TECHNICAL MEMORANDUM

Algorithm for Stochastic Storm Generation of Watershed Precipitation-Frequency Relationships for Synoptic Scale Mid-Latitude Cyclone and Tropical Storm Remnant Storm Types

MG Schaefer Ph.D. P.E.

1.0 OVERVIEW

This Technical Memorandum provides a brief summary of the algorithm that was developed for production of watershed precipitation-frequency (PF) relationships and uncertainty bounds for synoptic scale Mid-Latitude Cyclone (MLC) and Tropical Storm Remnant (TSR) storm types. These are storms characterized by low to moderate precipitation intensities extending over very large areas that can generate large total precipitation amounts over several days. The 48-hour duration is typically selected as the key duration which is representative of the time period when the majority of precipitation occurs for these synoptic scale storms. This algorithm utilizes stochastic generation methods for simulation of multiple storms per sample year from which the annual maximum watershed precipitation is produced. The roots of this procedure date to the PF study for the American River (Schaefer and Barker[25,26]) where a multivariate approach was first utilized.

This algorithm employs findings from regional point PF analyses and precipitation data from historical storms observed on the watershed and large storms observed in the region that have been transposed to the watershed of interest. The algorithm has numerous components which are discussed in the following sections. The analyses and findings for the 3,525-mi^2 watershed above Cherokee Dam on the Holston River (Figures 1 and 2) is used as a case study to provide context to description of the stochastic storm generation algorithm.

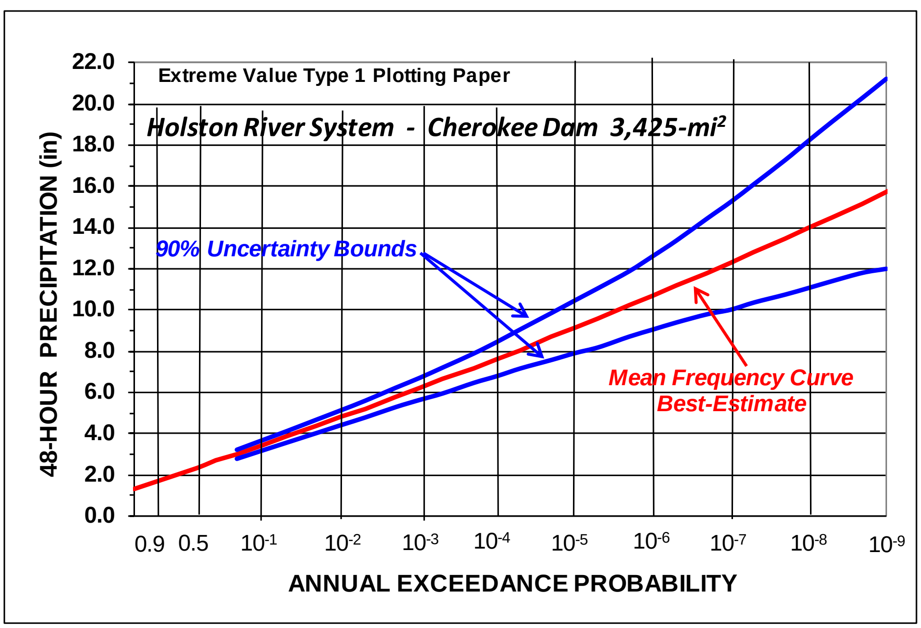

Figure 1 – Watershed Precipitation-Frequency Relationship and 90% Uncertainty Bounds for the Watershed above Cherokee Dam on the Holston River Tennessee

2.0 FRAMEWORK FOR STOCHASTIC STORM GENERATION FOR A WATERSHED

Precipitation-frequency analyses for both point precipitation and watershed precipitation are based on assembling datasets of annual maxima. The stochastic storm generation algorithm is therefore based on the use of annual maxima datasets, specifically the findings of regional PF analyses for point (station) precipitation annual maxima (MGS et al[18]) for development of watershed PF relationships. In particular, the stochastic storm generation algorithm was constructed to emulate synoptic scale storm events and storm characteristics which produce watershed precipitation annual maxima.

2.1 Synoptic Scale Watershed Precipitation

The basic concepts in the stochastic storm generation algorithm can best be described by considering the characteristics of synoptic scale precipitation that produce watershed precipitation annual maxima.

Consider the Cherokee watershed of 3,525-mi^2 located on the Holston River in the Tennessee Valley. Numerous synoptic scale Mid-Latitude Cyclones (MLCs) occur in a given year which affect the watershed where precipitation is measured at stations located in that watershed (Figure 2). One of these storms will produce the watershed 48-hour annual maxima, which is the largest 48-hour areal-average precipitation on the watershed in a given year. This storm may, or may not, produce the 48-hour annual maxima at each of the stations in the network of stations (Figure 2). In particular, it is common that a given storm produces the annual maxima at several stations but not all stations. This behavior requires that multiple storms be stochastically generated each sample year from which the areal-average watershed annual maxima is determined along with the annual maxima at each of the stations in the station network.

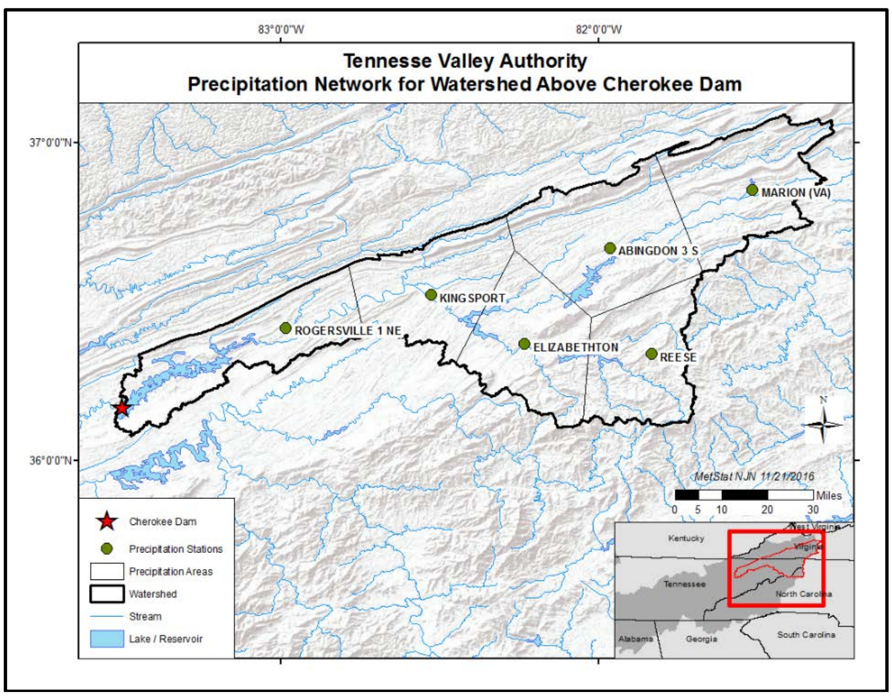

Figure 2 – Network of Six Precipitation Stations Used for Watershed above Cherokee Dam on the Holston River

2.2 Major Components of Stochastic Storm Generation Algorithm

The stochastic storm generation algorithm is constructed to emulate the natural processes for the occurrence of multiple synoptic scale storm events in a year that affect a given watershed. The stochastic storm generation approach has the following components which are described in greater detail in later sections of this TM.

- A key station, centrally located in the watershed, is selected along with a network of stations located throughout the watershed of interest;
- The storm spatial correlation structure for the synoptic scale storm type of interest (MLC or TSR) is determined from cross-correlation analyses of point precipitation from historical storms for the network of stations;
- A point-to-area relationship, similar in form to a Thiessen polygon approach, is developed for estimating areal-average 48-hour watershed precipitation using 48-hour point precipitation for the network of stations and findings of spatial storm analyses of historical storms (MetStorm[17]);
- The key station is used to stochastically generate 48-hour precipitation for multiple storms each sample year based on the findings of the regional point PF analyses for the storm type of interest (MLC or TSR);
- Concurrent 48-hour point precipitation is stochastically generated at each of the stations in the station network using correlation relationships with point precipitation at the key station and the storm spatial correlation developed from historical storms;
- The areal-average 48-hour watershed precipitation is estimated for each storm in a sample year based on the point precipitation at each of the stations in the network and the point-to-area relationship described above;
- The watershed annual maximum is selected as the largest areal-average 48-hour watershed precipitation amongst the storms generated for the given sample year;
- The collection of storms in a given sample year is examined to determine the 48-hour point precipitation annual maxima for each of the stations in the network of stations;
- The watershed PF relationship is derived from the dataset of watershed annual maxima produced by the stochastic storm generation process; and
- A point PF relationship is computed from the dataset of annual maxima for each of the stations in the station network. The stochastically generated point PF relationship for each station is compared with the point PF relationship for that station obtained from the regional PF analysis. The similarity of this comparison is used as one measure to validate the success of the stochastic storm generation algorithm in developing the watershed PF relationship.

2.3 Station Network Example for Cherokee Watershed

As noted above, a network of precipitation stations, located within or very near the watershed of interest, are used for describing the spatial distribution of precipitation within the watershed. Figure 2 depicts the network of six stations used for the watershed above Cherokee Dam on the Holston River. The station at Abingdon 3S, VA was selected as the key station and is used along with substations arrayed about the watershed for describing the spatial distribution of precipitation and for developing a relationship between point precipitation and areal-average precipitation for the watershed. The term substation is used here to distinguish stations in the network other than the key station.

The number of stations required in the network is dependent upon the size of the watershed and may be viewed in the conventional context of a Thiessen polygon network. Parsimony suggests limiting the stations to the number of stations that are adequate for producing reliable point-to-area predictor relationships (Section 5). Experience to-date indicates that two to four stations are usually adequate for watersheds up to 1,000-mi^2, six stations for several thousand square miles and eight to ten stations for watersheds of about 8,000-mi^2. The same network of precipitation stations is used in the stochastic storm generation process for generation of point precipitation and computation of areal-average precipitation for the watershed.

2.4 Data Sources

A variety of precipitation data are required for use in conducting precipitation analyses which include:

- Precipitation annual maxima data series for each station in the study area domain for the storm type of interest for the key duration (48-hours for Cherokee watershed)
- A network of precipitation stations located within and/or very near the watershed of interest that are sufficient in number and spacing to be representative of the spatial distribution of synoptic scale precipitation on the watershed
- Concurrent precipitation for the key duration for each station in the station network for each historical storm that occurred on the watershed or was transposed to the watershed of interest

2.5 Supporting Analyses – Prior Studies

Findings from several point and spatial precipitation analyses for the storm type of interest are needed to support the methods used for stochastic storm generation. The findings from these analyses are typically produced as separate products before analyses are conducted that specifically support the stochastic storm generation algorithm. This includes:

1. Findings from regional point precipitation-frequency analyses (Hosking and Wallis[12], Schaefer[25,27,28,29], MGS et al[18], L-RAP[14]) that includes spatial mapping of the at-site means, regional L-Cv, regional L-Skewness and identification of the regional probability distribution
2. Point PF relationship for precipitation annual maxima series (AMS) data for the key duration at the key station and point PF relationships for all other stations in the station network
3. Point PF relationship for precipitation for the key duration at the key station based on a peaks-over-threshold (POT) analysis for data from storms that occur many times in a given year
4. Findings of spatial storm analyses (MetStorm[17]) for areal-average watershed precipitation for the key duration resulting from storms that occurred on the watershed
5. Results from spatial storm analyses (MetStorm[17]) for areal-average watershed precipitation for the key duration resulting from major storms that have occurred in the climatic region where the watershed is located and have been transposed to the watershed of interest

2.6 Supporting Analyses – Stochastic Storm Generation Algorithm

A number of analyses are needed to specifically support the stochastic storm generation algorithm.

1. Probability-plot of historical watershed-average precipitation annual maxima for the key duration based on findings of spatial analyses of historical storms
2. Predictor equations for watershed-average precipitation developed using both linear multiple regression and a Thiessen polygon approach for precipitation for the key duration for stations in the station network
3. Cross-correlation analyses for all station pairings in the precipitation network for precipitation at the key duration and for transformed precipitation data (Tukey[32], Box and Cox[2]) to describe the spatial correlation structure of synoptic scale storms

2.7 Stochastic Storm Generation Methods

Several probabilistic concepts and Monte Carlo sampling methods are used in the development of a watershed PF relationship and uncertainty bounds using the stochastic storm generation approach. These include:

- Application of the Total Probability Theorem (Benjamin and Cornell[1], Kuczera and Nathan et al[19]) using precipitation at the key duration (48-hours) from the key station as the independent variable
- Stratified sampling of precipitation annual maxima series data (AMS) for the key duration for the key station
- Generation of multiple storms per year with precipitation for the key duration less than the annual maxima at the key station using standard Monte Carlo methods and the point PF relationship from the POT analysis
- Lane's multivariate spatial disaggregation method (Salas et al[22]) for stochastic generation of point precipitation for the key duration at all stations which preserves all cross-correlation relationships between stations in the station network
- Latin-hypercube sampling (McKay et al[16]) of L-moment parameters for the key station to account for epistemic uncertainties in the at-site mean, regional L-Cv, regional L-Skewness and the regional probability distribution
- Bootstrap resampling methods (Efron[4]) to account for uncertainties in the cross-correlation estimates for the station network (storm spatial correlation structure)
- Characterization of uncertainties in alternative multiple regression and Thiessen polygon predictor equations for areal-average watershed precipitation for the key duration

2.8 Stochastic Storm Generation Flow Chart

The sequence of tasks for stochastic storm generation are summarized in the Flow Chart depicted in Figure 3. The inner (blue) loop is used for computing a watershed PF relationship for one plausible scenario of inputs and probabilistic model parameters and typically involves simulation of 10,000 sample years of storms with 3 to 12 storms per sample year. The outside loop is used for computing alternative plausible watershed PF relationships that account for epistemic uncertainties through different scenarios of plausible inputs and probabilistic model parameters. The outer loop typically involves about 200 scenarios to capture the effects of uncertainties which provides for computation of a mean watershed PF relationship and 90% uncertainty bounds. This approach typically results in stochastic generation of about 14 million separate storms.

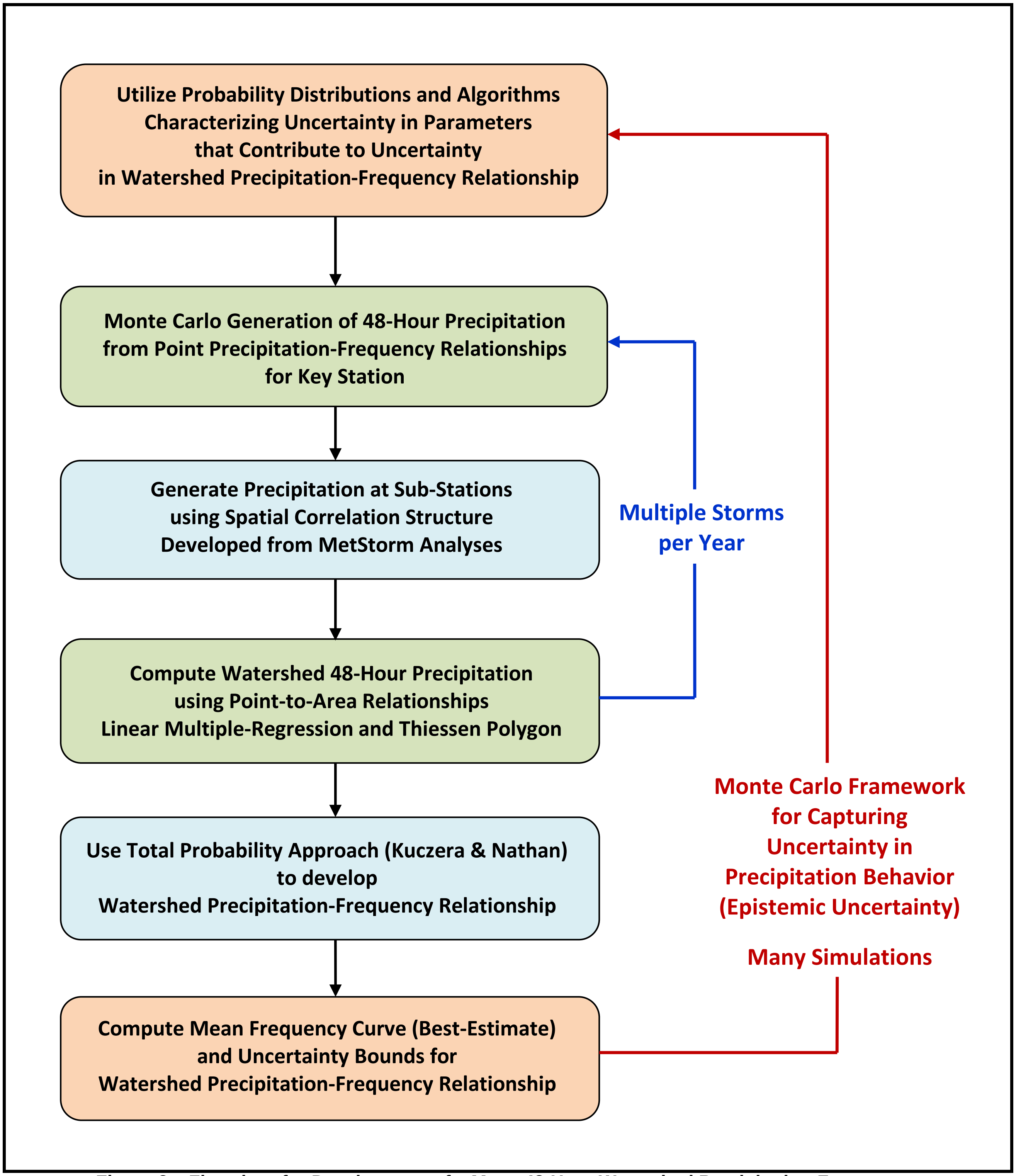

Figure 3 – Flowchart for Development of a Mean 48-Hour Watershed Precipitation-Frequency Relationship and Uncertainty Bounds for Synoptic Scale Mid-Latitude Cyclone and Tropical Storm Remnant Storm Types

3.0 POINT PRECIPITATION-FREQUENCY RELATIONSHIPS FOR THE KEY STATION

As discussed previously, the key station should be centrally located in the watershed to provide a good indicator of synoptic scale precipitation that likely extends over most, or all, of the watershed. Figure 4 depicts the 48-hour point PF relationship for the Abingdon 3S key station based on regional PF analyses conducted for the Tennessee Valley Study Area (MGS et al[18]). Figure 5 depicts the at-site Peaks-Over-Threshold solution for the Abingdon 3S station where a threshold of 0.45-inches was used which corresponds to a storm magnitude that is exceeded about 15 times per year. The point PF relationship in Figure 3 is used for generation of precipitation annual maxima at the key station. The POT PF relationship in Figure 5 is used for generation of multiple storms per sample year at the key station which are smaller in magnitude than the precipitation annual maxima generated from the point PF relationship shown in Figure 4.

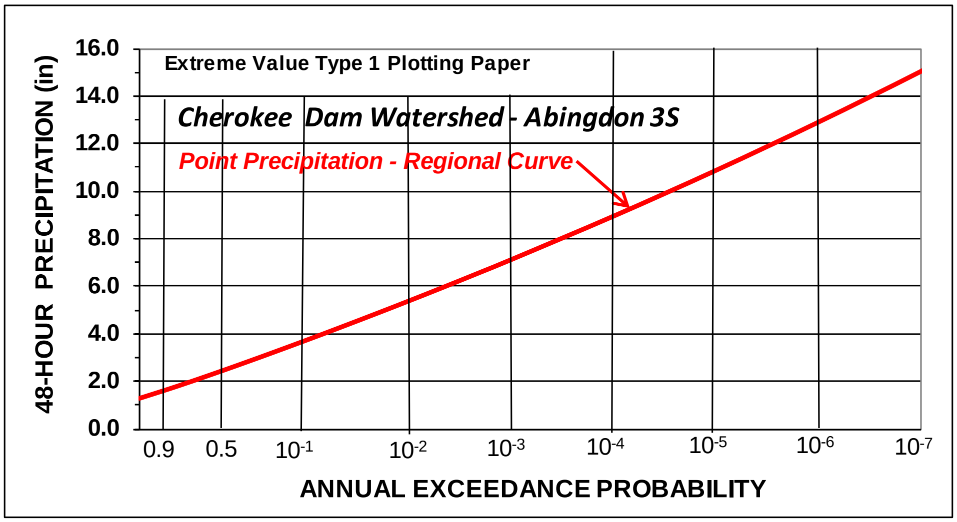

Figure 4 – Point Precipitation-Frequency Relationship for 48-Hour Precipitation Annual Maxima for Abingdon 3S Precipitation Station

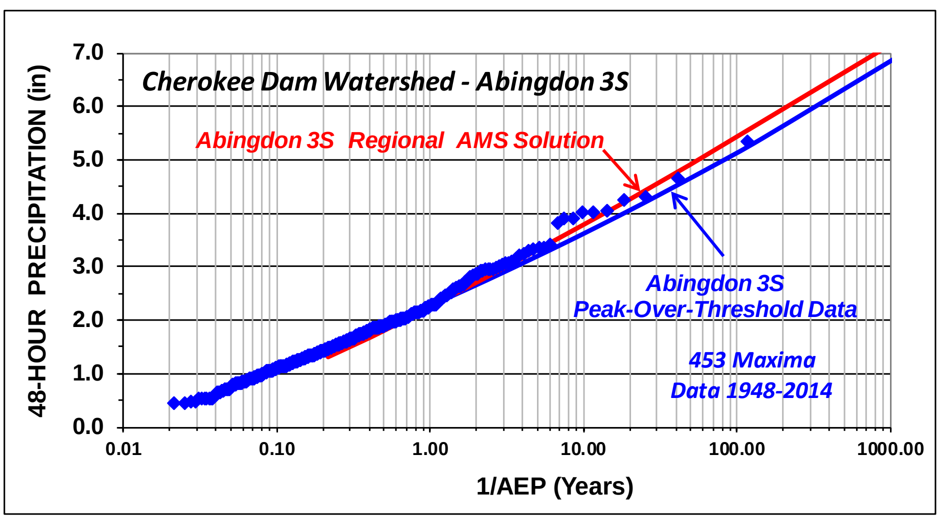

Figure 5 – Point Precipitation-Frequency Relationship for 48-Hour Precipitation Maxima Exceeding 0.45-Inches for Abingdon 3S Precipitation Station

4.0 SPATIAL ANALYSES OF WATERSHED-AVERAGE PRECIPITATION

Spatial storm analyses (MetStorm[17]) are conducted for historical storms that occurred on the watershed to develop an annual maxima series dataset for 48-hour precipitation. The findings of this analysis are used to calibrate the outputs from the stochastic storm generation algorithm to reasonably replicate the watershed PF behavior of historical storms. Figure 6 depicts a probability-plot of 48-hour watershed-average precipitation annual maxima for the watershed above Cherokee Dam.

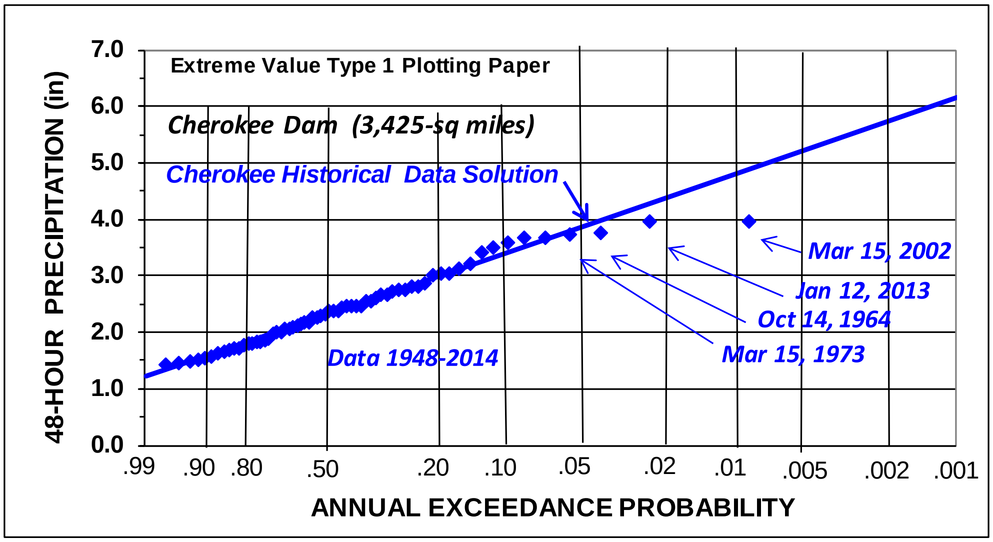

Figure 6 – Probability-Plot of 48-Hour Watershed-Average Precipitation Annual Maxima for the Watershed above Cherokee Dam

5.0 PREDICTOR EQUATIONS FOR WATERSHED-AVERAGE PRECIPITATION

Two predictor equations are developed for estimation of watershed-average precipitation using the findings of spatial storm analyses (MetStorm[17]). The analyses are conducted for historical storms that occurred on the watershed and major storms that occurred within the study area and are transposed to the watershed of interest. Two predictor equations are developed to provide for the ability to characterize uncertainties in the point-to-area predictor equations.

The first predictor equation has the form of a Thiessen polygon approach (Figure 2) which is a linear combination of point precipitation for stations in the network (Equation1). This yields a weighted-average estimate of watershed-average precipitation (P_w) where the weights reflect the areal contribution for the precipitation at each station. These weights are adjusted to account for spatial differences in areal-weighted at-site means for the various polygons relative to the station at-site mean for a given polygon.

The second predictor equation is obtained from multiple linear regression which has the same mathematical form as the Thiessen polygon predictor (Equation 1) and where ε is the error term which accounts for the unexplained variance in the prediction. The error term is expressed as a standardized residual relative to the predicted value to produce homoscedasticity across the range of precipitation magnitudes. The error term is modeled by a 3-parameter Gamma distribution which provides for the ability to model the Normal distribution and distributions with minor skewness.

Equation 1 has the form:

$$P_w = \theta + w_1P_1 + w_2P_2 + . \ . \ . \ . \ + w_nP_n + \ \varepsilon \qquad (1)$$

where: P_w is the 48-hour watershed-average precipitation; θ is a bias correction for the Thiessen approach and the intercept term for multiple regression; w_i are weights for the Thiessen approach and regression coefficients; $P_1 \ . \ . \ . \ P_n$, are concurrent precipitation at stations within the station network; and ε is the error term which accounts for the unexplained variance and is expressed as a standardized residual relative to the predicted value.

Figure 7 depicts the regression relationship for 48-hour watershed-average precipitation using a single station, the key station, for the Cherokee watershed. This example shows the relatively high predictive power of a single station centrally located in the watershed, which is the primary criterion for selection of the key station.

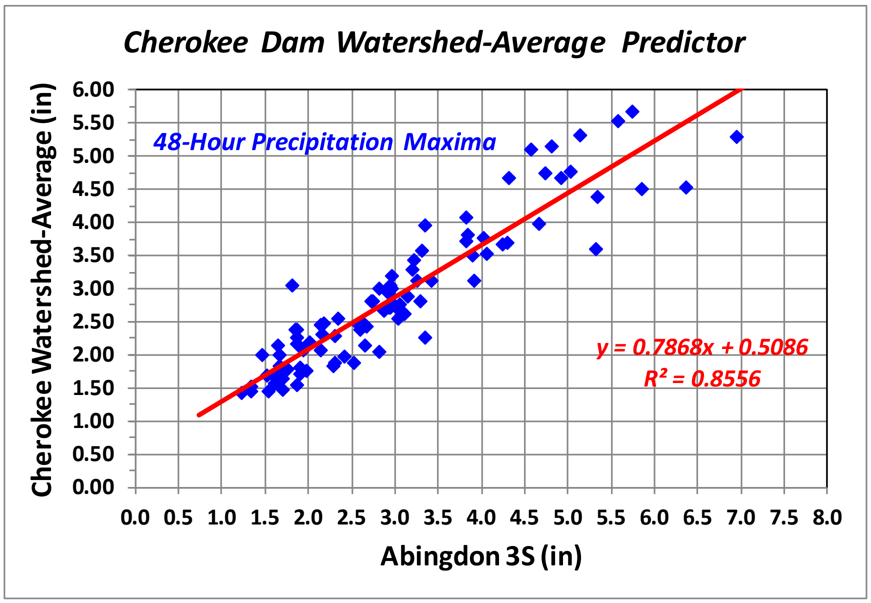

Figure 7 – Linear Regression of 48-Hour Watershed-Average Precipitation for Cherokee Watershed using 48-Hour Precipitation at the Key Station at Abingdon 3S Virginia

Figures 8a and 8b depict comparisons of observed 48-hour watershed-average precipitation for the Cherokee watershed with predicted values using a network of six stations (Figure 2). The very high R-squared measures reflect the excellent explanatory power that has been found in this form of the predictor equation for synoptic scale MLC and TSR storm types.

Table 1 lists comparison of Thiessen weights and regression coefficients for the six stations in the station network for the Cherokee watershed. A review of Table 1 values shows similarity of the magnitude of Thiessen weights and regression coefficients and the low levels of the relative standard error of measure. The predictive power of this form of the point-to-area predictors results in relatively low levels of uncertainty (Figure 17) for this component of the stochastic storm generation process.

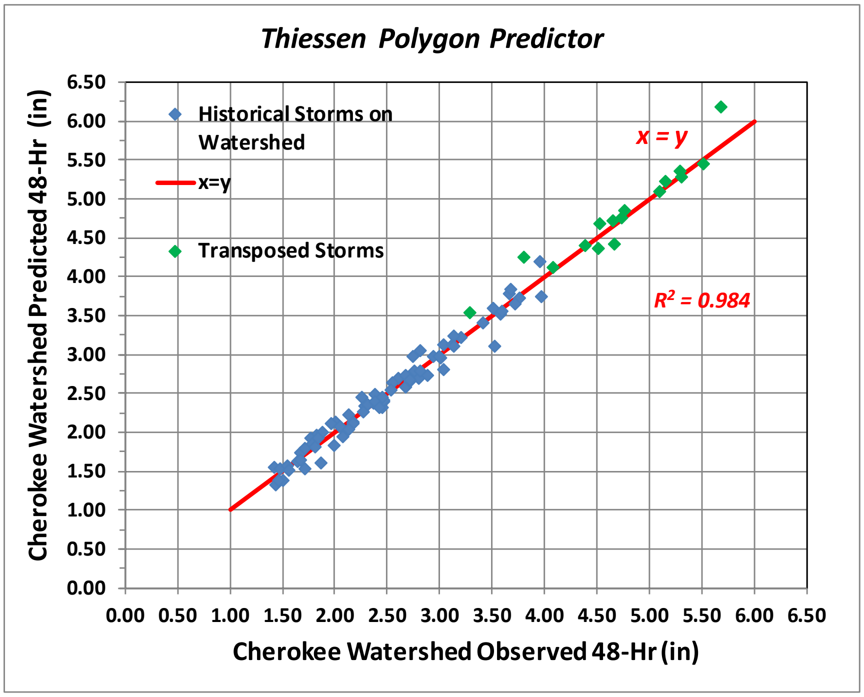

Figure 8a – Comparison of 48-Hour Watershed-Average Precipitation from Spatial Storm Analyses with Predicted 48-Hour Watershed-Average Precipitation using a 6-Station Thiessen Polygon Approach for the Cherokee Watershed

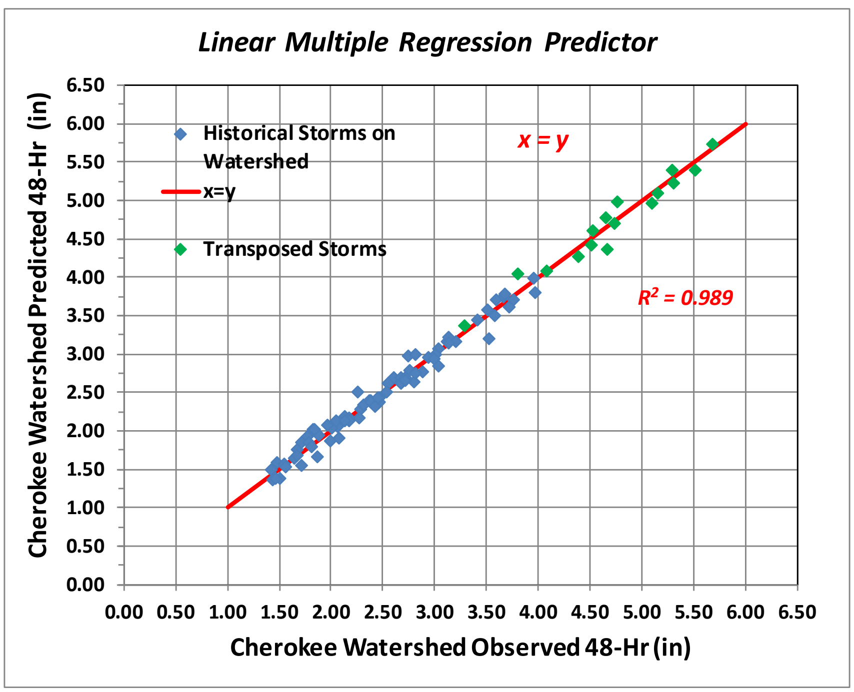

Figure 8b – Comparison of 48-Hour Watershed-Average Precipitation from Spatial Storm Analyses with Predicted 48-Hour Watershed-Average Precipitation using a 6-Station Multiple Linear Regression Approach for the Cherokee Watershed

Table 1 – Comparison of Thiessen Weights and Multiple Regression Coefficients for Cherokee Watershed

Measure or Station	Thiessen Polygon	Multiple Linear Regression
Bias Correction (θ)	-0.0600	-0.0037
Abingdon 3S, VA	0.2061	0.2321
Marion, VA	0.1810	0.1693
Reese, TN	0.1380	0.1851
Elizabethton, TN	0.1892	0.0905
Kingsport, TN	0.1486	0.1213
Rogersville, TN	0.1775	0.2131
Relative Standard Error (ε)	5.0%	4.4%

6.0 SPATIAL CORRELATION STRUCTURE

Cross-correlation information is needed to preserve the spatial correlation structure for precipitation throughout the watershed. In the stochastic storm generation process, point precipitation at the key station is generated first and then point precipitation at the substations in the station network are stochastically generated based on correlation with precipitation at the key station and using the spatial cross-correlation structure amongst all stations.

This is accomplished using a form of the multi-variate spatial disaggregation algorithm that was developed by Lane for use in time series modeling and described in Salas et al[22]. This approach typically requires transformation of precipitation using the *Ladder of Powers* approach (Tukey[32]) or by a Box-Cox[2] transform to yield homoscedasticity and residuals (error term) that are near to being Normally distributed. A square root transform has been found to be suitable for the synoptic scale MLC storm type (Figure 9a). It should be noted that the 48-hour precipitation for each station and the 48-hour precipitation maxima for the watershed have the same starting and ending time (concurrent timing of precipitation).

Figures 9a and 9b depict examples of cross-correlation for the key station and other stations in the network for the Cherokee watershed after the 48-hour concurrent precipitation have received a square root transform. The general spatial behavior is for cross-correlation coefficients to decay (reduce) with distance between stations. There may also be preferential correlation in some directions geographically as orientation of fronts and storm movement tend to have common patterns in a given region. Cross-correlation is also enhanced in mountainous terrain due to the orographically induced precipitation.

The Lane disaggregation procedure utilizes correlation relationships between the key station at Abingdon 3S and substations (correlation coefficients highlighted yellow in Table 2) and cross-correlation between standardized residuals at the substations (highlighted blue in Table 3). The standardized residuals for cross-correlation between substations are computed by division of the residuals from the key station and substation correlation relationship by the standard deviation for the substation. The standardized residuals for cross-correlation between substations in the spatial correlation structure are modeled using the 3-parameter Gamma distribution, typically with zero skewness (Normally distribution) or with minor skewness.

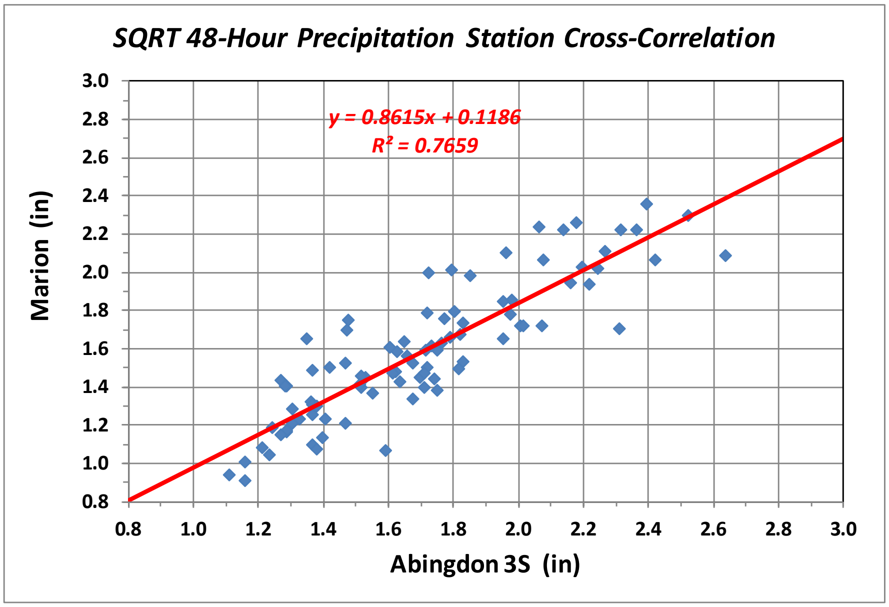

Figure 9a – Correlation of 48-Hour Precipitation at Marion Virginia with 48-Hour Precipitation at the Key Station at Abingdon 3S Virginia (Square Root Transform)

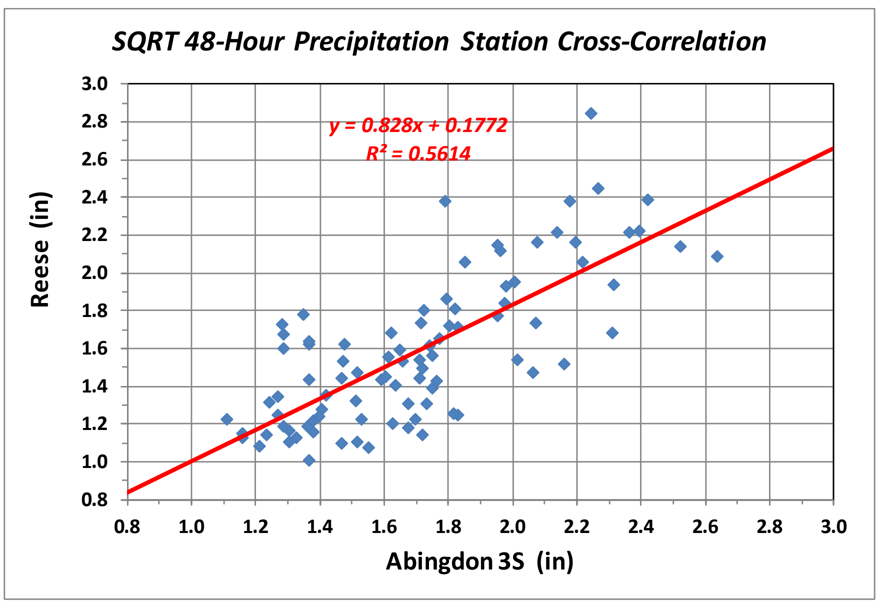

Figure 9b – Correlation of 48-Hour Precipitation at Reese Tennessee with 48-Hour Precipitation at the Key Station at Abingdon 3S Virginia (Square Root Transform)

The stochastic generation of point precipitation at the substations using the Lane procedure can be viewed as incorporating two elements; correlation of substations with the key station and cross-correlation between the sub-stations. The Lane spatial disaggregation procedure involves multi-variate matrix computations for multiple substations but the procedure can be initially viewed from the perspective of a two-station network. First, the correlation relationship between each substation and the key station is preserved which is expressed by the first two terms in Equation 2a. Second, the cross-correlation between substations is preserved in a clever formulation through correlation of the standardized residuals which involves matrix computations to account for the spatial cross-correlation structure and random variates to account for aleatoric uncertainty as indicated by the third term in Equation 2a.

As described above, the first two terms in Equation 2a account for the correlation between the key station and substation (dependency) and the third term accounts for the unexplained variance (random element). For the case of a two-station network such as shown in Figure 9b, a stochastic generation formula would have the form:

$$PT_{Reese} = \alpha + \beta * PT_{Abingdon3S} + SQRT(1-\rho^2) * \sigma_{Reese} * G(0,1,\gamma) \quad (2a)$$

where: PT_{Reese} is the square root transformed point precipitation at the Reese substation; α is the intercept parameter; β is the slope parameter; $PT_{Abingdon3S}$ is the square root transformed point precipitation at the key station; ρ is the correlation coefficient; σ_{Reese} is the standard deviation of the square root transformed point precipitation at the Reese substation; and $G(0,1,\gamma)$ is a random variate drawn from the 3-parameter Gamma distribution with a mean of zero, standard deviation of unity, and a coefficient of skewness γ.

Numerous simulations from Equation 2a will yield a scatter-plot similar to that seen in Figure 9b that will preserve both the dependent portion of the relationship (red line) and the scatter about that relationship. Point precipitation at the Reese substation (P_{Reese}) is then computed by reversing the square root transform as:

$$P_{Reese} = PT_{Reese}^2 \quad (2b)$$

Expanding the Lane procedure to stochastic generation of point precipitation at multiple substations (P_{sub}) involves matrix computations to determine the D matrix which is the solution to:

$$[D][D]^T = [R_{cc}] \quad (3a)$$

where: [D] is a lower triangular matrix solution for matrix $[R_{cc}]$; and $[R_{cc}]$ is a lower triangular matrix of cross-correlation coefficients for substation standardized residuals (Table 3).

The matrix computation of substation precipitation (P_{sub}) is computed from the transformed PT_{sub} which is expressed as:

$$[PT_{sub}] = [\alpha] + [\beta]\, PT_k + [D][G] \quad (3b)$$

$$P_{sub_i} = PT_{sub_i}^{\,2} \quad (3c)$$

where: PT_{sub} is a column matrix of square root transformed point precipitation at the substations; α is a column matrix of intercept parameters for the substations; β is a column matrix of slope parameters for the substations; PT_k is the square root transformed point precipitation at the key station; [D] is a lower triangular matrix from Equation 3a; [G] is a column matrix of random variates drawn from the 3-parameter Gamma distribution $G(0,1,\gamma)$ with a coefficient of skewness γ; and P_{sub_i} is the point precipitation at a specific substation.

The benefit of the Lane spatial disaggregation approach is to preserve the spatial correlation structure for precipitation by preserving the cross-correlation between precipitation at all stations in the network as indicated by the correlation coefficients shown in red in Table 2. This is accomplished by preserving the cross-correlation relationships of the standardized residuals in the lower triangular cross-correlation matrix which are highlighted in blue in Table 3.

Table 2 – Cross-Correlation Matrix of Square Root Transformed 48-Hour Precipitation for Stations within Six Station Network for Cherokee Watershed

Cross-Correlation Matrix for 48-Hour Precipitation for Six Station Network for Cherokee Watershed						
SQRT Transformed	**Abingdon 3S**	Marion	Reese	Elizabethton	Kingsport	Rogersville
Abingdon 3S VA	1.0000					
Marion VA	0.8751	1.0000				
Reese TN	0.7493	0.7876	1.0000			
Elizabethton TN	0.7636	0.7757	0.8101	1.0000		
Kingsport TN	0.7126	0.6517	0.5289	0.7282	1.0000	
Rogersville TN	0.6730	0.6160	0.4525	0.6166	0.8274	1.0000

Table 3 – Cross-Correlation Matrix for Standardized Residuals from Correlation Relationships for Stations within Six Station Network for Cherokee Watershed

Cross-Correlation Matrix for Standardized Residuals for Six Station Network for Cherokee Watershed						
	Abingdon 3S	Marion	Reese TN	Elizabethton	Kingsport	Rogersville
Abingdon 3S VA	n/a					
Marion VA	n/a	1.0000				
Reese TN	n/a	0.4116	1.0000			
Elizabethton TN	n/a	0.3437	0.5563	1.0000		
Kingsport TN	n/a	0.0827	-0.0109	0.4063	1.0000	
Rogersville TN	n/a	0.0755	-0.1058	0.2150	0.6703	1.0000

7.0 MULTIPLE STORMS PER YEAR

The stochastic storm generation algorithm simulates many synoptic scale storms per sample year to emulate the natural processes. The truncated Poisson distribution is used to select the number of storms to be generated in each sample year. The conventional Poisson distribution has a lower bound of zero and the truncated Poisson distribution has a lower bound of one which assures at least one storm per year during the simulations. An analysis of historical storms in the Holston River system yielded a mean value of 6 storms per year. Figure 10 depicts the truncated Poisson distribution with a mean value of 6 storms per year which shows a range of from 1 storm to about 13 storms per year.

Watershed-average precipitation is computed from the point precipitation stochastically generated for the key station and substations for each storm and the point-to-area predictor (Section 5). The watershed annual maximum is then determined as the largest watershed-average precipitation from the collection of storms generated in a given sample year. Annual maxima at the substations are determined from the maximum point precipitation from multiple storms stochastically generated at each substation in a given sample year.

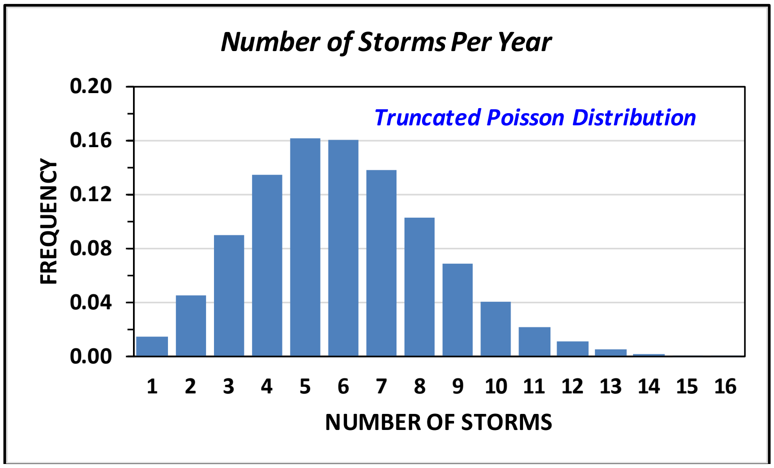

Figure 10 – Probability Distribution for the Truncated Poisson Distribution Applicable to the Cherokee Watershed with a Mean Value of Six Storms per Year

8.0 TOTAL PROBABILITY THEOREM

The Total Probability Theorem (Equation 4) is used to compute annual exceedance probabilities for specific magnitudes of watershed-average precipitation (Benjamin and Cornell[1]). The law of total probability states that given n mutually exclusive events A_1 through A_n whose probabilities sum to unity, then:

$$P(B) = P(B|A_1)\,P(A_1) + P(B|A_2)\,P(A_2) + \;\ldots\ldots + P(B|A_n)\,P(A_n) \qquad (4)$$

where: B is the event of interest and $P(B|A_i)$ is the conditional probability of B given A_i.

In this application, the 48-hour point precipitation at the key station is represented by the variable A and the 48-hour watershed-average precipitation is represented by the variable B. For the case of stochastic storm generation, the 48-hour precipitation at the key station is the dominant variable with regard to explanatory power as seen by the high coefficient of determination (R^2) value in Figure 7. The Total Probability procedure used here is to divide the range of precipitation magnitudes at the key station into intervals (bins) using the Extreme Value Type 1 variate (x-axis) for creation of the bins (Figure 11). Note the bins have equal intervals based on the Extreme Value Type 1 variate, but are not equal intervals in terms of precipitation nor incremental probabilities.

The law of total probability for a continuous probability distribution for watershed-average precipitation (Kuczera and Nathan et al[19]) can be expressed as:

$$p(Q > q) = \sum_{i=1}^{m} p\,(Q > q\;|R_i)\,p[R_i] \qquad (5)$$

where: $p\,(Q > q)$ is the probability of watershed-average precipitation (Q) being greater than a specified magnitude (q) and summed over all precipitation bins (i to m); $p[R_i]$ is the probability of precipitation (R) being within a specific precipitation bin i; and $p(Q > q)|R_i$ is the conditional probability of Q > q given the precipitation magnitude being within precipitation bin *i*, which can be estimated as the number of exceedances of *q* divided by the number of simulations within a precipitation bin.

The Cumulative Distribution Function (CDF) for watershed-average precipitation can be determined by computing exceedance probabilities ($p(Q > q)$) for a series of q values (Figure 12) over the range of magnitudes of watershed-average precipitation (Q) using Equation 5.

The Total Probability Theorem is also used to compute the point PF relationship for each of the substations. In this application, the precipitation annual maxima for each substation is determined from the multiple storms that are generated each sample year. The point PF values at a given substation are represented by variable B in Equation 4 and by Q,q in Equation 5.

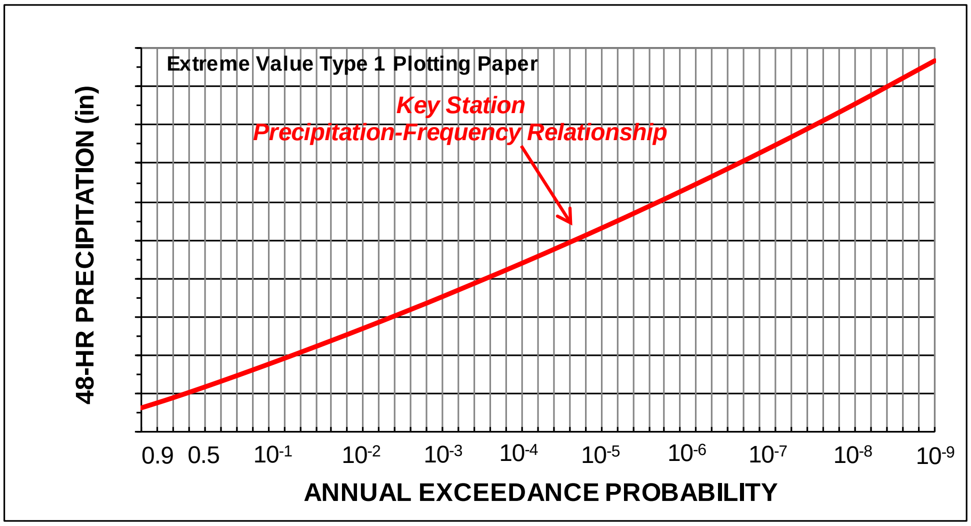

Figure 11 – Division of 48-Hour Precipitation at Key Station into Sampling Intervals (Bins)

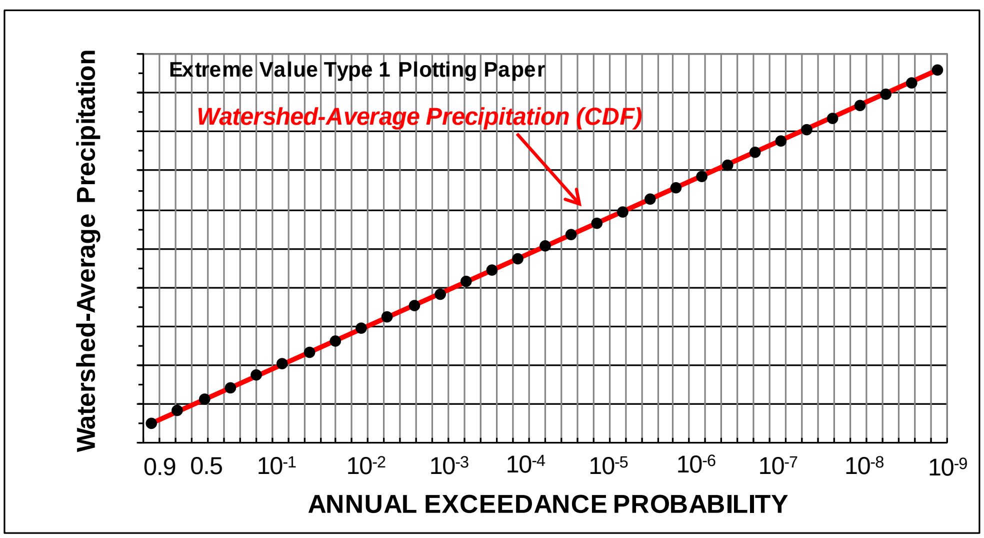

Figure 12 – Cumulative Distribution Function for 48-Hour Watershed-Average Precipitation Defined for Selected Values of 48-Hour Watershed-Average Precipitation

9.0 CALIBRATION AND VALIDATION

Procedures are available for calibration and validation of the watershed-average PF relationship produced by the stochastic storm generation procedure (Flowchart Figure 3).

9.1 Calibration of Watershed-Average Precipitation-Frequency Relationship

The watershed-average PF relationship produced by the stochastic storm generation procedure can be calibrated to the historical watershed-average PF relationship by adjustment of the bias correction term (θ) in Equation 1. Experience indicates that only minor adjustments are usually needed to merge the stochastically generated watershed PF relationship with the historical watershed PF relationship. Figure 13 depicts an example of this type of calibration for the Cherokee watershed where a minor adjustment of θ= -0.11-inch was added to the stochastically generated watershed PF relationship to match the historical data for the Cherokee watershed.

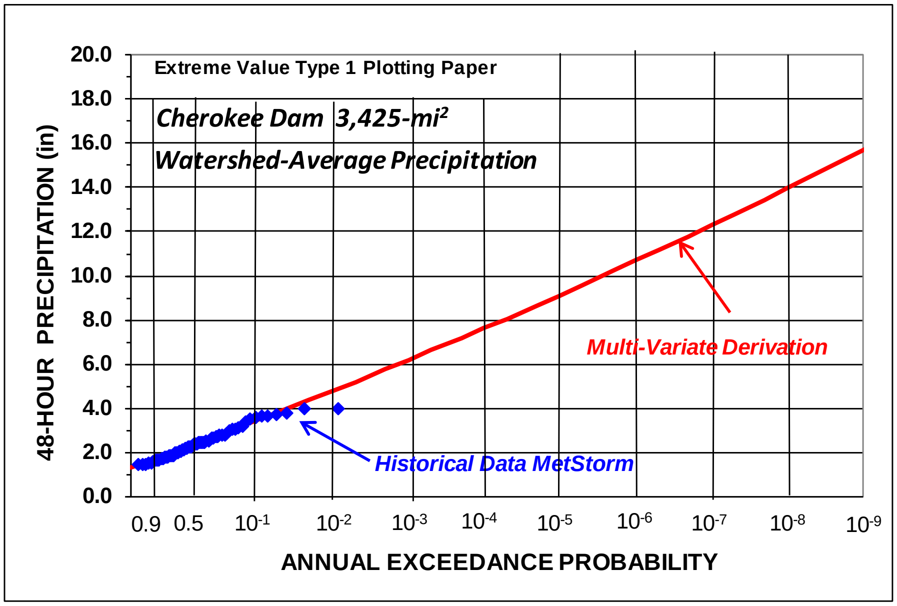

Figure 13 – Depiction of Stochastically Generated 48-Hour Watershed-Average Precipitation Relationship Match to Historical 48-Hour Watershed-Average Precipitation Relationship for Cherokee Watershed

Minor adjustments can also be made to the stochastically generated values of point precipitation at the substations. The intercept parameter in the correlation relationships (α Equation 3b) can be adjusted to preserve the at-site mean at a substation, where only very minor adjustments are typically required to match at-site mean values. Minor adjustments can also be made to the skewness of the standardized residuals in the spatial correlation structure (γ Equation 3a). This adjustment affects the shape of the upper tail of the point PF relationships. Adjustments to the substation at-site means and upper tail behavior are considered fine-tuning to better replicate the point PF relationships at the substations that were determined through the regional PF analysis. These adjustments to the simulation of point precipitation at substations generally result in only minor changes to the watershed PF relationship.

Figures 14a through 14e depict comparisons of point PF relationships obtained from the stochastic storm generation algorithm with regional point PF analyses for the five substations in the Cherokee station network. The point PF relationship for the Abingdon 3S key station (Figure 4) exactly matches the point PF relationship from the regional PF analyses because it was used directly in the stochastic storm generation algorithm. A review of Figures 14a through 14e generally shows strong similarity in point PF relationships and demonstrates the capability of the stochastic storm generation algorithm to generate both point PF relationships and the resultant watershed-average PF relationship for the Mid-Latitude Cyclone and Tropical Storm remnant storm type.

It is important to note that this high level of replication was produced using the regional point PF relationship at the key station and the spatial correlation structure observed in historical storms.

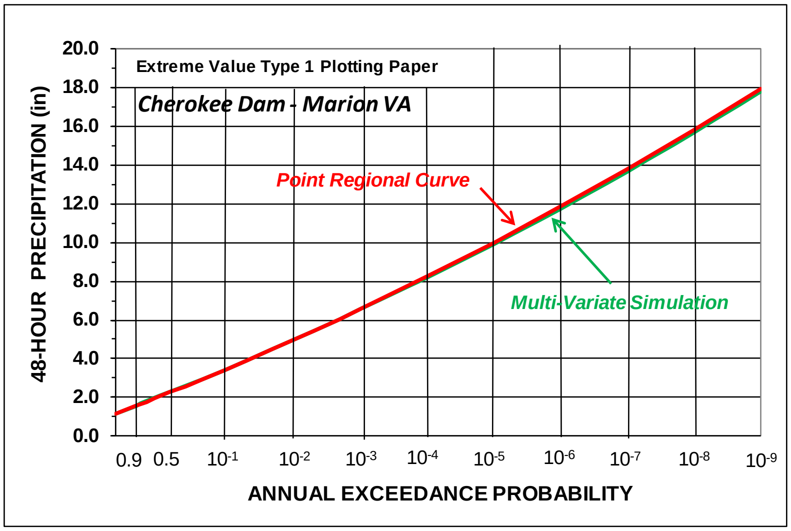

Figure 14a – Comparison of Stochastically Generated Point Precipitation at the Marion VA Substation as Compared to the Point PF Relationship Obtained from the Regional Point PF Analysis

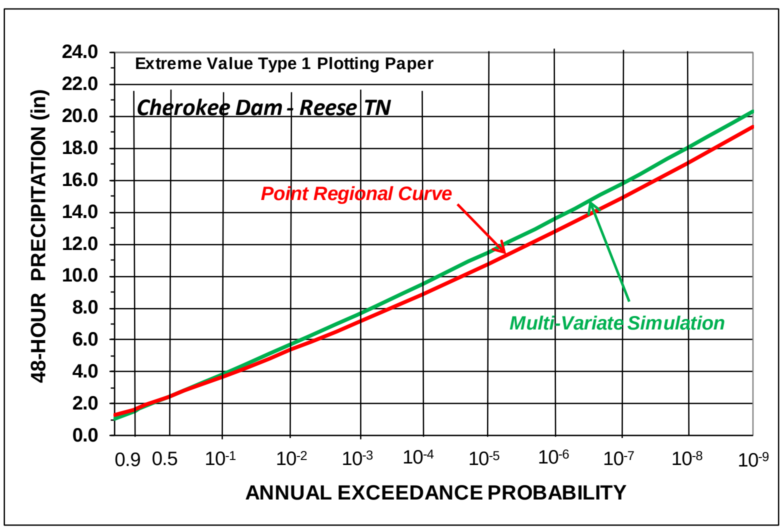

Figure 14b – Comparison of Stochastically Generated Point Precipitation at the Reese TN Substation as Compared to the Point PF Relationship Obtained from the Regional Point PF Analysis

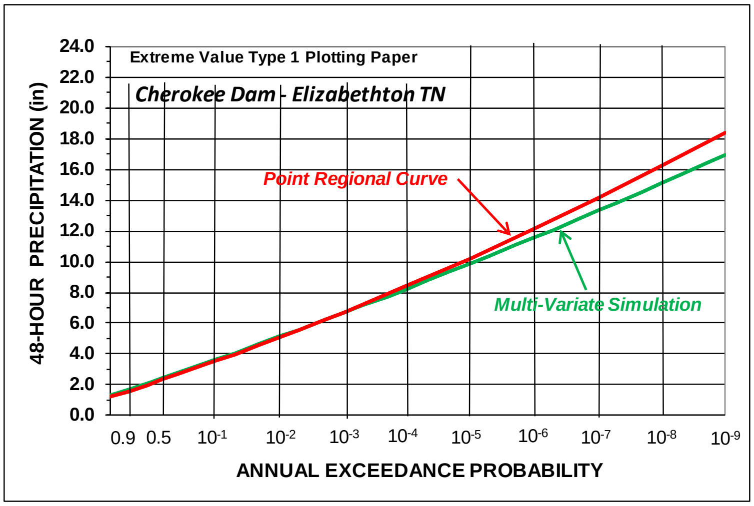

Figure 14c – Comparison of Stochastically Generated Point Precipitation at the Elizabethton TN Substation as Compared to the Point PF Relationship Obtained from the Regional Point PF Analysis

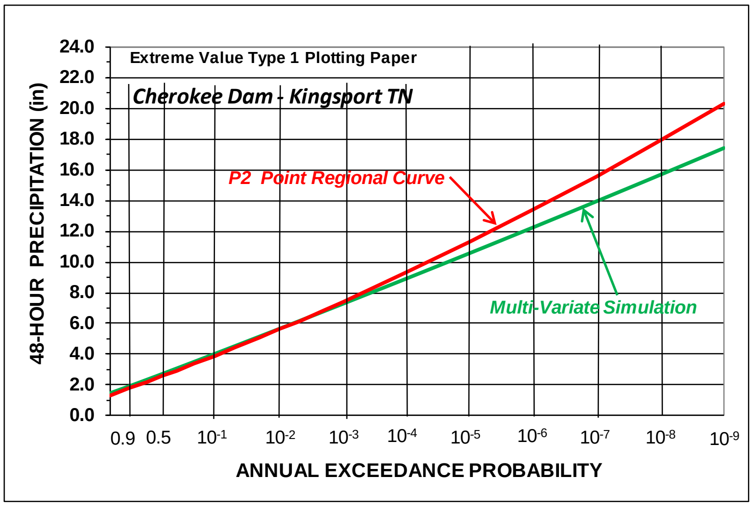

Figure 14d – Comparison of Stochastically Generated Point Precipitation at the Kingsport TN Substation as Compared to the Point PF Relationship Obtained from the Regional Point PF Analysis

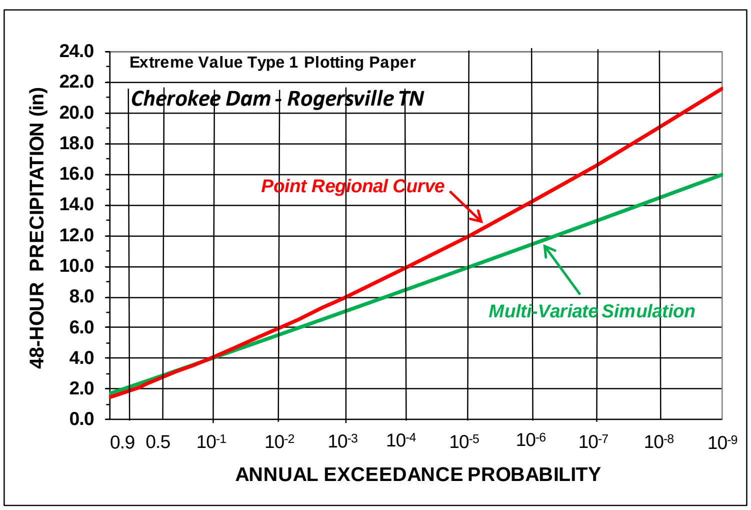

Figure 14e – Comparison of Stochastically Generated Point Precipitation at the Rogersville TN Substation as Compared to the Point PF Relationship Obtained from the Regional Point PF Analysis

9.2 Validation of Watershed Precipitation-Frequency Relationship

Suitability of the watershed PF relationship is validated by several measures and considerations including:

- Point PF relationship at the Key Station is preserved (Figure 4)
- Spatial correlation structure is preserved for network of stations for both historical storms observed on the watershed and the largest storms observed in the region transposed to the watershed of interest (Table 2)
- Point-to-Area relationships (Multiple Regression and Thiessen) have very high explanatory power supporting computation of watershed-average precipitation (Figures 8a and 8b)
- Stochastically generated watershed PF relationship is calibrated to historical watershed-average annual maxima as represented by the historical watershed PF relationship (Figure 13)
- Stochastic generation of multiple storms per year results in point PF relationships at substations in the station network that reasonably replicate the point PF relationships determined from the regional PF analysis (Figures 14a - 14e)

10.0 UNCERTAINTY CONSIDERATIONS

The ability to characterize uncertainties is a major component of the algorithm for stochastic storm generation of a watershed PF relationship. Aleatoric and epistemic uncertainties are inherent to all aspects of the natural processes of synoptic scale MLC and TSR precipitation and the algorithm created to simulate the natural processes. The aleatoric uncertainties are addressed in the simulation procedures in the inner (blue) loop of the Flow Chart shown in Figure 3 and epistemic uncertainties are addressed in the outer loop of the Flow Chart. Epistemic uncertainties for each of the components of the algorithm are discussed in the following sections.

10.1 Uncertainties in Key Station Precipitation-Frequency Relationship

The point PF relationship at the key station is a primary driver for stochastic storm generation. There are four epistemic uncertainties that are addressed as part of the algorithm that include: the key station at-site mean; regional L-Cv and L-Skewness and the regional probability distribution.

Key Station At-Site Mean – Epistemic uncertainty in the at-site mean at the key station is modeled by a Normally distributed random variable with mean zero and a standard deviation based on the record length at the key station and regional information from the relative standard error of spatially mapped at-site means. This typically represents about a 4-5% standard error relative to the at-site mean. This information is obtained from the relative standard error computed in spatial mapping of at-site means for the project study area.

Regional L-Cv and Regional L-Skewness – Sampling properties of L-Moment ratios are near Normally distributed (Hosking and Wallis[12]). Epistemic uncertainties in the regional values of L-Cv and L-Skewness are based on the relative standard error of regional solutions for L-Cv and L-Skewness that are used in spatial mapping (MGS et al[18]). A Normally distributed random variable with mean zero and a standard deviation equal to the relative standard error are used in stochastic simulations. Relative standard errors are commonly on the order of 2% and 6% for regional L-Cv and L-Skewness, respectively, when very large regional datasets of precipitation annual maxima and storm typing are employed.

Regional Probability Distribution – The choice of the 4-parameter Kappa distribution (Hosking[11], Hosking and Walllis[12]) provides the capability to emulate alternative probability distributions near the chosen 3-parameter regional probability distribution. Specifically, the second shape parameter (*h*, of the Kappa distribution (Equation 6) can be treated as a fixed value which yields a 3-parameter distribution. In this context, the Generalized Logistic (GL), Generalized Extreme Value (GEV), Gaucho and Generalized Pareto (GP) are seen as special cases of the 4-parameter Kappa distribution where the second shape parameter has values of -1, 0, +0.5 and +1, respectively. An L-Moment ratio diagram is depicted in Figure 15 which shows the regional L-Moment Ratio pairings for 38 homogeneous regions for the MLC storm type and the regional probability distribution being very near the GEV, as indicated by the centroid of the data.

The quantile function for the 4-parameter Kappa distribution is:

$$q(F) = \xi + \frac{\alpha}{\kappa}\left\{1-\left(\frac{1-F^h}{h}\right)^{\kappa}\right\} \qquad (6)$$

where: ξ, α, κ, and h are location, scale and two shape parameters respectively.

In practice, the 4-parameter Kappa distribution is selected as the regional probability distribution and the second shape parameter (h) is determined in increments of 0.05, which effectively yields a 3-parameter distribution. There is mathematical reasoning (Gumbel[6]) to support the case that the GEV is approached asymptotically based on the mathematical form of the probability distribution that describes the precipitation magnitudes generated by a specific meteorological process and the number of storms generated in a given year from which the annual maxima is selected. This consideration supports the choice of a regional probability distribution that is near, but not yet asymptotic to, the GEV.

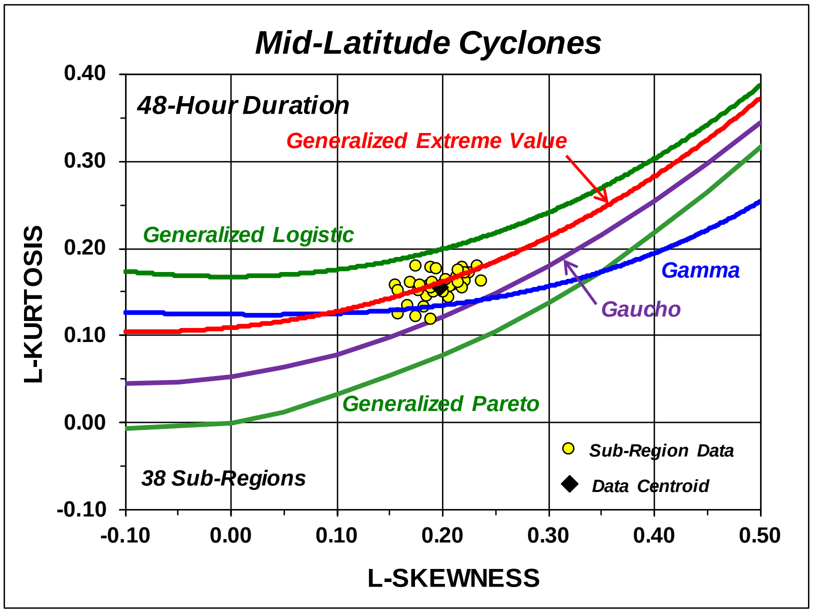

Figure 15 – L-Moment Ratio Diagram Showing L-Skewness and L-Kurtosis Pairings for 38 Homogeneous Regions for the Mid-Latitude Cyclone Storm Type

Epistemic uncertainty about the chosen 3-parameter regional probability distribution is modeled by preserving the functional relationship between L-Moment ratios for L-Skewness and L-Kurtosis for the fixed value of the second shape parameter while allowing for uncertainty (variability) in both L-Skewness and L-Kurtosis. This can be viewed as having a L-Moment ratio curve near and parallel to one of the 3-parameter distributions shown in Figure 15, such as the GEV. Uncertainty about L-Skewness is described in the prior section and uncertainty in L-Kurtosis is characterized by a Normally distributed random variable. An Equivalent Independent Record Length (EIRL) measure (MGS et al[18]) is used in estimating the standard deviation for L-Kurtosis where epistemic uncertainties in L-Kurtosis are similar in magnitude to that for L-Skewness.

10.2 Uncertainties in Point-to-Area Predictors of Watershed-Average Precipitation

Two predictors of watershed-average precipitation are developed to provide for characterization of uncertainty in the point-to-area estimation of watershed-average precipitation. These include the linear multiple regression and Thiessen Polygon approaches described previously. Epistemic uncertainty is characterized by using a weighted-average of the two predictions (Equation 7) where the mean value of the weights is based on the unexplained variance for each predictor (Kuczera[13]). The standard deviation of the weights is set wide to allow a wide range of weightings and modeled using the 4-parameter Beta distribution (Benjamin and Cornell[1]). For the case of the Cherokee watershed, the mean weights are 0.53, 0.47 for the multiple regression and Thiessen polygon approaches respectively, and epistemic uncertainty weightings vary from fully multiple regression (w_{mr} =1.00) to fully Thiessen Polygon (Figure 16).

$$P_w = w_{mr} {}^* P_{Wmr} + w_{tp} P_{Wtp} \qquad (7)$$

where: P_w is the watershed-average precipitation; w_{mr} is the weight for the multiple linear regression estimate of watershed-average precipitation P_{Wmr}; w_{tp} is the weight for the Thiessen Polygon estimate of watershed-average precipitation P_{Wtp}; and weights w_{mr} and w_{tp} sum to unity.

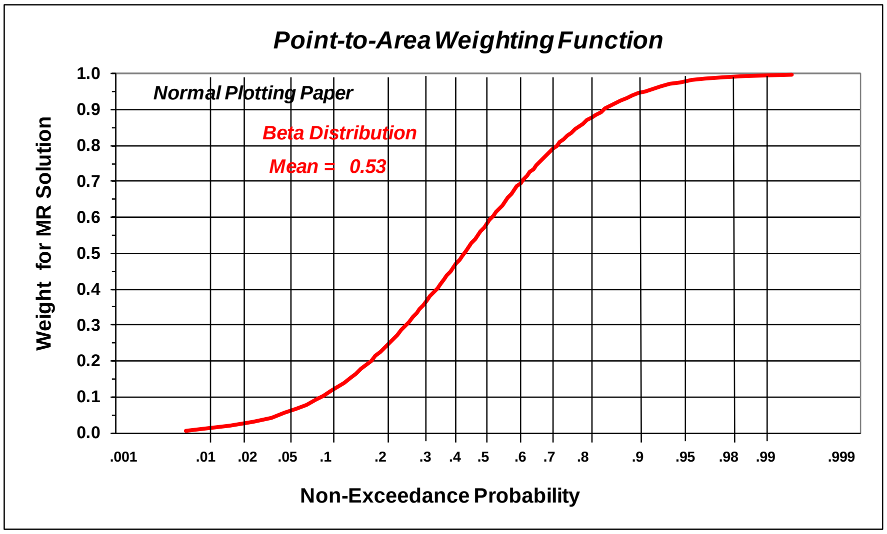

Figure 16 – Characterization of Uncertainty in Weighting Factor for Multiple Linear Regression Predictor of Watershed-Average Precipitation

10.3 Uncertainties in the Spatial Correlation Structure

Epistemic uncertainty in the spatial correlation structure is modeled using a bootstrap resampling procedure. This is accomplished by stochastically generating a sample set of point precipitation values for the substations using the spatial correlation structure determined from the historical storms on the watershed and the large storms in the region that were transposed to the watershed. The aleatoric uncertainty which arises from the standardized residuals term in Equation 3a results in alternative solutions of the spatial correlation structure. These alternative solutions for the spatial correlation structure are used to emulate the epistemic uncertainty in the spatial correlation structure.

10.4 Relative Contributions of Components to the Total Uncertainty

A separate set of stochastic storm simulations are conducted to identify the relative contribution of each component to the total uncertainty. This analysis is conducted by first computing the total uncertainty variance where Latin hypercube sampling allows all components to vary through their respective ranges of uncertainty. Subsequent analyses are then conducted wherein the value of one component of the stochastic storm generation algorithm is fixed at its mean value and the resultant uncertainty variance is compared to the simulations where uncertainties are considered for all components. Figure 17 shows a stacked histogram where the relative contribution of uncertainty for each component for the Cherokee watershed is depicted. The square root of the uncertainty variance on the ordinate of Figure 17 corresponds to the standard deviation for the width of uncertainty bounds shown in Figure 18.

It is important to note that the relative contribution of uncertainty varies with the annual exceedance probability of watershed precipitation. In particular, the uncertainty in L-Skewness and the identification of the regional probability distribution are the primary contributors to uncertainty for extreme precipitation. These two components have the greatest influence on the shape of the upper tail of the watershed PF relationship.

Further reduction in epistemic uncertainty for L-Skewness and the regional probability distribution is possible in the future. The use of storm typing and large regional studies have the potential to provide an increased understanding in the range of L-Skewness values and identification of the regional probability distribution and associated probabilistic behavior for synoptic scale storm processes.

Figure 19 depicts the watershed point PF relationship for areal-average values of the at-site mean for the watershed, and regional L-Cv and regional L-Skewness for the Cherokee watershed. The watershed point PF relationship has a common level of annual exceedance probability at all locations throughout the watershed for a given value of 48-hour precipitation. The difference between the watershed point PF relationship and the areal-average watershed PF relationship can be viewed in the context of an Areal Reduction Factor (ARF). It should be noted the ARFs for PF relationships are not fixed values but vary with annual exceedance probability. Smaller areal reduction occurs for common storm events and larger areal reduction occurs for extreme storm events.

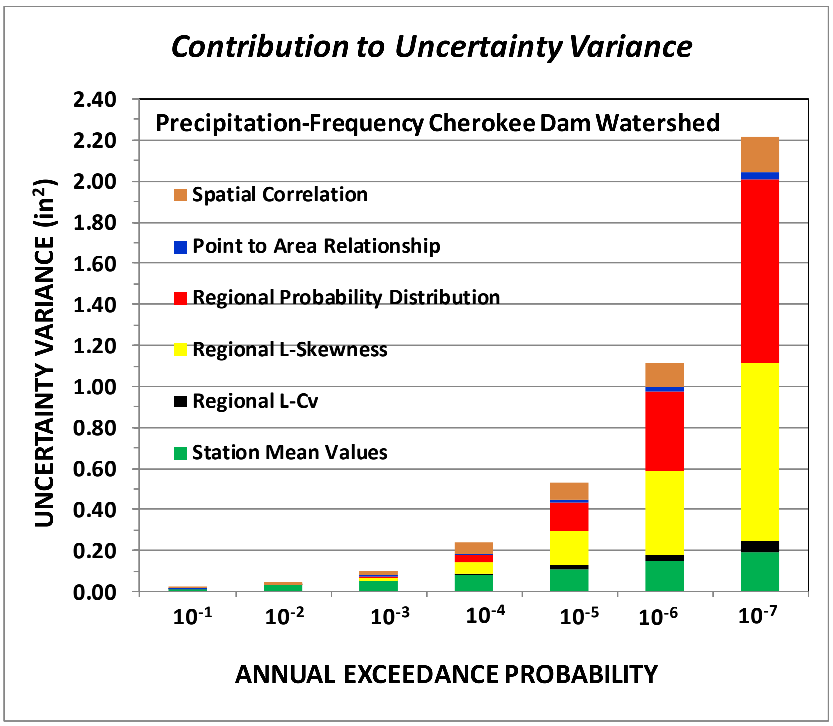

Figure 17 – Stacked Histogram Showing Relative Contribution of Various Sources of Uncertainty to the Total Uncertainty in the Watershed Precipitation-Frequency Relationship for the Cherokee Watershed TN

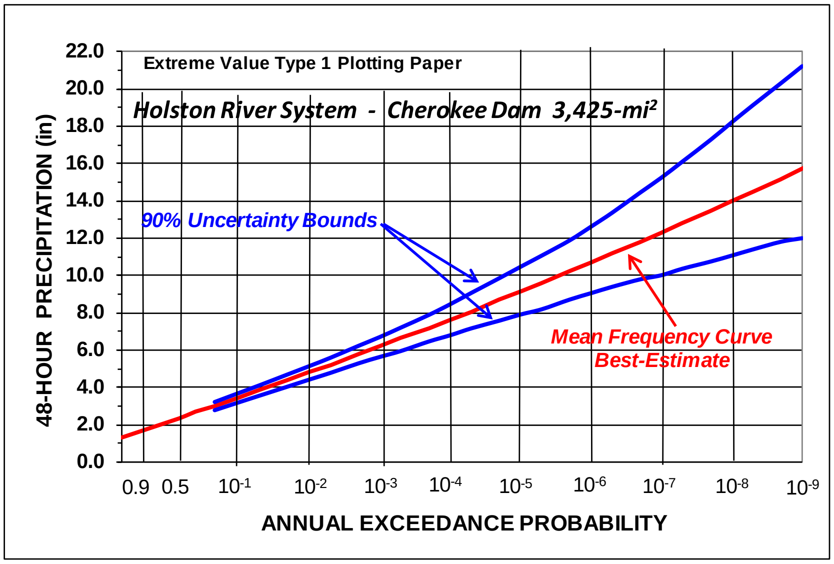

Figure 18 – Watershed Precipitation-Frequency Relationship and 90% Uncertainty Bounds for the Cherokee Watershed on the Holston River TN

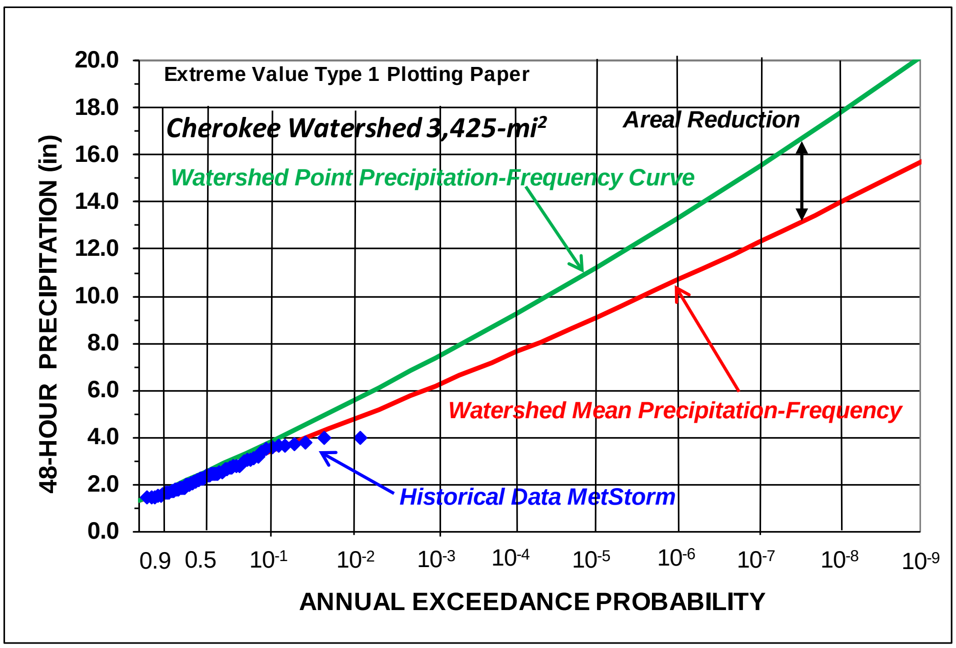

Figure 19 – Comparison of 48-Hour Watershed Point Precipitation-Frequency Relationship and 48-Hour Watershed-Average Precipitation-Frequency Relationship for the Cherokee Watershed on the Holston River TN

11.0 REFERENCES AND RELATED ARTICLES

1. Benjamin JR and Cornell CA, Probability and Statistics for Civil Engineers, McGraw-Hill, 1970.
2. Box GEP and Cox DR, An Analysis of Transformations, Journal of the Royal Statistical Society, Series B, *26*, 211-252, 1964.
3. Cunnane C, Unbiased Plotting Positions - A Review, Journal of Hydrology, 37, 205-222, 1978.
4. Efron B, Bootstrap Methods: Another Look at the Jackknife, The Annals of Statistics, 7 (1): pp1–

26, 1979.

5. Fiorentino MS, Gabrielle S, Rossi F and Versace P, Hierarchical Approach for Regional Frequency Analysis, *Regional Frequency Analysis*, Edited by VP Singh, pp35-49, Reidel, Norwell Massachusetts, 1979.
6. Gumbel EJ, Statistics of Extremes, Columbia University Press, 1958.
7. Hosking JRM, The Theory of Probability Weighted Moments, IBM Research ReportRC12210, October 1986.
8. Hosking JRM, L-Moments: Analysis and Estimation of Distributions using Linear Combinations of Order Statistics, Journal Royal Statistical Society, Ser B, 52, pp105-124, 1990.
9. Hosking JRM, and Wallis JR, Some Statistics Useful in Regional Frequency Analysis, Water Resources Research, Vol 29, No. 2, pp271-281, Feb 1993.
10. Hosking JRM, and Wallis JR, The Effect of Inter-Site Dependence on Regional Flood Frequency Analysis, Water Resources Research, Vol 24, pp588-600, 1988.
11. Hosking JRM, The Four Parameter KAPPA Distribution, IBM Research Report RC13412, IBM Research Yorktown Heights NY, 1988.
12. Hosking JRM, and Wallis JR, Regional Frequency Analysis - An Approach Based on L-Moments, Cambridge Press, 1997.
13. Kuczera G, Combining Site-Specific and Regional Information: An Empirical Bayes Approach, Water Resources Research, Volume 18, No. 2, pp306-314, April 1982.
14. L-RAP, L-Moments Regional Analysis Program, developed by MG Schaefer and BL Barker, MGS Software LLC, Olympia WA, 2005.
15. Langbein WB, Annual Floods and the Partial Duration Flood Series, Transaction American Geophysical Union, Vol 30, pp879-881, 1949.
16. McKay MD, Conover WJ, and Beckman RJ, A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code, Technometrics, 221, pp239-245, 1979.
17. MetStorm, Storm Analysis Software developed by MetStat Inc, Fort Collins CO, 2015.
18. MGS Engineering Consultants, MetStat, Applied Climate Services and Riverside Technology, Regional Precipitation-Frequency Analyses for Mid-Latitude Cyclones, Mesoscale Storms with Embedded Convection, Local Storms and Tropical Storm Remnant Storm Types in the Tennessee Valley Watershed, prepared for Tennessee Valley Authority, January 2015.
19. Nathan RJ, Weinmann EM and Hill P, Use of Monte Carlo Simulation to Estimate the Expected Probability of Large to Extreme Floods, 28th International Hydrology and Water Resources Symposium, Wollongong, NSW, The Institution of Engineers, Australia, 2003.
20. Nathan R, Jordan P, Scorah M, Lang S, Kuczera G, Schaefer M and Weinmann E, Estimating the Exceedance Probability of Probable Maximum Precipitation, Journal of Hydrology, Apr 2016.
21. Oregon Climate Service, Mean Annual Precipitation Maps for United States, prepared with

PRISM Model for NRCS, Corvallis Oregon, 2005.

22. Salas JD, Delleur JW, Yevjevich V and Lane WL, Applied Modeling of Hydrologic Time Series, Water Resources Publications, 4th Printing, 1997.
23. Schaefer MG, Regional Analyses of Precipitation Annual Maxima in Washington State, Water Resources Research, Vol.26, No.1, pp119-132, January,1990.
24. Schaefer MG, Taylor GH and Wallis JR, Spatial Mapping of 100-Year and 1000-Year Precipitation for 6-Hour, 24-Hour and 72-Hour Durations for Southwest British Columbia, MGS Engineering Consultants, Inc., prepared for BCHydro Power Supply and Engineering, March 2001.
25. Schaefer MG and Barker BL, Precipitation Magnitude-Frequency Characteristics for the American River Watershed, prepared for US Army Corps of Engineers, Hydrologic Engineering Center, MGS Engineering Consultants, January 2000.
26. Schaefer MG and Barker BL, Stochastic Modeling of Extreme Floods on the American River at Folsom Dam Flood-Frequency Curve Extension, MGS Engineering Consultants for US, US Army Corps of Engineers, Hydrologic Engineering Center, Davis CA, September 2005.
27. Schaefer MG, Barker BL, Taylor GH and Wallis JR, Regional Precipitation-Frequency Analysis and Spatial Mapping of Precipitation for 24-Hour and 2-Hour Durations in Western Washington, prepared for Washington State Department of Transportation, Report WA-RD 544.1, MGS Engineering Consultants, March 2002.
28. Schaefer MG, Barker BL, Taylor GH and Wallis JR, Regional Precipitation-Frequency Analysis and Spatial Mapping of Precipitation for 24-Hour and 2-Hour Durations in Eastern Washington, prepared for Washington State Department of Transportation, MGS Engineering Consultants, January 2006.
29. Schaefer MG, Barker BL, Taylor GH and Wallis JR, Regional Precipitation-Frequency Analysis and Spatial Mapping of 24-Hour Precipitation for Oregon, prepared for Oregon State Department of Transportation, MGS Engineering Consultants, June 2007.
30. Stedinger JR, Vogel RM and Foufoula-Georgiou E, Frequency Analysis of Extreme Events, Chapter 18, Handbook of Hydrology, D Maidment (editor), McGraw-Hill Inc, NY 1993.
31. SEFM, Stochastic Event Flood Model User's Manual, MGS Software LLC Olympia WA, developed by Schaefer MG and Barker BL, 2001, 2009, 2015.
32. Tukey JW, Exploratory Data Analysis, Addison-Wesley, Reading, MA, 1977
33. Weiss LL, Ratio of True to Fixed Interval Maximum Rainfall, Journal Hydraulics, ASCE, 90(HY1), pp77-82, 1964.
34. Wyss GD and Jorgenson KH, A Users Guide to LHS: Sandia's Latin Hypercube Sampling Software, Sandia National Laboratories, report SAND98-0210, February 1998.