Quantification of uncertainties in the 100-year flow at an ungaged site near a gaged station and its application in Georgia☆

Huidae Cho*, Emma Bones

Water Resources Engineer, Dewberry, 2835 Brandywine Road Suite 100, Atlanta, GA 30341

Abstract

The Federal Emergency Management Agency has introduced the concept of the "1-percent plus" flow to incorporate various uncertainties in estimation of the 100-year or 1-percent flow. However, to the best of the authors' knowledge, no clear directions for calculating the 1-percent plus flow have been defined in the literature. Although information about standard errors of estimation and prediction is provided along with the regression equations that are often used to estimate the 1-percent flow at ungaged sites, uncertainty estimation becomes more complicated when there is a nearby gaged station because regression flows and the peak flow estimate from a gage analysis should be weighted to compute the weighted estimate of the 1-percent flow. In this study, an equation for calculating the 1-percent plus flow at an ungaged site near a gaged station is analytically derived. Also, a detailed process is introduced for calculating the 1-percent plus flow for an ungaged site near a gaged station in Georgia as an example and a case study is performed. This study provides engineers and practitioners with a method that helps them better assess flood risks and develop mitigation plans accordingly.

Keywords: Uncertainty analysis, 100-year flow, Regression equation, Flood, Risk assessment, Floodplain

☆NOTICE: This is the author's version of a work that was accepted for publication in the Journal of Hydrology. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in the Journal of Hydrology 539, 640–647 (August 2016), doi:10.1016/j.jhydrol.2016.05.070.

CITATION: Cho, H., Bones, E., 2016. Quantification of uncertainties in the 100-year flow at an ungaged site near a gaged station and its application in Georgia. Journal of Hydrology 539, 640–647.

*Corresponding author.
Email address: hcho@isnew.info (Huidae Cho)

1. Introduction

This study introduces a new uncertainty concept in flood insurance studies by the Federal Emergency Management Agency (FEMA), investigates how uncertainties in flow prediction are propagated through the flow weighting method developed by U.S. Geological Survey (USGS), and derives an analytical solution that quantifies those uncertainties.

The National Flood Insurance Program (NFIP) was created by the United States Congress through the passage of the National Flood Insurance Act of 1968 (FEMA, 2002). The Act was created in response to Hurricane Betsy which caused over one billion dollars in damage to the Gulf States. It was the first hurricane to cause damages in excess of one billion dollars and received the nickname "Billion Dollar Betsy" (Holladay and Schwartz, 2010). The devastation caused by poorly communicated risk can also be noted in recent flooding in South Carolina in October 2015, where 17 people were killed by flood waters (CNBC Weather, 2015), and in Missouri in January 2016, where 7100 buildings were affected by flooding and 25 people in Illinois and Missouri were killed by flood waters (ABC News, 2016). Currently, the NFIP insures over 5.5 million properties, affecting over 10,000 communities (Holladay and Schwartz, 2010), and since insurance payouts became widespread in 1978, it has provided 51.7 billion dollars in funds for homeowners (FEMA, 2015).

As the NFIP requires the purchase of flood insurance by property owners, a standard had to be established "to be used as the basis for risk assessment, insurance rating, and floodplain management" (FEMA, 2002). Based on these criteria, the "1-percent annual chance flood" (i.e., 100-year or 1% flood) was recommended for use as the NFIP standard (FEMA, 2002). The Department of Housing and Urban Development (HUD) initially oversaw the mapping of the 1% floodplain in flood-prone communities until 1979 when the responsibilities of the NFIP were taken over by FEMA. Since the inception of the NFIP in 1968, the 1% floodplain has been used to communicate the extent of the risk associated with flooding. However, there are many uncertainties associated with predicting the 1% floodplain, and when those uncertainties are accounted for, flooding risk has the potential to expand linearly or non-linearly depending on the channel geometry (Jung and Merwade, 2015). Since uncertainty analysis for floodplain mapping provides more resilient and reliable information for flood risk management (Ntelekos et al., 2006; Xu and Booij, 2007; Jung and Merwade, 2015) and, as computer models and topography have greatly improved since 1968, FEMA has realized the importance of evaluating those uncertainties in order to communicate the potential risk to communities. To facilitate the communication of risk, FEMA has added the "1-percent plus" flood elevation to the Risk Mapping, Assessment, and Planning (Risk MAP) flood risk products for all riverine analyses (FEMA, 2011, 2013). FEMA (2013) defines the "1-percent plus" flood elevation as follows:

> The 1% plus flood elevation is defined as a flood elevation derived by using discharges that include the average predictive error for the regression equation discharge calculation for the Flood Risk Project. This error is then added to the 1% annual chance discharge to calculate the new 1%

plus discharge. The upper 84-percent confidence limit is calculated for Gage and rainfall-runoff models for the 1% annual chance event.

Statistically, the logarithmic 1% plus flow is one standard error of prediction away from the mean logarithmic estimate of the 1% flow, which is equivalent to the upper 84% prediction limit in a one-tailed test. Modelers should derive the 1% plus flood elevation by using the 1% plus flow, which indicates how uncertain the estimate of the 1% flow is. Although the benefits of including uncertainty analyses in flood risk assessments have been discussed previously in the literature (Ntelekos et al., 2006; Xu and Booij, 2007; Jung and Merwade, 2012, 2015), it is not well established how to obtain the 1% plus flow for estimating the corresponding flood elevation. Ames (2006) proposed a bootstrap approach for obtaining confidence limits of low flow from data and the same approach could be applied to flood studies if there were enough peak flow data from which a large number of new "realizations" of peak flow data could be generated. However this resampling technique is not applicable when the area of interest is not gaged and no records of peak flows are available. Other than this similar work of Ames (2006) that addresses uncertainties in low flow estimates intead of peak flow estimates, the literature review has revealed no instructions on how to compute the 1% plus flow to the best of the authors' knowledge. Therefore, with no guidance for modelers, they are unable to properly communicate the flood risk to communities, endangering the persons and properties in flood-prone areas.

In order to accurately calculate the 1% plus flows and properly communicate risk to those affected by floods, this study derives an equation for the 1% plus flow for regression analyses weighted with historical gage records using the error propagation method (Birge, 1939; Ku, 1966; Tellinghuisen, 2001). Additionally, a case study is presented that develops the 1% plus flow based on the method discussed in this study. The floodplains for the 1% and 1% plus flows are delineated and compared to demonstrate how the 1% plus floodplain can be an effective tool to communicate risk to communities and their residents.

2. Background

2.1. 1-percent flow

For FEMA flood studies, modelers often use USGS Scientific Investigations Reports (SIR) for estimating the magnitude and frequency of floods for ungaged watersheds. USGS publishes regional flood-frequency equations for different exceedance probabilities including the 1% chance flow for different states. For example, Gotvald et al. (2009) derived the regional flood-frequency equations for rural ungaged streams in Georgia. USGS takes the logarithm of peak flows and performs a regression analysis using various watershed parameters including the drainage area, slope, percents of the watershed falling in different hydrologic regions, etc. Using the regression equations derived in this way, hydrologic modelers can estimate flows at ungaged sites for a 1% chance of exceedance. A typical equation for such log-linear regression analyses is as follows:

$$\log Q_P = \log K + \sum_{i=1}^{n} A_i \log X_i \tag{1}$$

where $\log(\cdot)$ is the logarithm function of base 10, Q_P is the flow of a P% probability, K and A_i's are regression coefficients, and X_i's are independent variables describing the watershed. One can obtain the final equation for Q_P by taking the exponential of both sides of Eq. (1) after a log-linear regression analysis as follows:

$$Q_P = K \prod_{i=1}^{n} X_i^{A_i}. \tag{2}$$

Eq. (2) can be used to estimate the P% flow at ungaged sites. USGS computes a $100(1-\alpha)$% prediction interval for the true peak flow at an ungaged site using the following inequation:

$$Q_P/C < Q_P < CQ_P \tag{3}$$

where C is defined as

$$C = 10^{t_{(\alpha/2,n-p)} S_{p,i}} \tag{4}$$

where $t_{(\alpha/2,n-p)}$ is the critical value of the Student's t-distribution at an α level and degrees of freedom $n-p$, where n is the number of observations used for the regression analysis and p is the number of regression variables; and $S_{p,i}$ is the standard error of prediction for site i. Gotvald et al. (2009) defines $S_{p,i}$ as follows:

$$S_{p,i} = \left[\gamma^2 + \mathbf{x}_i \mathbf{U} \mathbf{x}_i^T\right]^{0.5} \tag{5}$$

where γ^2 is the model error variance, $\mathbf{x}_i$ is a row vector of a 1 as the first element followed by regression equation parameter values for site i, $\mathbf{U}$ is the covariance matrix for the regression coefficients, and $\mathbf{x}_i^T$ is the transpose of $\mathbf{x}_i$.

When there is a streamflow gage at the outlet of the study watershed, one can analyze historical records of annual peak flows to estimate the P% flow without having to use the regression equation. Assuming that annual peak flows follow the log-Pearson Type III distribution as recommended by the Interagency Advisory Committee on Water Data (1982), one can fit a distribution curve to historical annual peak flows by adjusting its statistical parameters. The PeakFQ program (Flynn et al., 2006) implements this parameter estimation procedure, and calculates flows and confidence intervals for different probabilities. The user can obtain the upper one-tailed 84% prediction limit of the 1% flow by changing the confidence intervals parameter to 0.84 (i.e., 84%). This upper 84% prediction limit is located one standard prediction error away from the mean estimate of the 1% flow and corresponds to the definition of the 1% plus flow.

2.2. Weighted flow estimate for an ungaged site near a gaged station

Since the area of interest is most likely ungaged, the regression analysis is needed to estimate peak flows. However, if there is a gaged station near the ungaged site of interest, it is recommended by the Interagency Advisory Committee on Water Data (1982) to incorporate the gage data into the regression analysis to produce results that more closely resemble the real-world data. According to Gotvald et al. (2009), a gaged station is considered near an ungaged site if the drainage area of the ungaged site is within the range of 50% to 150% of the drainage area of the gaged station. To properly incorporate gaged data into flow estimates at nearby ungaged sites, Gotvald et al. (2009) combined two estimates of the peak flow from the gaged station and ungaged site by weighting both flows with the drainage area difference between the two locations, and defined the weighted estimate of the peak flow for a P% chance exceedance at the ungaged site as follows:

$$Q_{P(u)w} = \left[\frac{2\Delta A}{A_{(g)}} + \left(1 - \frac{2\Delta A}{A_{(g)}}\right)\frac{Q_{P(g)w}}{Q_{P(g)r}}\right] Q_{P(u)r} \tag{6}$$

where ΔA is the absolute value of the difference between the drainage areas of the gaged station and ungaged site, $A_{(g)}$ is the drainage area for the gaged station, $Q_{P(g)w}$ is the weighted estimate of the peak flow for the P% chance exceedance for the gaged station, and $Q_{P(g)r}$ and $Q_{P(u)r}$ are the peak-flow estimates for the P% chance exceedance at the gaged station and ungaged site, respectively, calculated using the applicable regional regression equations.

Gotvald et al. (2009) also combined two estimates of the peak flow from the regression analysis and gage analysis at the gaged station using the variances of prediction of both methods, and defined the logarithm of the weighted estimate of the peak flow at the gaged station using the following equation:

$$\log Q_{P(g)w} = \frac{V_{p,P(g)r} \log Q_{P(g)s} + V_{p,P(g)s} \log Q_{P(g)r}}{V_{p,P(g)s} + V_{p,P(g)r}} \tag{7}$$

where $Q_{P(g)s}$ is the estimate of the peak flow at the gaged station from the log-Pearson Type III analysis for the P% chance exceedance, and $V_{p,P(g)r}$ and $V_{p,P(g)s}$ are the variances of prediction for the P% chance exceedance in log units at the gaged station derived from the applicable regional regression equations and the log-Pearson Type III analysis, respectively.

In order to solve for $Q_{P(u)w}$, which will provide the weighted flow at the area of interest, Eqs. (6) and (7) are combined to obtain the following equation:

$$
\begin{aligned}
Q_{P(u)w} &= \left[\frac{2\Delta A}{A_{(g)}} + \left(1 - \frac{2\Delta A}{A_{(g)}}\right) \frac{10^{\frac{V_{p,P(g)r}\log Q_{P(g)s} + V_{p,P(g)s}\log Q_{P(g)r}}{V_{p,P(g)s}+V_{p,P(g)r}}}}{Q_{P(g)r}}\right] Q_{P(u)r} \\
&= \left[\frac{2\Delta A}{A_{(g)}} + \left(1 - \frac{2\Delta A}{A_{(g)}}\right) \frac{10^{\log Q_{P(g)s}\frac{V_{p,P(g)r}}{V_{p,P(g)s}+V_{p,P(g)r}}}\, 10^{\log Q_{P(g)r}\frac{V_{p,P(g)s}}{V_{p,P(g)s}+V_{p,P(g)r}}}}{Q_{P(g)r}}\right] Q_{P(u)r} \\
&= \left[\frac{2\Delta A}{A_{(g)}} + \left(1 - \frac{2\Delta A}{A_{(g)}}\right) \frac{Q_{P(g)s}^{\frac{V_{p,P(g)r}}{V_{p,P(g)s}+V_{p,P(g)r}}}\, Q_{P(g)r}^{\frac{V_{p,P(g)s}}{V_{p,P(g)s}+V_{p,P(g)r}}}}{Q_{P(g)r}}\right] Q_{P(u)r} \\
&= \left[\frac{2\Delta A}{A_{(g)}} + \left(1 - \frac{2\Delta A}{A_{(g)}}\right) Q_{P(g)s}^{\frac{V_{p,P(g)r}}{V_{p,P(g)s}+V_{p,P(g)r}}}\, Q_{P(g)r}^{\frac{V_{p,P(g)s}}{V_{p,P(g)s}+V_{p,P(g)r}}-1}\right] Q_{P(u)r} \\
&= \left[\frac{2\Delta A}{A_{(g)}} + \left(1 - \frac{2\Delta A}{A_{(g)}}\right) Q_{P(g)s}^{\frac{V_{p,P(g)r}}{V_{p,P(g)s}+V_{p,P(g)r}}}\, Q_{P(g)r}^{-\frac{V_{p,P(g)r}}{V_{p,P(g)s}+V_{p,P(g)r}}}\right] Q_{P(u)r} \\
&= \left(a + bQ_{P(g)s}^{c} Q_{P(g)r}^{-c}\right) Q_{P(u)r}
\end{aligned}
\tag{8}
$$

where $a = \frac{2\Delta A}{A_{(g)}}$, $b = 1 - a$, and $c = \frac{V_{p,P(g)r}}{V_{p,P(g)s}+V_{p,P(g)r}}$. By letting $Q'_{P(u)w} = \log Q_{P(u)w}$, $Q'_{P(g)s} = \log Q_{P(g)s}$, $Q'_{P(g)r} = \log Q_{P(g)r}$, and $Q'_{P(u)r} = \log Q_{P(u)r}$, and, by taking the logarithm of both sides of Eq. (8), the following equation is obtained:

$$
\begin{aligned}
Q'_{P(u)w} &= \log\left[a + b\left(\frac{10^{Q'_{P(g)s}}}{10^{Q'_{P(g)r}}}\right)^{c}\right] + Q'_{P(u)r} \\
&= \log\left[a + b\cdot 10^{c\left(Q'_{P(g)s} - Q'_{P(g)r}\right)}\right] + Q'_{P(u)r}.
\end{aligned}
\tag{9}
$$

3. Derivation of the 1-percent plus flow

By using the equations provided by USGS SIR reports, the 1% flow can be determined for an ungaged site located near a gaged station. However, no direction is provided for how to determine the predictive uncertainties associated with the 1% flow. As explained earlier, the predictive uncertainties are necessary to determine the 1% plus flow as required by FEMA. This section analytically derives an equation to compute the 1% plus flow at ungaged sites near a gaged station using the propagation of errors (Birge, 1939).

Assuming that $Q'_{P(g)s}$ (the log-Pearson Type III analysis), $Q'_{P(g)r}$, and $Q'_{P(u)r}$ (the logarithm of two regression flows at the gaged station and ungaged site, respectively) are uncorrelated, cross terms in the propagated standard deviation equation (Birge, 1939; Ku, 1966; Tellinghuisen, 2001) are canceled out and the following equation is obtained:

$$\sigma'_{P(u)w} = \left[\left(\frac{\partial Q'_{P(u)w}}{\partial Q'_{P(u)r}} \right)^2 {\sigma'_{P(u)r}}^2 + \left(\frac{\partial Q'_{P(u)w}}{\partial Q'_{P(g)s}} \right)^2 {\sigma'_{P(g)s}}^2 + \left(\frac{\partial Q'_{P(u)w}}{\partial Q'_{P(g)r}} \right)^2 {\sigma'_{P(g)r}}^2 \right]^{0.5} \tag{10}$$

where $\sigma'_{P(u)w}$, $\sigma'_{P(u)r}$, $\sigma'_{P(g)s}$, and $\sigma'_{P(g)r}$ are the standard deviations of $Q'_{P(u)w}$, $Q'_{P(u)r}$, $Q'_{P(g)s}$, and $Q'_{P(g)r}$, respectively.

The standard errors of $Q'_{P(u)r}$ and $Q'_{P(g)r}$ can be defined using the equation for the standard error of prediction in Gotvald et al. (2009) as follows, respectively:

$$\sigma'_{P(u)r} = S_{P(u)r} \tag{11}$$

and

$$\sigma'_{P(g)r} = S_{P(g)r} \tag{12}$$

where $S_{P(u)r}$ and $S_{P(g)r}$ are the standard errors of prediction for the ungaged site and gaged station, respectively, for the P% chance exceedance, and can be calculated using Eq. (5).

The output from PeakFQ can be used, once the confidence interval has been adjusted to 0.84 (i.e., 84%) in the "Output Options" tab, to calculate the standard error of $Q'_{P(g)s}$ as follows:

$$\sigma'_{P(g)s} = \log Q^{84\%UL}_{P(g)s} - Q'_{P(g)s} = \log \frac{Q^{84\%UL}_{P(g)s}}{Q_{P(g)s}} \tag{13}$$

where $Q^{84\%UL}_{P(g)s}$ is the 84% upper confidence limit of $Q_{P(g)s}$ from the PeakFQ output.

Further, the partial derivatives $\frac{\partial Q'_{P(u)w}}{\partial Q'_{P(u)r}}$, $\frac{\partial Q'_{P(u)w}}{\partial Q'_{P(g)s}}$, and $\frac{\partial Q'_{P(u)w}}{\partial Q'_{P(g)r}}$ can be expanded as follows, respectively:

$$\frac{\partial Q'_{P(u)w}}{\partial Q'_{P(u)r}} = \frac{\partial \left[\log \left(a + b \cdot 10^{c\left(Q'_{P(g)s} - Q'_{P(g)r}\right)} \right) + Q'_{P(u)r} \right]}{\partial Q'_{P(u)r}} = 1, \tag{14}$$

$$
\begin{aligned}
\frac{\partial Q'_{P(u)w}}{\partial Q'_{P(g)s}} &= \frac{\partial \left[\log\left(a + b\cdot 10^{c\left(Q'_{P(g)s} - Q'_{P(g)r}\right)}\right) + Q'_{P(u)r}\right]}{\partial Q'_{P(g)s}} \\
&= \frac{\partial \log\left(a + b\cdot 10^{c\left(Q'_{P(g)s} - Q'_{P(g)r}\right)}\right)}{\partial\left(a + b\cdot 10^{c\left(Q'_{P(g)s} - Q'_{P(g)r}\right)}\right)} \frac{\partial\left(a + b\cdot 10^{c\left(Q'_{P(g)s} - Q'_{P(g)r}\right)}\right)}{\partial\left[c\left(Q'_{P(g)s} - Q'_{P(g)r}\right)\right]} \\
&\quad \times \frac{\partial\left[c\left(Q'_{P(g)s} - Q'_{P(g)r}\right)\right]}{\partial Q'_{P(g)s}} \\
&= \frac{bc\cdot 10^{c\left(Q'_{P(g)s} - Q'_{P(g)r}\right)}}{a + b\cdot 10^{c\left(Q'_{P(g)s} - Q'_{P(g)r}\right)}} = \frac{bc\left(Q_{P(g)s}Q^{-1}_{P(g)r}\right)^c}{a + b\left(Q_{P(g)s}Q^{-1}_{P(g)r}\right)^c},
\end{aligned}
\tag{15}
$$

and

$$
\begin{aligned}
\frac{\partial Q'_{P(u)w}}{\partial Q'_{P(g)r}} &= \frac{\partial \left[\log\left(a + b\cdot 10^{c\left(Q'_{P(g)s} - Q'_{P(g)r}\right)}\right) + Q'_{P(u)r}\right]}{\partial Q'_{P(g)r}} \\
&= \frac{\partial \log\left(a + b\cdot 10^{c\left(Q'_{P(g)s} - Q'_{P(g)r}\right)}\right)}{\partial\left(a + b\cdot 10^{c\left(Q'_{P(g)s} - Q'_{P(g)r}\right)}\right)} \frac{\partial\left(a + b\cdot 10^{c\left(Q'_{P(g)s} - Q'_{P(g)r}\right)}\right)}{\partial\left[c\left(Q'_{P(g)s} - Q'_{P(g)r}\right)\right]} \\
&\quad \times \frac{\partial\left[c\left(Q'_{P(g)s} - Q'_{P(g)r}\right)\right]}{\partial Q'_{P(g)r}} \\
&= \frac{-bc\cdot 10^{c\left(Q'_{P(g)s} - Q'_{P(g)r}\right)}}{a + b\cdot 10^{c\left(Q'_{P(g)s} - Q'_{P(g)r}\right)}} = \frac{-bc\left(Q_{P(g)s}Q^{-1}_{P(g)r}\right)^c}{a + b\left(Q_{P(g)s}Q^{-1}_{P(g)r}\right)^c}.
\end{aligned}
\tag{16}
$$

By letting $d = \frac{bc\left(Q_{P(g)s}Q^{-1}_{P(g)r}\right)^c}{a+b\left(Q_{P(g)s}Q^{-1}_{P(g)r}\right)^c}$ and plugging Eqs. (11)–(16) into Eq. (10), the equation for the standard deviation of $Q'_{P(u)w}$ is obtained as follows:

$$
\sigma'_{P(u)w} = \left\{\left(S_{P(u)r}\right)^2 + d^2\left[\left(\log\frac{Q^{84\%UL}_{P(g)s}}{Q_{P(g)s}}\right)^2 + \left(S_{P(g)r}\right)^2\right]\right\}^{0.5}. \tag{17}
$$

The logarithmic 1% plus flow is defined as the upper 84% confidence bound of the logarithmic 1% flow in a one-tailed test. When there is a gage station near an ungaged site of interest, Eq. (6) should be used with $P = 1\%$ to compute the weighted flow estimate for the 1% flow. In this case, the probability of the true logarithmic 1% flow being less than the upper confidence bound in a one-tailed test has to be 84% (i.e., one standard deviation away from the logarithmic mean). That is,

$$
\log Q_{1\%+(u)w} = \log Q_{1\%(u)w} + \sigma'_{1\%(u)w} \tag{18}
$$

where $Q_{1\%+(u)w}$ is the 1% plus flow, $Q_{1\%(u)w}$ is the 1% flow, and $\sigma'_{1\%(u)w}$ is the standard deviation of $\log Q_{1\%(u)w}$. From Eq. (18), the following equation for the 1% plus flow is obtaind:

$$Q_{1\%+(u)w} = 10^{\sigma'_{1\%(u)w}} Q_{1\%(u)w}. \tag{19}$$

4. Application

To provide an example for the 1% plus flow calculations given above, the Chestatee River in Lumpkin County, Georgia, shown in Figure 1 is used as a case study. The Chestatee River is located in a rural, mountainous area of north Georgia and its drainage area ranges from 88 to 622 km^2 within Lumpkin County. There are total 73 drainage basins in this study area including one USGS gage station and 72 ungaged sites. The USGS gage station 02333500 is located on the Chestatee River in Dahlonega, Georgia, and has a drainage area of 396 km^2. Therefore, according to the Interagency Advisory Committee on Water Data (1982), drainage areas along the Chestatee River ranging from 198 to 594 km^2 (i.e., 50%–150% of 396 km^2) should be adjusted using the gage data. For this case study, 55 drainage basins fit this criterion including the gage station while the other 18 drainage basins are ungaged and not affected by the gage data. In this section, a location with a drainage area of 348 km^2 shown in Figure 1 was examined as sample calculations, but all the 73 drainage basins were analyzed using the same method and their results are shown and discussed in Section 5.

Figure 2 shows the flowchart for calculating the 1% plus flow for the study area. The regression equation for the 1% flow for the study area is defined in Gotvald et al. (2009) in the United States system of units, which is converted to the international system of units as follows:

$$\begin{aligned} Q_{1\%} = {} & 10^{[-1.7935+0.0289(\mathrm{PCT}_1)+0.02711(\mathrm{PCT}_2)+0.01963(\mathrm{PCT}_3)+0.0258(\mathrm{PCT}_4)+0.0286(\mathrm{PCT}_5)]} \\ & \times \mathrm{DA}^{[0.594+0.00119(\mathrm{PCT}_2)+0.00139(\mathrm{PCT}_3)]} \end{aligned} \tag{20}$$

where $Q_{1\%}$ is the 1% flow in m^3/s, DA is the drainage area in km^2, and PCT_1, PCT_2, PCT_3, PCT_4, and PCT_5 are the area percentages in hydrologic regions 1, 2, 3, 4, and 5, respectively. Table 1 shows the parameter values for the 1% flow regression equation for the gage station and ungaged site. The drainage area for the gage station, $A_{(g)}$, is 396 km^2 and the difference in drainage area between the two locations, ΔA, is 48 km^2.

Table 1: Regression parameters for the 1% flow regression equation for the gage station and ungaged site. DA in km^2 and PCT_1–PCT_5 in %.

Location	DA	PCT_1	PCT_2	PCT_3	PCT_4	PCT_5
Gage station	396	69.9	30.1	0	0	0
Ungaged site	348	65.7	34.3	0	0	0

The additional variables required to complete the flowchart in Figure 2 are the average variance of prediction of the 1% regression equation, $V_{p,1\%(g)r} = 0.0305$,

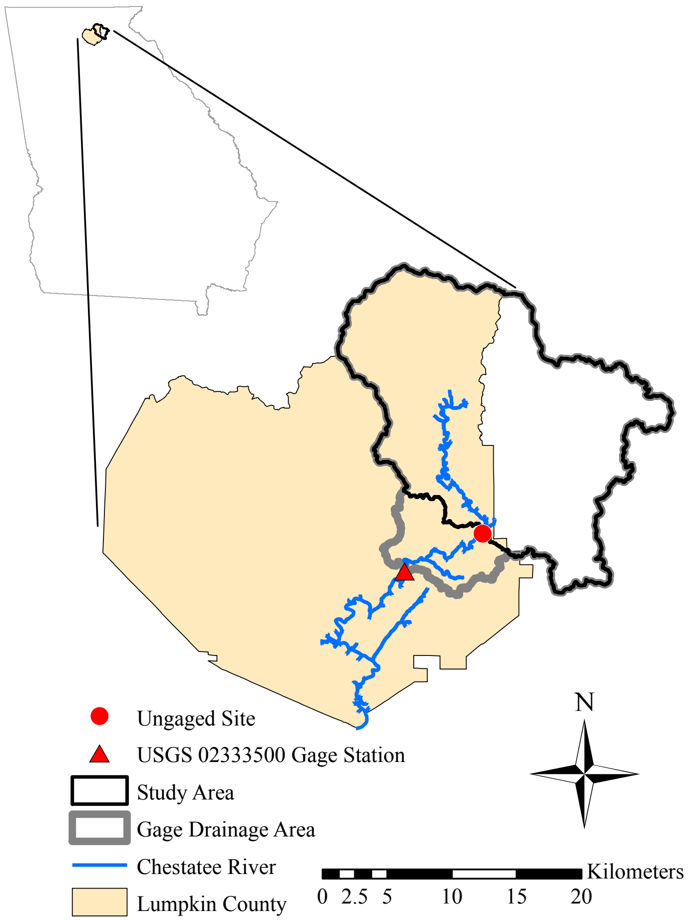

Figure 1: Ungaged site on the Chestatee River in Lumpkin County used for sample calculations.

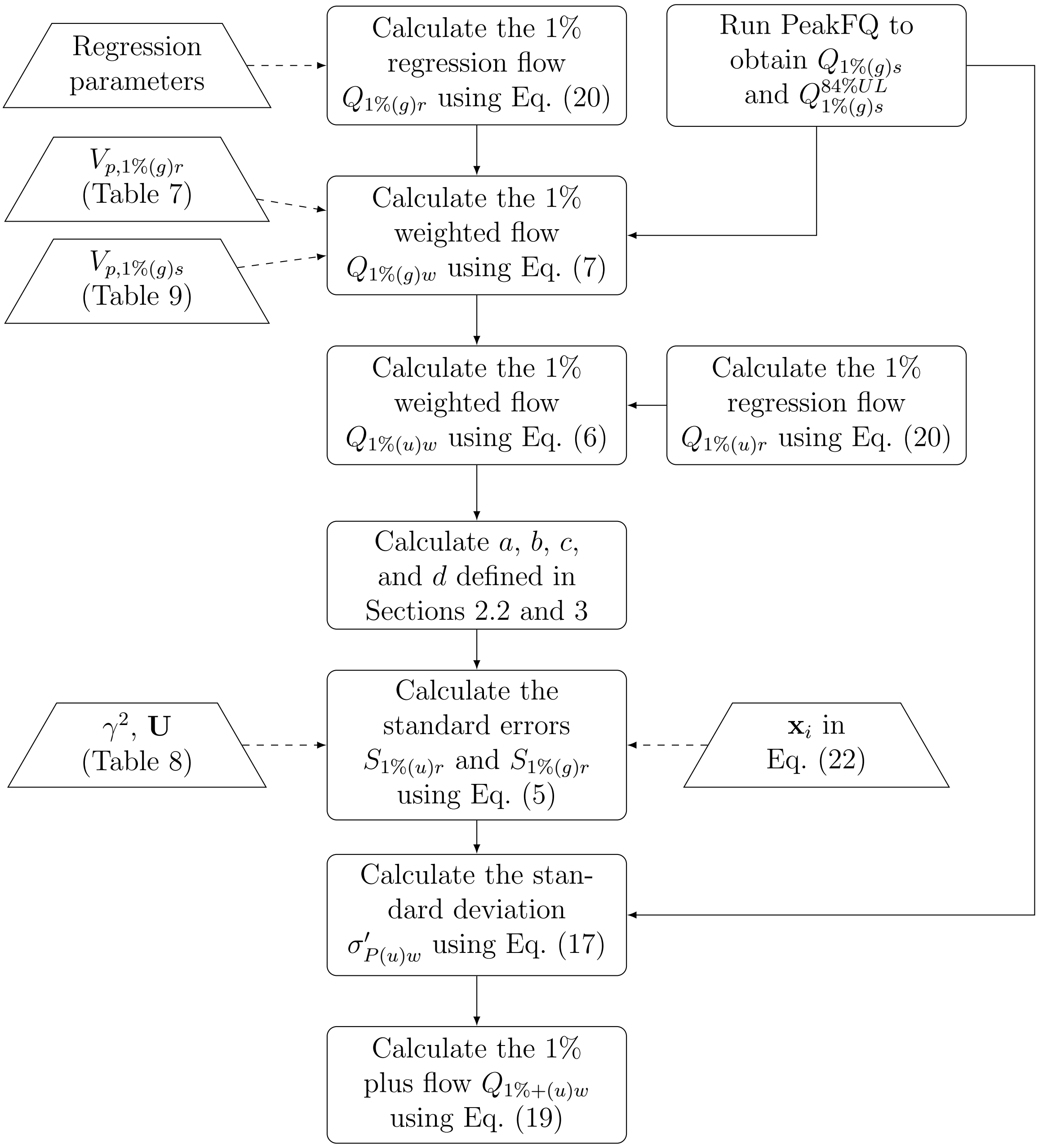

Figure 2: Flowchart for calculating the 1% plus flow for the study area. Table numbers in parentheses refer to Gotvald et al. (2009).

the variance of prediction values for the USGS 02333500 gage station, $V_{p,1\%(g)s} = 0.0038$, the model error variance of the 1% regression equation, $\gamma^2 = 0.0294$, and the covariance matrix for the regression coefficients, $\mathbf{U}$ in Eq. (21), which are all provided by Gotvald et al. (2009) in Tables 7, 9, 8, and 8, respectively.

$$\mathbf{U} = \begin{bmatrix} 1.50\text{e}{-3} & -3.01\text{e}{-4} & -4.79\text{e}{-6} & -9.19\text{e}{-6} & -6.05\text{e}{-6} & -2.51\text{e}{-6} & 2.10\text{e}{-6} & 1.66\text{e}{-6} \\ -3.01\text{e}{-4} & 1.28\text{e}{-4} & -2.00\text{e}{-7} & 1.76\text{e}{-6} & 2.16\text{e}{-6} & 5.74\text{e}{-8} & -1.11\text{e}{-6} & -1.37\text{e}{-6} \\ -4.79\text{e}{-6} & -2.00\text{e}{-7} & 8.49\text{e}{-8} & 8.16\text{e}{-8} & 4.14\text{e}{-8} & 1.86\text{e}{-8} & -4.95\text{e}{-9} & 5.64\text{e}{-9} \\ -9.19\text{e}{-6} & 1.76\text{e}{-6} & 8.16\text{e}{-8} & 4.18\text{e}{-7} & 7.54\text{e}{-8} & 1.95\text{e}{-8} & -1.42\text{e}{-7} & -1.56\text{e}{-8} \\ -6.05\text{e}{-6} & 2.16\text{e}{-6} & 4.14\text{e}{-8} & 7.54\text{e}{-8} & 1.21\text{e}{-6} & 5.40\text{e}{-9} & -2.73\text{e}{-8} & -5.62\text{e}{-7} \\ -2.51\text{e}{-6} & 5.74\text{e}{-8} & 1.86\text{e}{-8} & 1.95\text{e}{-8} & 5.40\text{e}{-9} & 2.85\text{e}{-7} & 4.71\text{e}{-10} & 6.63\text{e}{-9} \\ 2.10\text{e}{-6} & -1.11\text{e}{-6} & -4.95\text{e}{-9} & -1.42\text{e}{-7} & -2.73\text{e}{-8} & 4.71\text{e}{-10} & 7.52\text{e}{-8} & 1.54\text{e}{-8} \\ 1.66\text{e}{-6} & -1.37\text{e}{-6} & 5.64\text{e}{-9} & -1.56\text{e}{-8} & -5.62\text{e}{-7} & 6.63\text{e}{-9} & 1.54\text{e}{-8} & 3.16\text{e}{-7} \end{bmatrix} \tag{21}$$

For the study area, the parameter vector for site i, $\mathbf{x}_i$, is defined as:

$$\begin{aligned} \mathbf{x}_i = [&1,\ \log \mathrm{DA} - 0.4133,\ \mathrm{PCT}_1,\ \mathrm{PCT}_2,\ \mathrm{PCT}_3,\ \mathrm{PCT}_5, \\ &(\log \mathrm{DA} - 0.4133) \cdot \mathrm{PCT}_2,\ (\log \mathrm{DA} - 0.4133) \cdot \mathrm{PCT}_3]. \end{aligned} \tag{22}$$

By performing the calculations shown in Figure 2 at each applicable ungaged site in the watershed, a series of flows was generated for the 1% plus flow along the Chestatee River whose flows are weighted by the nearby gage station. These flows were then taken into the Hydrologic Engineering Center's River Analysis System (HEC-RAS) (Brunner, 2010) to generate the 1% plus flood elevations, which were used to delineate the floodplain for the 1% plus flow using ArcGIS (Esri, 2011).

5. Results and discussion

The parameter vectors for the ungaged site and gage station are $\mathbf{x}_{(u)r} = [1, 2.128, 65.7, 34.3, 0, 0, 73.03, 0]$ and $\mathbf{x}_{(g)r} = [1, 2.18, 69.9, 30.1, 0, 0, 65.76, 0]$, respectively. Also, the PeakFQ program calculates the estimated peak flow from the log-Pearson Type III analysis $Q_{1\%(g)s} = 670\,\mathrm{m^3/s}$ and its 84% upper confidence limit $Q_{1\%(g)s}^{84\%UL} = 765\,\mathrm{m^3/s}$. Tables 2 and 3 summarize calculation results of the coefficients, the standard errors of prediction, and the 1% plus flow for the study area.

Table 2: Calculation results of a, b, c, d, and the standard errors of prediction.

a	b	c	d	$S_{1\%(u)r}$	$S_{1\%(g)r}$
0.24	0.76	0.889	0.72	0.1731	0.1731

Table 3: Calculation results of the 1% plus flow in $\mathrm{m^3/s}$ for the study area.

$Q_{1\%(g)r}$	$Q_{1\%(g)w}$	$Q_{1\%(u)r}$	$Q_{1\%(u)w}$	$Q_{1\%+(u)w}$
477	646	445	564	929

Figure 3 provides a comparison of the flow calculated using the regression equation in Eq. (20), the weighting equation in Eq. (6), and the equation for the 1%

plus flow in Eq. (19). The regression equation predicts the lowest flows which are $193\,\mathrm{m}^3/\mathrm{s}$ lower than the gage data at the drainage area of $396\,\mathrm{km}^2$. The weighting equation increased the predicted flows, with the largest difference between the flows from the weighting and regression equations occurring at the gage station where the flows are most heavily weighted towards the gage data. The effect of the gage data decreases as the location of an ungaged site moves away from the gage station and completely vanishes at 50% and 150% of the drainage area of the gage station because the first term in Eq. (6) becomes 1 and the second term is canceled out, yielding $Q_{1\%(u)w} = Q_{1\%(u)r}$. However, Figure 3 does not show ungaged sites exactly at the 50% and 150% locations because there was no need to introduce flow changes at these exact locations for modeling purposes.

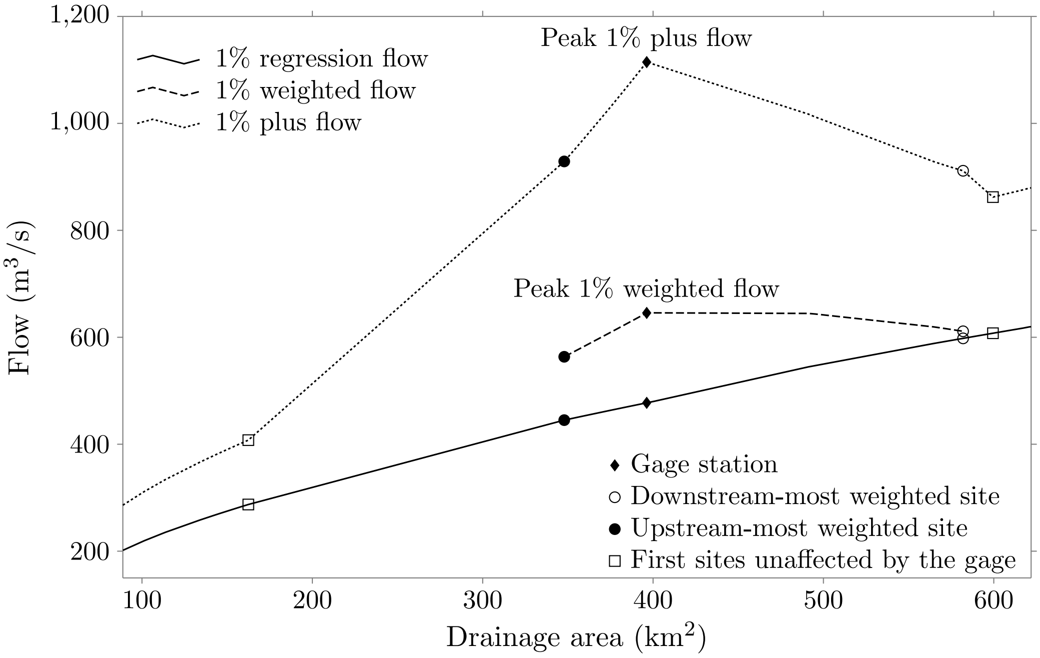

Figure 3: Discharge comparison along the Chestatee River.

The weighting method using Eq. (6) introduced a peak 1% weighted flow at the gage station as shown in Figure 3 because the 1% flow estimated using gage analysis, $Q_{1\%(g)s}$, is much higher than the flow at the same location obtained using the regression equation, $Q_{1\%(g)r}$, such that the weighted flow $Q_{1\%(g)w}$ in Eq. (7) becomes larger than the regression flow at the downstream limit of the weighting method. Because of this peak 1% weighted flow at the gage station, the 1% weighted flow decreases as the drainage area of an ungaged site becomes larger until it reaches 150% of the gage's drainage area, after which no more weighting is applied and the 1% regression flow starts increasing once again. Similarly, if a gage station measured lower flows than those predicted by the regression equation, the 1% weighted flow can generate a valley between 50% and 150% of the gage's drainage area. A peak or valley in the weighted flow causes estimated flows at downstream sites to be lower than those at upstream sites.

The predictive uncertainties of the 1% flows range from 49% at the downstream-most weighted site ($583\,\mathrm{km}^2$) to 73% at the gage station (the peak 1% plus flow at

396 km^2 shown in Figure 3), to 65% at the upstream-most weighted site (348 km^2). The ungaged sites affected by the gage data have a higher uncertainty when compared to those sites not affected. This phenomenon seems to be contradictory because uncertainties should decrease as more data are collected and used for prediction. In this case, the ungaged sites within the effect of the gage station use regional flow data used for developing the regression equation and flow data measured at the gage station while the ungaged sites farther from the gage station only use the regional flow data. However, these contradictory observations can be explained by the weighting process of the regression flow and gage analysis. As the difference between the flows measured at the gage station and the flows predicted using the regression equation increases, so does uncertainty in the weighting process of the two flows because the expected flow becomes more uncertain when the two flows diverge. Combining the two flows does not necessarily reduce uncertainty, but can even increase it because of uncertainties in the weighting method itself. This uncertainty is propagated to the 1% plus flow and is further aggravated by structural errors in the flow estimation and gage analysis methods, and uncertainties in gage data. However, if only one method of flow estimation was used without any weighting, collecting more data could help reduce predictive uncertainties.

Figure 4 shows a sample of the 1% and 1% plus floodplains, which were mapped using the 1% weighted flows and the 1% plus flows, respectively. Flood elevations were taken from the output of the HEC-RAS model and the Triangulated Irregular Network (TIN) of the water surface elevation was created by linear interpolation. The ground elevation TIN was then subtracted from the water surface elevation TIN to find intersecting polygons between the two TINs where the water depth is 0. These intersecting polygons represent a floodplain. This procedure was repeated for both of the 1% weighted and 1% plus flows to map the two floodplains.

Despite the large uncertainty range of the 1% flows from 49% to 73%, the area difference between the two floodplains is only 13% between the downstream-most and upstream-most weighted sites. Jung and Merwade (2015) previously explained this phenomenon as an impact of the channel geometry on the change in flooding area caused by flow changes. At the location shown in Figure 4, water surface elevations varied by 1.8 m on average between the 1% and 1% plus flood elevations, which is a 30% increase of water depth, yet the change in floodplain area was 11% because of the steepness of the channel geometry. The 11% increase in floodplain area has different impacts on the three buildings shown in Figure 4. Building 1 was outside the 1% floodplain, but was flooded by the 1% plus flow whereas building 2 was not affected by the increased flood elevation and building 3 was always inside both floodplains. In a location with less mountainous terrain and larger floodplains, the 1% plus floodplain could have much larger impacts.

In this study, an equation for calculating the uncertainties associated with 1% flow estimation including the flow weighting process, regression equations, and gage analysis was analytically derived. Using the derived equation, a new method was introduced as a flowchart for calculating the 1% plus flow, which is required to derive the 1% plus flood elevation that FEMA has recently added to their Risk MAP flood risk products for all riverine studies. It was shown that uncertainties in the regression equations and gage analysis are propagated and aggravated through

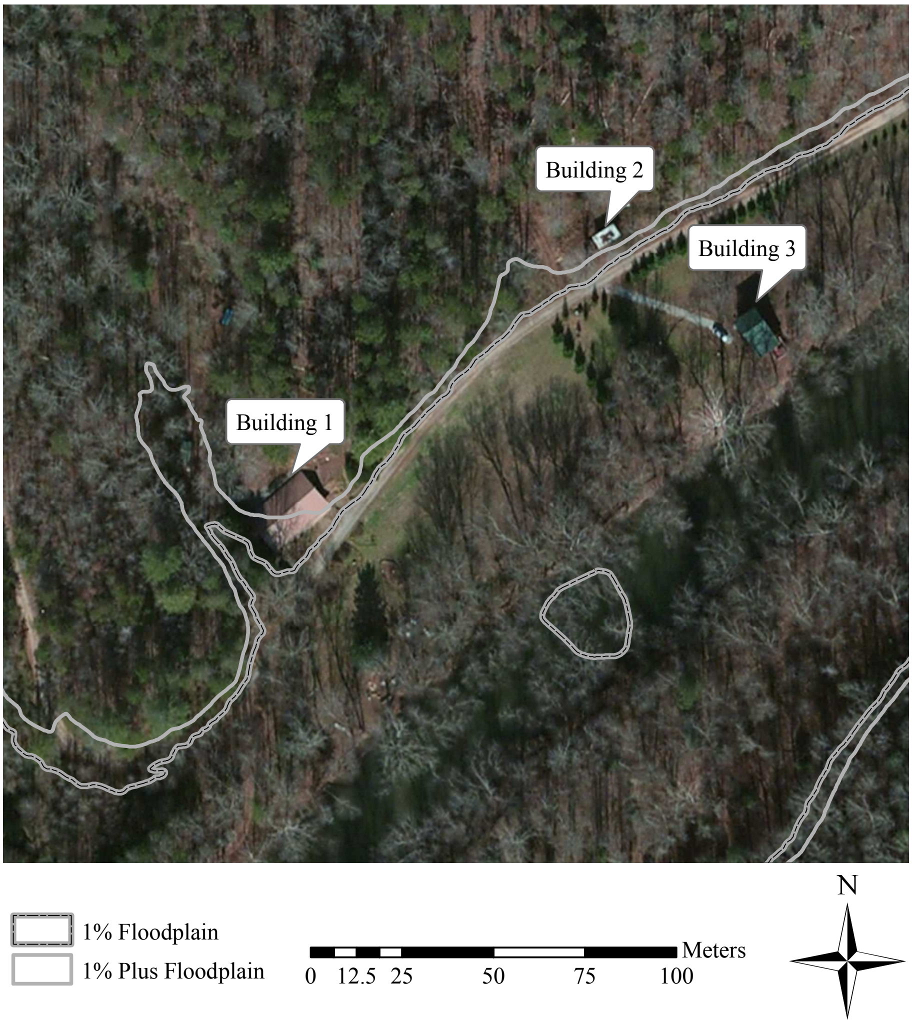

Figure 4: Section of the 1% and 1% plus floodplains along the Chestatee River. Esri (2016) was used to show the imagery basemap.

the flow weighting process, but the impact of increased uncertainties on the 1% floodplain is greatly affected by the geometry of the study area. Although the watershed in the case study is located in north Georgia, the method developed in this study can be used in other states than Georgia when appropriate regional regression equations and variances are used for calculating the 1% plus flow using the flowchart in Figure 2. This method can be a useful tool to quantify uncertainties associated with the flow weighting method introduced by USGS using the regression equation and gage analysis.

With the ability to quantify uncertainty surrounding flood predictions using the proposed method, the 1% plus floodplain can be provided to communities as a tool to help communicate to the public the additional risk caused by the uncertainty of the flow prediction method and collected data. Local governments can employ the method to encourage more property owners vulnerable to serious flood risks outside the traditional 1% floodplain to invest in flood insurance for their homes and businesses, and to identify additional structures that may need to be evacuated in a flood event to prevent loss of life.

In addition to risk communication to the public, communities can better plan for and manage their flood risk by further understanding the uncertainties of the modeling results and floodplains provided to them. Furthermore, government agencies such as FEMA and USGS can identify and prioritize areas with excessive floodplain uncertainties that require improved data collection for better gage analysis and revision of the regional regression equations. For instance, the 1% plus floodplain developed using the proposed method can help FEMA administrators identify highly sensitive communities where small uncertainties can propagate into large changes in the floodplain that could affect many more property owners and residents. FEMA can utilize these studies when investing more funding into more detailed and accurate studies for these highly sensitive communities, creating the most accurate flood map possible in order to efficiently and effectively communicate risk during flood events. The process of quantifying the risk uncertainty and identifying areas in need of improved studies will allow FEMA to spend its funds more effectively on updated studies that will provide the most benefit to the largest number of people.

6. Conclusions

This study discussed the importance of evaluating uncertainties in 1% flow estimation and introduced the 1% plus flow that FEMA has recently added to the Risk MAP flood risk products for all riverine analyses. The concept of the 1% plus flow is a new idea for FEMA that offers the potential for greatly improved risk assessment, study prioritization, and future cost-savings. When a gaged station is available nearby, the 1% flow estimated using regression equations is recommended to be weighted by the gage analysis. Uncertainties in the regression equation and gage analysis propagate to the final 1% flow estimate in this weighting process. An equation for calculating the standard error of prediction in 1% weighted flow estimation was analytically derived and was used to develop a process for calculating the 1% plus flow for a case study. Because of the steep nature of the river geometry, the impact of increased uncertainty on the 1% floodplain was not significant in terms of floodplain area, but the impact of increased water depths on nearby buildings was not necessarily negligible. Since a small increase in uncertainty could have a larger impact on the 1% floodplain in a relatively flat terrain, care needs to be taken when assessing uncertainties in flood studies. The method developed in this study can be used as a tool by communities to analyze and manage their own risk based on the uncertainty of data provided to them by FEMA during the Risk MAP process. As the use of the 1% plus flow becomes more widespread, it can assist communities to better plan for and manage their flood risk by further understanding the uncertainties of the floodplains provided to them. Further, the proposed method can help FEMA and communities to identify and prioritize areas with excessive floodplain uncertainties in order to pinpoint areas that require improved data generation to create more precise floodplains.

Acknowledgements

The authors would like to thank Mathini Sreetharan and Wilbert O. Thomas, Jr. for their review of our early work and fruitful discussions. They do not necessarily agree with our work and the inferences drawn in this study.

References

ABC News, January 2016. Latest: Flooding causing fewer road closures in Missouri. `http://abcnews.go.com/US/wireStory/latest-flooding-causing-fewer-road-closures-missouri-36096516`, accessed in January 2016.

Ames, D. P., 2006. Estimating 7Q10 confidence limits from data: A bootstrap approach. Journal of Water Resources Planning and Management 132 (3), 204–208.

Birge, R. T., 1939. The propagation of errors. The American Physics Teacher 7 (6), 351–357.

Brunner, G. W., January 2010. HEC-RAS River Analysis System User's Manual Version 4.1. US Army Corps of Engineers, Institute for Water Resources, Hydrologic Engineering Center, 609 Second Street, Davis, CA 95616-4687.

CNBC Weather, October 2015. South Carolina's rain and floods, by the numbers. `http://www.cnbc.com/2015/10/08/south-carolinas-rain-and-floods-by-the-numbers.html`, accessed in January 2016.

Esri, 2011. ArcGIS Desktop: Release 10.

Esri, 2016. ArcGIS imagery basemap under license.

FEMA, August 2002. National flood insurance program description. `http://www.fema.gov/media-library-data/20130726-1447-20490-2156/nfipdescrip_1_.pdf`, accessed in October 2015.

FEMA, December 2011. Operating guidance no. 6-11: User guidance for flood risk datasets and products. `http://www.fema.gov/media-library-data/1387818737990-e6470a5e8596a1262699534a13e9c324/Operating+Guidance+6-11-User+Guidance+for+Flood+Risk+Datasets+and+Products+(Dec+2011).pdf`, accessed in October 2015.

FEMA, August 2013. FP 204-078-1: Federal insurance and mitigation administration policy. `http://www.fema.gov/media-library-data/445f109e9cd88ea083891aa196a51167/Policy+Memo+-+Standards+for+Flood+Risk+Projects.pdf`, accessed in September 2015.

FEMA, October 2015. Loss statistics from Jan 1, 1978 through report "AS OF" date below. `http://bsa.nfipstat.fema.gov/reports/1040.htm`, accessed in October 2015.

Flynn, K. M., Kirby, W. H., Hummel, P. R., 2006. User's Manual for Program PeakFQ, Annual Flood-Frequency Analysis Using Bulletin 17B Guidelines: U.S. Geological Survey Techniques and Methods 4-B4, Chapter 4 of Book 4, Section B. U.S. Geological Survey.

Gotvald, A. J., Feaster, T. D., Weaver, J. C., 2009. Magnitude and frequency of rural floods in the southeastern United States, 2006: Volume 1, Georgia. US Geological Survey Scientific Investigations Report 2009-5043, U.S. Department of the Interior, U.S. Geological Survey.

Holladay, J. S., Schwartz, J. A., April 2010. Flooding the market: The distributional consequences of the NFIP. `http://policyintegrity.org/documents/FloodingtheMarket.pdf`, accessed in October 2015.

Interagency Advisory Committee on Water Data, 1982. Guidelines for determining flood flow frequency Bulletin #17B of the Hydrology Subcommittee. Tech. rep., U.S. Department of the Interior, Geological Survey Office of Water Data Coordination, Reston, Virginia 22092.

Jung, Y., Merwade, V., 2012. Uncertainty quantification in flood inundation mapping using generalized likelihood uncertainty estimate and sensitivity analysis. Journal of Hydrologic Engineering 17 (4), 507–520.

Jung, Y., Merwade, V., 2015. Estimation of uncertainty propagation in flood inundation mapping using a 1-D hydraulic model. Hydrological Processes 29, 624–640.

Ku, H. H., 1966. Notes on the use of propagation of error formulas. Journal of Research of the National Bureau of Standards—C. Engineering and Instrumentation 70C (4), 263–273.

Ntelekos, A. A., Georgakakos, K. P., Krajewski, W. F., 2006. On the uncertainties of flash flood guidance: Toward probabilistic forecasting of flash floods. Journal of Hydrometeorology 7, 896–915.

Tellinghuisen, J., 2001. Statistical error propagation. The Journal of Physical Chemistry A 105, 3917–3921.

Xu, Y., Booij, M. J., 2007. Propagation of discharge uncertainty in a flood damage model for the Meuse river. In: Begum, S., Stive, M. J. F., Hall, J. W. (Eds.), Flood Risk Management in Europe. Vol. 25 of Advances in Natural and Technological Hazards Research. Springer Netherlands, Ch. 16, pp. 293–310.