the lower and upper bounds of each month's confidence interval, the interval's range, and the percentage this range represents relative to the respective monthly 7Q10. From the table, we can see that the monthly interval ranges are small, varying from 5% to 12%. Most of the developed monthly 7Q10s are within their respective 95% confidence interval except for March, July, and December.

Table 2 shows the results by numerical approach based on Monte Carlo simulation by Jiang et al [3]. Comparison shows that the results by the two approaches are very close for this example. The percentage of error for the numerical approach is in general smaller than for the analytical approach. This result stems from the fact that the numerical method usually runs limited Monte Carlo simulations. When the number of Monte Carlo simulation runs becomes larger, the two approaches' results should be closer.

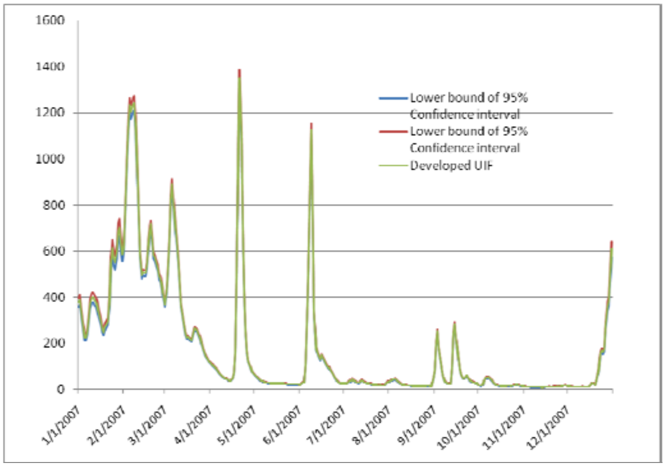

Figure 1: The calculated UIF and the lower bound and upper bound of its 95% Confidence Interval at Claxton.

Table 1: The calculated 95% Confidence Interval of developed 7Q10s by the analytical method at Claxton.

Month	Lower Bound 95% C.I. (CFS)	Upper Bound 95% C.I. (CFS)	Range 0f 95% C.I.	95% C.I.as a Percent of 7Q10	Calculated 7Q10
1	43.8	46.2	2.4	5%	45
2	58.3	61.4	3.0	5%	60
3	145.6	155.4	9.7	6%	142
4	32.6	35.4	2.8	8%	34
5	14.6	15.9	1.3	8%	15
6	6.0	6.4	0.3	5%	6
7	6.4	6.8	0.4	6%	7
8	8.1	8.8	0.7	8%	8
9	5.7	6.3	0.6	10%	6
10	3.9	4.4	0.5	12%	4
11	2.8	3.1	0.3	8%	3
12	6.1	6.5	0.3	5%	5

Case Study 2. This case is at the planning node at Penfield on the Oconee River. Upstream of this node is the Athens basic node. Consequently flow is routed from the Athens node to the Penfield node is necessary for calculating the local increment flow. There is missing observed flow data at the Penfield node. Therefore, data filling is needed and was done by regression. Uncertainties involved in UIF development at Penfield node include measurement error from the observed flow, data filling error, routing error and measured error of water use data in the reach upstream of the Penfield node.

Table 2: The calculated 95% Confidence Interval of developed 7Q10s by the numerical method at Claxton.

Month	Lower Bound 95% C.I. (CFS)	Upper Bound 95% CI. (CFS)	Range 0f 95% C. I.	95% C.I.as a Percent of 7Q10	Calculated 7Q10
1	40.7	42.5	1.8	4%	45
2	53.9	56.3	2.4	4%	60
3	111	117.3	6.3	5%	142
4	31.7	33	1.3	4%	34
5	14.1	15	0.9	6%	15
6	5.51	5.76	0.3	4%	6
7	6.39	6.66	0.3	4%	7
8	7.94	8.41	0.5	6%	8
9	5.63	5.88	0.3	4%	6
10	3.84	3.98	0.1	4%	4
11	2.81	2.94	0.1	4%	3
12	5.94	6.2	0.3	4%	5

(Note: in the tables, C.I. means Confidence Interval)

Figure 2 shows the developed UIF, the 95% confidence interval lower bound, and the 95% confidence interval upper bound for 2007.The figure shows that the developed UIF is within the 95% confidence interval. The average UIF flow at Penfield is 1238 cfs and the average confidence interval range is 131 cfs.. The ratio of the average confidence interval range and the average UIF is 10%, which is very small. Thus the error range of the developed UIF is 10% with 95% confidence.

Table 3 shows the monthly 7Q10s calculated from the developed UIFs and their respective 95% confidence intervals. Column 2 to 5 represents the lower and upper bounds of each month’s confidence interval, the interval’s range, and the percentage this range represents relative to the respective monthly 7Q10. From the table, we can see that the monthly interval ranges are small, varying from 11% to 28%. Most of the developed monthly 7Q10s are within their respective 95% confidence interval except for June and August.

Table 4 shows the results by numerical approach based on Monte Carlo simulation [3]. Comparison shows that the results by the two approaches are still very close for this example. The percentage of error for the numerical approach is in general smaller than for the analytical approach. This result stems from the fact that the numerical method usually runs limited Monte Carlo simulations. When the number of Monte Carlo simulation runs becomes larger, the two approaches’ results should be closer.

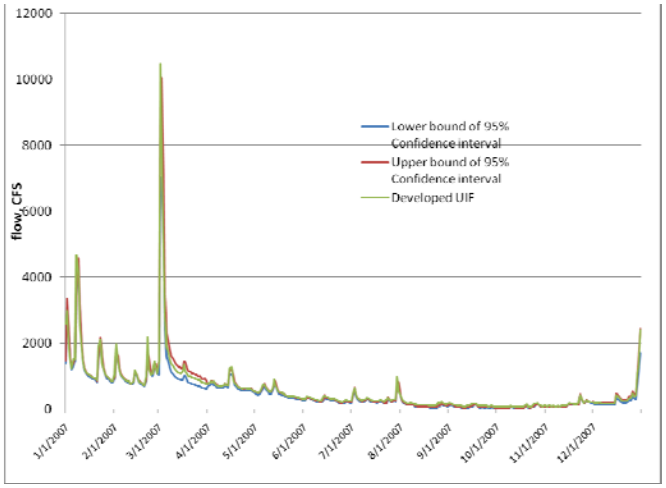

Figure 2: The calculated UIF and the lower bound and upper bound of its 95% Confidence Interval at Penfield

Table 3: The calculated 95% Confidence Interval of developed 7Q10s by the analytical method at Penfield

Month	Lower Bound 95% C.I. (CFS)	Upper Bound 95% C.I. (CFS)	Range 0f 95% C.I.	95% C.I.as a Percent of 7Q10	Calculated 7Q10
1	513	584	71.2	12%	539
2	591	671	80.2	12%	609
3	657	749	91.5	12%	684
4	595	666	71.5	11%	625
5	350	403	52.9	13%	395
6	257	289	32.0	11%	244
7	196	228	32.7	14%	193
8	114	157	43.2	28%	106
9	91	111	20.3	18%	106
10	146	165	19.7	12%	154
11	207	243	35.7	15%	229
12	319	401	82.7	21%	314

Table 4: The calculated 95% Confidence Interval of developed 7Q10s by the numerical method at Penfield

Month	Lower Bound 95% C.I. (CFS)	Upper Bound 95% C.I. (CFS)	Range 0f 95% C.I.	95% C.I.as a Percent of 7Q10	Calculated 7Q10
1	470.7	535.8	65.1	12%	539
2	568	608	40.0	7%	609
3	624.3	685.8	61.5	9%	684
4	592.2	626.4	34.2	5%	625
5	358.1	379.1	21.0	6%	395
6	207.9	249	41.1	17%	244
7	164.8	197.8	33.0	17%	193
8	98.9	108.6	9.7	9%	106
9	91	101.5	10.5	10%	106
10	125.9	151.8	25.9	17%	154
11	193.1	221	27.9	13%	229
12	293.4	317.1	23.7	7%	314

SUMMARY

This paper proposed an approximate analytical approach to quantify uncertainties during unimpaired flow development. Analysis starts from quantification of each variable or component in UIF calculation based on the USGS gage rating for measurement data, estimated water use error, error estimates resulting from the routing process, and data filling process. The lower and upper bounds of variables were obtained. The uniform distribution for variables and components of UIF calculation was assumed. Then the uniform distribution was used for quantifying the uncertainty of unimpaired flows based on the Central Limit Theorem. Use of the uniform distribution is more conservative than the use of the normal distribution.. The 95% confidence interval can then be calculated for UIFs and resultant 7Q10s.

Two case studies, one at the Claxton node and one at the Penfield node have been studied with different uncertainty factors. Results shown that the 95% confidence intervals are narrow as compared to the magnitude of their UIFs and 7Q10. This result give confidence that the error range is relatively small and within accepted levels for the developed UIFs and 7Q10s. The comparison with the numerical solution shows that the results by both the analytical method and the numerical method are close and consistent. However, the analytical approach is simpler and more efficient. These case studies suggested that the proposed analytical method provides an approximate and efficient approach for evaluating and quantifying the uncertainty in unimpaired flow calculation.

REFERENCES

[1] Georgia Department of Natural Resources. 2010, Unimpaired Flows Data Report.

[2] Haan, Charles T., 2002, Statistical Methods in Hydrology, Iowa State Press.

[3] Liang, Hailian etc., 2011, Uncertainty Analysis Of Unimpaired Flow And Monthly 7q10, Georgia Water Resources Conference.