NOAA Atlas 14 Volume 2 Documentation Review

February 17 – March 17, 2005
Ohio River Basin and Surrounding States

NOAA's National Weather Service
Office of Hydrologic Development
Hydrometeorological Design Studies Center
Silver Spring, Maryland
June 2, 2005

Introduction

This document is a consolidated summary of all the review comments with our response. The wording of the comments was unchanged to make sure the meaning was not misconstrued and so individual reviewers could identify their comments. We have distilled the comments into 4 separate categories:

1. **Format or typo's**
2. **Clarification considerations**
3. **Technical questions**
4. **General comments**

1 Format or typo's

1.1 Page 45 - Figure 4.8.2 is in text ahead of Figure 4.8.1. Why? I assume because of size. See pages 45, 46, and 47.

Response: Yes, the size of Figure 4.8.1 and the general flow of the document influenced the placement of Figure 4.8.1. For readability purposes, we are going to leave it as is.

1.2 Page 1: 2. Preface to Volume 2, 2nd paragraph, 1st sentence: Administration's is shown as "Administration=s" on both my printout and on my computer. Please check if this is a simple "typo" or a text translation problem from the PFDS to my computer.

Page 1: Paragraph starting with "The work …": has the word "Administration=s" should be "Administration's"

Page 7: Figure 4.1.1: Sub-title at bottom of page indicates "Hourly and N-Minute Stations". This needs to be deleted; not what is shown on this figure. See correct title at top of figure. Suggest replacement with 'Daily Stations' (see figure 4.1.2).

Page 9: Record length, 6th sentence: "N- minute records used in the analysis had 14 to nearly 100 years of data …". Table 4.1.2 indicates there is one N-minute station that had 105 data yrs of record. Would delete the word "nearly" and replace with 'just over' or simply state '105 years of data'.

Page 61: 1st paragraph: Appendix reference should read A.2 and not A.3.

Page 68: Last paragraph: Reference to Appendices agrees with what is obtained from the PFDS but not what is shown in your "Table of Contents".

Page 5: 3rd paragraph, 1st sentence: Text refers to "Both the spatial interpolation and the point estimates were subject to external peer reviews (see Section 6 and Appendix 5)." This is true as to what I see directly from the PFDS; however, please look at the 'Table of Contents" in the documentation which states these peer reviews is

contained in Appendices A.5 (Point) and A.6 (Spatial). Need to correct your "Table of Contents" to reflect what is really shown in the text and actual Appendices.

Page 19: missing the words "to the" between "proportional" and "number" in the last sentence of the paragraph beginning, "The regional approach employs…"

Page 20: Paragraph beginning, "Regional frequency analysis…": last sentence. "Data" is a plural term so it should read, "…if a station's data are consistent with…"

Page 28: 6th line from the top. "…for a region based, but this did not…" This sentence doesn't make sense. Remove "based"?

Page 41: bottom, reference of Chow, 1988 is given. This should be Chow et al. Also the "Chow" should not be in the parentheses.

Page 47: top text box: Last word in box should be "contours", plural.

Response: Thank you very much for catching these editorial mistakes. They have been corrected.

1.3 Page 9 - Why under Record Length the average data years for daily stations is 54 and 37 for hourly stations, but in Table 4.1.2 the average is 63 and 40.

Page 9: Record length, 3rd sentence: Text indicates "… average 54 data years in length for daily stations and 37 data years for hourly (Table 4.1.2)." However, Table 4.1.2 indicates an average record length (data yrs) of 63 for daily and 40 for hourly. Since both locations refer to "data years", I don't understand why different values are listed.

Page 13: Table 4.1.6: Depending on what set of data years are correct, you may have to adjust the numbers show in the 3rd column of Table 4.1.6.

Response: The numbers in the paragraph were residual text left from the Volume 1 documentation. The values in the tables are correct. The information in the paragraph has been corrected.

1.4 Page 55: Bottom two paragraphs. The authors have changed a writing style with the "≥" and "≤" symbols. I'd prefer the text equivalent, (e.g. "In these 21 cases, the precipitation frequency estimates less than 12 hours …")

Response: In the interest of being concise and using as few words as possible, it was decided to keep the symbols.

1.5 References: Maybe the problem is with my computer/printer or with the PDF converter, but these pages had VERY POOR formatting so that very large gaps between words appear or else words are overstriking other words. Please check that you're not using some unusual fonts that might not be universally available.

Response: It may be an issue with your computer/printer. The formatting seems OK. We are using a basic font, Times New Roman. Please let us know if you continue to have problems.

2 Clarification considerations

2.1 In looking for references, I found that some are listed under References and some in Appendix A.4, References, e.g., USDA-NRCS, 1998. However, I assume those in Appendix A.4 apply only to that Appendix report – although they may be listed also in the main text.

Response: We wanted to minimize any alterations of the Final Report contained in Appendix 4 from the original form submitted to us by Christopher Daly and George Taylor. Therefore, we left their Appendix in tact. References found within both the Final Report and the main sections of the documentation are also referenced in the reference section of NOAA Atlas 14 Volume 2.

2.2 Page 59 - Section 5. I assume point rainfall values can be easily downloaded in the same fashion as in Volume 1. The sample outputs, Figures 5.4.1 through 5.4.4, appear adequate.

Response: Yes, that is correct.

2.3 Page 16 - Tables 4.1.7 and 4.1.8 list "wet season" months by regions. However, the various homogeneous regions are not defined until Figures 4.4.1 and 4.4.2. Is that correct? I have no problems with the regionalizations – it just took more reading to find them.

Page 14 - Section 4.1.3. Tables 4.1.7 and 4.1.8 refer to homogeneous regions, yet one has to go to Figure 4.4.1 and 4.4.2 to observe their geographic locations – which are not covered until Section 4.4.

Page 16: Tables 4.1.7 and 4.1.8 refer to many, many regions but the maps for these regions do not appear nor is there a reference as to where to find these maps. The maps should appear here first. I assume these are the maps that currently appear on pages 24 and 25.

Response: Due to the complexity of the material and dependence of one part of the procedure on another, this was necessary. A sentence clearly identifying the section and figures of the regions has been added.

2.4 Page 4: 3.3 Approach, 2nd paragraph: "This Atlas introduces…". Would start as: "NOAA Atlas 14 introduces…". This avoids the thought that your statement here only applies to Volume 2 of the Series. Page 5: 1st paragraph, 1st sentence: Change "This Atlas" to read NOAA Atlas 14.

Page 13: 1st paragraph, 2nd , 3rd & 5th sentences: Use of words "this Atlas". See my previous comments 2 & 3 for additional reference. Again I have some confusion with your using the words "this Atlas". Here your referring to a comparison between T.P.40 and NOAA Atlas 14 (1st sentence) record length and specifically, in the 5th sentence, minimum length of daily and hourly station records used in each study. In particular you state that "in this Atlas" the minimum length of record used was increased to 30 data years for daily and 20 data years for hourly which is true for NOAA Atlas 14, Vol. 2 only. It is not true for NOAA Atlas 14, Vol. 1 where the minimum station record length used was 20 data years for daily and 15 data years for hourly (see section 4.1.1: Record length of NOAA Atlas 14, Vol. 1). My suggestion is to make sure your either referring to the NOAA Atlas 14 series as a whole or to specifically what volume of NOAA Atlas 14 when you use the reference "this Atlas".

Response: Good idea. These changes where made to both NOAA Atlas 14 Volumes 1 and 2.

2.5 Page 10: N-minute data, 1s paragraph, 9th line down: "Observed variations in n-minute ratios over average recurrence intervals, 2-year to 1,000-year, were preserved in the final n-minute ratios." What is meant by "preserved"? My cursory examination of the ratios shown in both tables 4.1.3 & 4.1.4 indicates that the ratios increase in a normal manner with increasing duration for a particular recurrence interval and decrease in a normal fashion with increasing recurrence interval for a particular duration. Would delete sentence or further explain meaning of sentence.

Response: Agreed. We have deleted the sentence.

2.6 Page 21: Discordancy, last sentence of paragraph, reference to list of stations found in Appendix A.6: In going to the PFDS I see there are two A.6 tables or Appendices indicated; therefore, shouldn't the text indicate "Appendices A.6 instead of singular? Additionally, In the "Table of Contents" shown at the beginning of the documentation, "Station lists" are indicated as being in Appendix A.7. Need to justify what is shown in the "Table of Contents" of the documentation with what is really contained in the individual Appendices. There seems to be a problem starting with Appendix A.5.

Response: We had inadvertently left the peer review appendices listed in the Table of Contents as two separate appendices. There is actually only one for Volume 2. The station list appendix has been corrected in the Table of Contents to be correctly referred to as Appendix A.6. There is only one Appendix A.6, but it happens to be split in two parts, one for each of the two tables contained. This split was to accommodate the large size of the daily table, Table A.6.1. Therefore, we will leave "Appendix" as singular, but pluralize "lists". We also inadvertently listed two peer review appendices when there is only one in Volume 2. The Table of Contents has been corrected.

2.7 Page 23, bottom of page, Appendices reference: Reference to Appendix A.7 and A.8 I think are correct but doesn't agree with what you state in the "Table of Contents". Is it Appendices A.8 or Appendix A.8?

Response: The Table of Contents has been corrected. The reference was corrected to say "Appendix A.8".

2.8 Page A.1.1: Introduction, 2nd paragraph, 1st sentence: Suggest one rewrite this sentence as "Temporal distributions (using all cases) for each duration are presented in Figure A.1.1." The (using all cases) was added for clarity and Figures was changed from plural to singular.

Response: Thank you for the suggestions. We changed "Figures" to "Figure" in that sentence. However, since the concept of quartiles isn't introduced until the next sentence, we didn't feel it necessary to add the clarifier "(using all cases)".

2.9 In section 4.4, it was noted that the regions were initially chosen based on climatology but no methodology was described (first paragraph). Was it cluster analysis or principal component analysis or climate division or something else? I'm not referring to the later discussion about the L-moments tools - just the initial cut.

Response: The original delineation of regions was done by hand drawing upon knowledge of climatology, season(s) of highest precipitation, type of precipitation (e.g., general storm, convective, tropical storms or hurricanes, or a combination), topography and the homogeneity of such characteristics in a given geographic area. Granted, these are fairly subjective, but are based on professional knowledge and experience. The subsequent objective testing of these regions using L-moments further addressed the appropriateness of the regions and led to subdivisions, etc. The text has been modified to make this more clear.

2.10 Page 28, The Monte Carlo Simulation test – The first sentence in this paragraph is not totally clear. I think 1,000 simulations of L moments were done with sites having the same record lengths and average L moments as the observed data for the region being evaluated. I think this statement is more clear than what is given in the report. It might help to give the equation for the GZ goodness-of-fit test to show that an observed L kurtosis is being compared to a theoretical L kurtosis for a candidate three-parameter frequency distribution.

Page 28: first paragraph under "Monte Carlo Simulation". The authors are discussing the means of L-skewness and L-kurtosis. Could they provide an equation showing how GZ was determined?

Response: Thank you for your comment. It is not our intent to reproduce the mathematical basis for procedures that can be found in other references. The equations and ratios used for this Atlas are explicitly defined in text and in the captions, so we do not feel additional segregated equations are necessary. However, we reviewed and edited

this paragraph to make it more clear.

2.11 Page 29, Table 4.5.1 – Is GUM the Gumbel distribution? Is NOR the Normal distribution? On page 28, there is no reference to any distribution with the abbreviation GUM or NOR. The Gumbel and Normal are two parameter distributions and it is surprising that a two parameter distribution would outperform the candidate three parameter distributions.

There does not appear to be any description of the candidate frequency distributions. There should be a short appendix where these distributions are defined. This description could be taken from Hosking and Wallis (1997). It should be noted again here that the selection of the distributions is based on the AMS.

The logic for choosing the "best" frequency distribution is not obvious. There are three goodness-of-fit tests and sometimes the frequency distribution chosen is not the best for any of the tests, for example, daily regions 11 and 18. For daily region 11, the GNO distribution was selected, but it not the best distribution by any test and the GZ statistic is not statistically significant. Same comment for the GEV distribution for daily region 22. For the hourly regions, the GEV is always selected and it is often not the best distribution on the basis of the three goodness-of-fit tests.

There must be some other criteria being used to select the frequency distribution that is not obvious to the reader. Was it intended to have the same distributions for a given region of the study area? This additional reasoning in selecting the distributions should be provided. For the hourly regions, perhaps the National Weather Service decided the GEV distribution was best overall. Granted there is not a lot of difference in the goodness-of-fit statistics and the quantiles probably do not differ significantly. However, given only the goodness-of-fit tests it is difficult to determine why some distributions were selected.

Selection of Appropriate Probability Distributions: I'm out of my expertise here, but I don't understand the contents of pages 29-38. These pages provide a table that indicates in a region-by-region fashion the performance of different frequency distributions as determined from Monte Carlo simulations, the "real-data-check" and the RMSE test. The table shows top three performing distributions for each test. My understanding of this is that the authors were trying to determine what frequency distribution was most appropriate for each region. I would assume that the top performing distribution as measured by these three tests would then be selected as the frequency distribution to be used for that region. This doesn't always seem to be the case, but I don't see any discussion of this point. Take for example Region 9 on page 30. The GNO (Generalized Normal) distribution is the top ranked distribution for all three tests in this region, but the GEV (Generalized Extreme Value) distribution was selected. Scanning this table I see many other examples of this. I assume the authors have a good explanation for this, but if it was offered in the text I missed it. I think the authors need to discuss more clearly how they selected frequency distributions in cases where the "best" distribution was unclear because of the tests.

Response: First, you are correct that GUM is the Gumbel distribution, and NOR is the Normal distribution. The two-parameter distributions GUM and NOR have a fixed L-skewness value (L-Cs=0 for NOR, which is a special case of PE3 and GNO; L-Cs=0.1699 for GUM, which is a special case of GEV). For a region, the GUM or NOR could be selected by Monte Carlo Simulation, one of three methods for goodness-of-fit, when the calculated regional L-Cs happened to be located at near 0.1699 or 0.0, but they never appear to be the first choice. In general, the 3-parameter distributions were preferred over the 2-parameter distributions because both theoretical and practical frequency analysis has shown that 3-parameter distributions are more stable and flexible when modeling the data. It should be kept in mind that there is no theoretical way to prove which frequency distribution governs the data. When a 2-parameter distribution was selected by the goodness-of-fit computation, it was normally ignored, since they are special cases of the other 3-parameter distributions, and the next 3-parameter distribution would be considered instead. The Tables of goodness-of-fit will be recompiled omitting the references to the 2-parameter distributions.

The three methods used for goodness-of-fit help provide a statistical foundation for the selection of a distribution. However, the final choice was not directly bounded by the three methods. Since the nature of observed data are complicated and a regional approach is used for frequency analysis, the smoothness or continuity of quantiles both in time and space in a large area was a concern. To reduce bull's eyes and/or gradients in precipitation frequency estimates between regions, the selected distribution by the three methods was sometimes changed during a review of results on a macro-scale. An effort was also made to maintain some consistency of selected distribution from region to region. Normally, the change was made to the next higher or lower distribution following the order of the behavior of the tail of the distribution in terms of L-kurtosis. The order of the distribution in terms of L-kurtosis from high to low is: GLO, GEV, GNO, PE3, GPA for L-skewness between 0.1 and 0.3, and GLO, GEV, GNO, GPA, PE3 for L-skewness between 0.3 and 0.5. For example, when the GEV was ranked best by the three methods, it may have been changed up to GLO or down to GNO during the final review of quantiles on a macro-scale to make quantiles smooth. An effort was made to keep such changes to a minimum. A test of sensitivity was conducted before making any changes. The change in distribution was only allowed when it caused a less than 5% change in the 100-year precipitation frequency estimates and was generally driven by only a few stations.

Generally speaking, there is no theoretical evidence to suggest that all regions in a large area should or should not be fitted by a unique distribution. However, the pattern of selected distributions in a large area should be explainable by both local climatology and observed data in combination with the statistical behavior of distributions selected.

Thank you for pointing out that the method is not clear in the documentation. The paragraph just prior to the table touches on our reasons for the choice of frequency distributions. To make it more clear, a heading has been given to that paragraph ("Selecting the most appropriate distribution") and text has been added providing more detail.

2.12 Page 38, 4.6.1 Regional growth factors – It would be helpful to state that the mean annual maximum flood is the same as the first L moment. Are not the Regional Growth Factors (RGFs) simply estimated by taking the ratio of the quantiles for the frequency distribution of choice to the mean annual maximum flood? Is this the same as "applying appropriate higher order moments of the selected probability distributions for a region to the dimensionless distribution function"? This latter sentence is confusing to me.

In section 4.6.1 the regional growth factors were described. However, there was no discussion as to why this approach was used as compared to making the calculations at all durations and frequencies directly from the data. I think I know the answer but it's not clear from the documentation.

Response: Regional Growth Factors (RGFs) are not simply estimated by taking the ratio of the quantiles for the frequency distribution of choice to the mean annual maxima flood (precipitation). Rather, regional growth factors are estimated through a set of regional parameters for a dimensionless frequency distribution. The regional parameters are weighted averages, proportional to data size, of parameters at the N sites in the region. The parameters at each site are first estimated based on the observed data rescaled by its mean of the site. In short, to get regional growth factors, first rescale the mean at each station and calculate the station L-statistics; then calculate the weighted averaged regional L-statistics; then calculate the regional parameters of the dimensionless probability distribution (of the selected distribution) from the regional L-statistics; finally, estimate the regional growth factors by calculating the quantiles of the dimensionless probability distribution. Hosking and Wallis (1997) provide an extensive treatment of regional growth factors and the index-flood procedure in Section 1.3. The text was modified to make this more clear.

Section 4.2.1, the Overview for the "Regional approach based on L-moments" section briefly provides insight into the advantages of using a regional index-flood approach. The regional frequency analysis using the method of L-moments was used because it assists in selecting the appropriate probability distribution and the shape of the distribution, but precipitation frequency estimates (quantiles) are estimated uniquely at each individual station by using a scaling factor, which, in this project, is the mean of the annual maximum series, at each station. The

resulting quantiles are more reliable than estimates obtained based on single at-site analysis (Hosking and Wallis, 1997). The regional estimation adds robustness to the results, whereas calculating precipitation frequencies directly from the data at each station is less reliable.

2.13 Page 41, 4.6.3 Conversion factors for AMS to PDS – Although Chow et al., (1988) use the term annual exceedance series (AES) to refer to the N largest (where N is the years of record) rainfall values above a base, I think this is confusing. Why not define the partial duration series as being the N largest peaks above a base and just refer to this as the partial duration series (PDS). This would be more clear given you are using annual maximum series (AMS) to define the series of annual maximums. The use of both AMS and AES is a little confusing.

A logical question is – Since there is a unique relation between the PDS and AMS frequency estimates (equation at the bottom of page 41), why go through all the effort to define the frequency analysis in terms of AMS and then convert to PDS? The 1.58-year AMS frequency value is equivalent to the 1.0-year PDS frequency value and the 2-year AMS value is equivalent to the 1.45-year PDS value. If someone wants the 1-year PDS value for a design purpose, just use the 1.58-year AMS value. It seems like a lot of effort is expended in converting AMS to PDS and displaying those data in the atlas. Some motivation for this effort should be provided.

Response: Thank you for your comments. There has been confusion over the use of AMS (Annual Maxima series) and PDS (Partial Duration series) and AES (Annual Exceedance series) among engineering design communities and academic research field for years. PDS is commonly called POT (Peaks over Threshold) and all occurrences over a given threshold make up the series. However, in this Atlas and elsewhere, PDS was restricted to the highest N events where N is the number of data years of the given record. This special case of the PDS series is also known as AES (i.e., when the number of the top peaks equals the number of years of data). The term, AES is provided in NOAA Atlas 14 to offer a distinction regarding the data type that was actually employed in the project and distinguish it from the common PDS (a.k.a. POT). In previous NOAA publications such as Technical Paper 40, the precipitation frequency estimates were first calculated from AMS and then presented in PDS along with a table of conversion ratios to convert between AMS and PDS for frequent events. In these publications, the meaning of PDS actually refers to the AES and not POT.

The equation $T_{AES} = [\ln(\frac{T_{AMS}}{T_{AMS}-1})]^{-1}$ at the bottom of page 41 can be written as $T_{PDS} = [\ln(\frac{T_{AMS}}{T_{AMS}-1})]^{-1}$ for a general form, which is good and distribution-free for converting return periods between AMS and PDS. In derivation of the equation, there were some assumptions adopted [See P. 8-22, HANDBOOK OF APPLIED HYDROLOGY, Ven Te Chow, 1964; and P. 18-38, HANDBOOK OF HYDROLOGY, Maidment, 1993]. Because of the assumptions, we would need to ensure its validation to real data through substantial testing and so we did not use the equation to convert results for one type of data (PDS or AMS) to the other, preferring a more empirically-driven approach as suggested by Jon Hosking in personal communication. We provide the equation in the documentation as theoretical support that the empirically-derived conversion factors are well-founded. Given the nature of real data and the resulting internal consistency adjustments that may be applied to quantiles, we did not feel it appropriate to apply a pure mathematical computation which has inherent assumptions. Note also that the equation only gives the relation between the PDS and the AMS in terms of return periods, such as a 2-year AMS event is equivalent to a 1.44-year PDS event. It does not directly provide the relation between the PDS and the AMS in terms of quantiles such as the relation between a 2-year PDS and a 2-year AMS event. Therefore, the empirical ratios were calculated to more accurately convert between the AMS results and the PDS (or AES) results.

We studied the relation between the PDS and the AMS quantiles of the real data to develop a set of generalized conversion ratios that is applicable to the specified study area. We separately analyzed both AMS and PDS (AES) data for the Semiarid Southwest United States and for the Ohio River Basin area for daily data and several other durations (1-hour, 6-hour, 7-day). We found that the relation from 2-year up to 1,000-year between the PDS and

the AMS is quite stable and duration-free. The quantiles of the 24-hour duration were used to establish the following table for NOAA Atlas 14 Volume 2 (Table 4.6.2 in the documentation).

2-year	5-year	10-year	25-year	50-year	100-year	200-year	500-year	1,000-year
1.086	1.023	1.010	1.004	1.004	1.004	1.004	1.004	1.004

The PDS estimates were then generated by converting from their corresponding AMS results for all durations since the AMS results were considered to be more stable. To stay consistent with the previous studies such as NOAA Atlas 2 and to meet today's needs at lower frequencies, NOAA Atlas 14 is presented in terms of both PDS and AMS with PDS being the default.

2.14 Page 43, 4.7 Estimation of confidence limits – This section is very brief. My understanding of the estimation of confidence limits is as follows. The confidence limits are estimated from the same simulation data that was used for the heterogeneity and goodness-of-fit tests. This simulation analysis provides 1,000 estimates of the L moments that are then input into the frequency distribution of choice to provide 1,000 estimates of the quantiles. The upper (5 percent) and lower (95 percent) confidence limits are determined from these 1,000 quantile values. What impact on uncertainty does the choice of frequency distribution have?

Based on limited testing, it appears the distance between the estimated quantile and the upper confidence limit is less than the distance to the lower limit. For example, the upper 5% limit may be 9 percent greater than estimated quantile and the lower 95% limit may be 12% lower. Given that the quantiles (e.g., 100-year event) should have positive skew, it would seem that the plus error should be greater than the minus error.

The confidence limit estimates generally indicate that the 90 percent confidence interval is on the order of plus and minus 10 percent for the 100-year event (based on a limited number of examples). This implies that we know the true 100-year rainfall within 10 percent 90 percent of the time. This seems to be an underestimation of the uncertainty in the rainfall estimates. The confidence limits only consider the uncertainty in the estimation of the L moments (that translates to the quantiles) and do not include the errors in measuring or observing the rainfall events. This should be stated. Also, what about the errors in spatial interpolation? In Section 4.8.1 a mean standard error of 10 percent is reported for estimating the mean annual maximum grid. It does not appear that this uncertainty is included in the confidence limit computations.

Response: The quantiles are the estimates that were obtained through an approach based on observed data as described in Hosking and Wallis (1997). The 90% confidence limits provide a measure of uncertainty to show how the quantile estimates could vary at the 90% confidence level when an underlying frequency distribution was selected. We did not study the impact on the uncertainty between different distributions.

In our procedure, the confidence limits calculated through Monte Carlo Simulation are adjusted for two reasons before they become the final: first, parallel adjustments corresponding to the internal consistency adjustment of the quantile estimates (Section 4.6.2); second, adjustments as a result of the internal consistency check for the confidence limits themselves. Two methods have been used to investigate the 90% confidence limits: 1) the upper 5% and the lower 95% of the calculated 1,000 quantile values, 2) the (mean+1.64 SD) for upper and the (mean-1.64 SD) for lower under an assumption of Normal distribution for the variation about the mean. We found that the results from the two methods are very close to the calculated, unadjusted upper and lower confidence limits.

In the Monte Carlo Simulation, the four-parameter Kappa distribution was used during the generation of the 1,000 synthetic data sets at each station. The distribution originally selected for the quantiles in that region was used to estimate the quantiles in each trial. While the confidence limits do not include the error associated with mis-specification of the appropriate distribution to model the precipitation frequency estimates, we believe that not only do our techniques for selecting the most appropriate distribution avoid this error, but the use of the Kappa distribution for the simulation of the synthetic data provides a flexible distribution that incorporates several of the

three-parameter distributions we use for modeling the underlying distribution thus minimizing the error. The inherent difficulty in assessing the "true" distribution for precipitation frequency has long been debated and is not easily resolved even with state-of-the-art methods. Please see Sections 6.4 and 7.5 of Hosking and Wallis (1997) for more details on the generation of confidence limits.

The confidence limits lay nearly equidistant and reasonably close to the estimates on both sides. However, we found that after the adjustment procedures some confidence limits unevenly distributed on the two sides of a quantile at some stations like the ones you have found, in which the upper limits were correctly estimated and adjusted while the lower limits were calculated correctly but may be adjusted incorrectly. This may very well be related to an assumption that the confidence limits are both upward sloping with the quantiles from 2-year through 1,000-year at a station. The assumption of upward sloping values is appropriate for all the cases of the quantiles, as well as the upper confidence limits, but may not be good for the lower confidence limits at some stations. We plan to review the lower confidence limits computation. A new internal consistency adjustment that fits either the upward sloping or the downward sloping will be studied and the lower confidence limits will then be recomputed. Thank you for your valuable comments.

Note that the quantiles may not necessarily have a positive skew (i.e., that the plus error may not be greater than the minus error). Our investigation for many cases of data has shown that the assumption of positive skew is not true. We have found no theoretical evidence to prove that the quantiles should have positive skew.

You are correct. The confidence limits are for the quantiles only while making the assumption that the data and measurements have been fully quality controlled. They do not include errors associated with the spatial interpolation. We have added text to the documentation clarifying this.

2.15 One persistent problem with the documentation as it stands is that much of the techniques used by the authors centered on the use of ratios. This is fine, but many times I am not entirely certain I know which is the numerator or denominator in the ratio. The same could be said of many of the statistical tests used by the authors. It would be extremely helpful if the authors could present more equations throughout the report explaining very explicitly how particular quantities were calculated. I will to refer back to this comment with specific examples as I go through my list of small/minor comments.

Page 11: Table 4.1.3 gives N-minute ratios. Could the authors include an equation that explicitly defines these ratios?

Page 12: Same comment for Tables 4.1.4 and 4.1.5. Ratios need to be defined.

Page 40: Table 4.6.1 and surrounding text. Please provide the equation that defines the ratios presented in this table.

Page 41: Figure 4.6.2 and surrounding text. Again, I'd like to see the equation that gives the ratios presented in this figure.

Page 53: Table 4.8.3 – please provide equation that defines the ratios presented in this table.

Response: Thank you for your comment. It is not our intent to reproduce the mathematical basis for procedures that can be found in other references. The equations and ratios used for this Atlas are explicitly defined in text and in the captions, so we do not feel additional segregated equations are necessary.

2.16 The authors talk of the ratio 1.13 for the daily versus 24-hour values. Please give equation.

Page 14: I'm getting confused because the 1.13 referred to in comment 7 [stated above] corresponded to the 24 hr vs 1 day observations. Here the authors quote 1.16 for the 1.16 for the 1 hour and 1.05 for the 2 hour. I suspect I'm

confusing readings taken on minutes, hours, and days, but again, defining these ratios with an explicit equation and tabulating them clearly would be helpful. Placing all this information in text forces some very careful reading and leaves me uncertain that I'm understanding exactly what is being discussed.

Response: To help clarify, we have added a table.

2.17 Page 17: Heading refers to daily annual maximum time series. I'm assuming this means daily collected values and not 24-hour values? Yes. I think the terms "daily" and "hourly" (or whatever would be more appropriate) could be added to the glossary. I notice that "n-minute" is defined so these two periods should also be defined or am I missing something?

Response: We make the assumption that our users have a sufficient background knowledge to make the distinction between daily and hourly precipitation measurements. Also, when discussing durations throughout the text, we've tried to be clear by including text such as "daily (24-hour through 60-day)" to indicate the applicable durations. The term "n-minute" was defined because it is less clear what type of data/duration it is.

2.18 Page 18: if there was a block of missing years in the period of record for a gage, was N affected in any way? If only partial years are used, how is N calculated (rounding or truncating)?

Response: The N used in calculations during the analyses was the total number of data years (i.e., years for which an annual maximum was extracted) regardless if there were missing years in between. All stations used in the analysis were screened for large gaps of missing years (see Section 4.3) that created inconsistent data samples. During the data series extraction, criteria were applied so that an annual (or partial duration) maximum was extracted only if a year had sufficient data (see Section 4.1.3). If a year passed these criteria, then it was counted in N.

2.19 Page 20: bottom: You talk of using a statistical t-test for population testing. Can you present the equation for this? What level of confidence was used?

Page 21: Annual maximum series screening section: Again, could the t-test equation be laid out? What level of confidence was used?

Response: Thank you for your comment. It is not our intent to reproduce the mathematical basis for procedures that can be found in other references. The equations and ratios used for this Atlas are explicitly defined in text and in the captions, so we do not feel additional segregated equations are necessary. That said, however, we have added the t-test equations to our glossary and added the level of confidence where appropriate in the text. The equation of the t-test used for testing whether a difference between means of two samples is significant is shown here:

$$t = \sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}} \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{n_1 s_1^2 + n_2 s_2^2}}$$

, following a Student's t distribution with (n_1+n_2-2) degree of freedom.

Here, $\bar{x}_1$ and $\bar{x}_2$ are the means for sample 1 and sample 2, respectively. s_1^2 and s_2^2 are sample variances. n_1 and n_2 are sample sizes. At 90% confidence level (or significance level α = 10%), reject H_0: the means have no significant difference if $|t| > t_{n_1+n_2-2,\alpha/2}$. The t-test is a common test found in most statistical texts (e.g., STATISTICS FOR BUSINESS AND ECONOMICS, Paul Newbold, 1988, pp. 365-366; HANDBOOK OF HYDROLOGY, Maidment, 1993, p. 17.23; Lin Shaohong, BASIC PROBABILITY AND STATISTICS, People Education Press, Beijing, 1980, P. 160; COMMON STATISTICS METHODS, Mathematics Institute of Chinese Academy, Beijing, 1973, p. 24).

2.20 Page 21: Discordancy section: Can you give the equation used for calculating the distance between the cluster center and the unique L-measures for a particular station? (See comment 1). Is this measure suggested by Hosking and Wallace? If so it should be cited. If not, do you give equal weight to each of the Lmoments? Is this appropriate?

Response: The reference to Hosking and Wallis is provided in the text. In Section 4.2.2, we state that "Details and equations for the analysis may be found in other sources (Hosking and Wallis, 1997; Lin et al., 2004)." Our intent is not to re-publish material found in Hosking and Wallis, particularly the equations, and so we offer a brief verbal explanation of what the discordancy measure is.

2.21 Page 21: Bottom paragraph: Another t-test is mentioned. I don't want to be a pain about this, but I again think that the equation for this t-test needs to be provided along with stating the chosen level of significance used in the test.

Response: We have added the definition to the glossary. The t-test used to test for population correlation is shown here:

$$t = \left| \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \right|$$, following a Student's t distribution with (n-2) degrees of freedom

(Paul Newbold, STATISTICS FOR BUSINESS AND ECONOMICS, Prentice Hall, 1988, p. 441). At 90% confidence level (or significance level $\alpha = 10\%$), reject H_0: there is no correlation or the correlation is not significant at significance level of 10% if $|t| > t_{n-2,\alpha/2}$.

2.22 Page 22: Top paragraph. Please present an equation that defines relative error. It's not clear to me what quantity is being used to define relative error.

Response: Quantiles were calculated using all stations regardless of the whether or not they were cross correlated. A second set of quantiles were calculated using only stations that were not cross correlated with other stations. The relative error for the region was calculated by comparing the "results", which means comparing the final quantiles or RGFs of both sets. Text has been added to make this clear.

2.23 Page 22: Section 4.4 second paragraph: The authors mention that a threshold of 2 is conservative and reasonable. Could the authors present the equation for determining H1?

Page 22 - Why not provide the equation defining the H1 measure? This would give the reader a better understanding of the homogeneity test.

Response: The reference to Hosking and Wallis is provided in the text. In Section 4.2.2, we state that "Details and equations for the analysis may be found in other sources (Hosking and Wallis, 1997; Lin et al., 2004)." Our intent is not to re-publish material found in Hosking and Wallis, particularly the equations, and so we offer a brief verbal explanation of what the heterogeneity measure, H1, is.

2.24 Page 23. Item number 3. It's not clear to me if the "X" was intended to remain a variable in this sentence, or if the authors intended to come back and replace "X" by an actual number. If it was the former, then I find the sentence difficult to understand and wish it could be re-written. If it was the latter then the "X" needs to be replaced with the intended actual number.

Response: The "X" was meant to indicate the various number of days for the different durations (i.e., 4 days, 7

days, 10 days). It has been changed to "a given number" to make the sentence more clear.

2.25 Page 23: Item number 4. How was the value of 174 determined? Provide calculation. Also, since 223 errors were found rather than 174 does that suggest that something more is going on? This deserves a little more explanation.

Response: If you take the number of stations in the Ohio Project (2,784) and multiply by the average of 55 years of data, you get roughly 150,000 station-years. One might expect to find 150 1,000-year annual exceedances of any duration. Assuming a Poisson distribution and a mean of 150, the 95% probability mass falls within the interval of 127 to 174. Given the close proximity of some stations, you might expect this number to be higher. Because the regions were derived using 24-hour data, they may be less applicable for 60-day data. Thus, there may be slightly more than expected at the longer durations. However, given that the regions were statistically homogeneous and given the limited use of 60-day data and the resource constraints, we did not refine the regions specifically for the 60-day duration. We have modified the text based on your comment.

2.26 Page 26: 2nd bullet in 1st group: What is the definition of "enough" data years? 5? 10? 50? Not clear.

Response: That is a good point. These at-sites stations had at least 100 data years, but 50 years was the absolute minimum for consideration. The text has been changed.

2.27 Page 26, At-site stations – It is surprising that at-site frequency analyses were used for 2 stations out of 2,846 stations. It seems the improvement in accuracy is not warranted given the additional effort. The third bullet in the second group of bullets on page 26 should indicate the RMSE was more not less if the station was included in the regional analysis.

Page 26: I understand the need for "At-site" stations. Do I understand correctly that ultimately there were only two at-site stations A1 and A2, and that only one was actually within the ultimate prediction extent?

Page 26 - Table 4.4.1. It is interesting that two of your at-site stations 11-6610 and 22-1880 had very long records that did not seem to fit with surrounding stations. These stations were obviously examined carefully.

Response: Effort was made to keep at-site stations to a minimum because they can lead to spatial bull's eyes or steep gradients in the maps. There were only 2 stations that were analyzed as at-sites within the project area with only one of those falling within the published core area. After extensive investigation and thorough quality control, these stations were made at-site stations to account for observed extreme precipitation regimes that the regional method could not resolve. It may not make a whole lot of difference generally but it likely is important at the local level. The information presented by the data at these stations was sufficiently different from the surrounding stations to retain this level of local detail at these locations.

Their long records enabled them to produce accurate estimates independent of a region. The ultimate goal of putting a station as an at-site was to reduce the error in the estimates. Regarding the text in the bulleted list, the RMSE of the region generally became less when the station was no longer included in the region. The bullet was reworded to make its meaning more clear.

2.28 Page 38: Section 4.6.2. Introductory paragraph. Please make it clear that examples of practical adjustments will be discussed one-by-one below. At first I thought that just the first paragraph dealt with practical adjustments and I didn't know what this meant.

Response: Good suggestion. We added some text to that effect.

2.29 Page 47: top text box below the two boxes with figures: What are the sources of extraneous "noise" the authors refer to? I don't see why the smoothing process should produce noise.

Response: The extraneous "noise" here refers to the unnatural spatial division between different smoothing levels. Based on your question, we removed this reference to avoid confusion.

2.30 Pages 45-51 or so: Can you write out representative equations for each of the steps enumerated? This is done for step 3 but not the others. It would help to understand the contents of each of the grids produced.

Response: The equations are explicitly defined in the text and in the captions, so we do not feel additional segregated equations are necessary. The remaining steps only involve simple arithmetic calculations and one equation, the linear slope equation (y=mx+b).

2.31 Page 54: top paragraph: the authors say the upper/lower limit grids were slightly less stable. Less stable than what? I assume less stable than the mean grids? Please clarify.

Response: Yes, less stable than the mean grids. We have clarified the text.

2.32 Page 54, bottom paragraph: This 1% adjustment is a practical need and I understand its use. Could the authors state what percent of cells required this adjustment so we have any idea of how prevalent this problem was?

Response: Since these adjustments were so rare, a quantitative assessment was not conducted.

2.33 Page 56: Top paragraph: Could the authors present the equation used to determine the ratios at the co-located stations and then the equation used to apply this ratio to the daily-only 24-hour precipitation frequency estimates? This would be helpful.

Response: Text was added pointing the reader to the IDW equation in Section 4.8.2. The ratio is explicitly described in the text.

2.34 Page 67: "Exceedances" paragraph: Last line: Why is it "at least" 10 events. Shouldn't it be "approximately" 10. We're talking about statistics, I don't see why there should be certainty of at least 10 stations satisfying this query.

Response: Yes, good point. The text has been changed.

3 Technical questions

3.1 Page 66 - Areal reduction factors. How does one access NOAA Atlas 2 if it is not available in their office?

Response: NOAA Atlas 2 Volumes I, II, V, IX, X, and XI is available digitally from the HDSC web site: http://www.nws.noaa.gov/oh/hdsc/currentpf.htm

3.2 Page 66 - Section 7. Reduction from point to areal rainfall is a critical problem for investigation of drainage areas larger than 10 mi2 . Unfortunately, many of the large networks with extended periods of record have been abandoned. However, the amount of current radar data must contain a lot of point data that could be used to estimate areal distribution or variation. Although the point rainfall estimates from radar may be in error by a possible factor of 2, the areal distribution should still be reasonably estimated for all types of precipitation – convective, frontal, orographic, or cyclonic.

Response: Rocky Durrans of University of Alabama did a study for us and demonstrated that there were just too many spatial errors in the radar data for it to be used. Since he did the work, the radar rainfall estimation techniques have improved, so at some point we may be able to extract areal reduction factors. (Durrans, S.R., L.T.

Julian, and M. Yekta, Estimation of Depth-Area Relationships Using Radar-Rainfall Data, Journal of Hydrologic Engineering, ASCE, Vol. 7, No. 5, pp. 356-367, 2002)

3.3 In section 4.1.2 check out the reference to Weiss, L.L., 1964. Ratio of true to fixed-interval maximum rainfall. Journal of Hydraulics Division, ASCE 90: 77-82. He developed a theoretical framework for why the ratio works out to be 1.13 when going from 1-day to 24-hour rainfall and so forth. It provides a nice underpinning on why these different studies seem to come up with the same ratios regardless of location.

Response: Thank you for pointing out the reference. There is also a recent paper on the topic: C. B. Young and B. M. McEnroe, 2003. Sampling adjustment factors for rainfall recorded at fixed time intervals, Journal of Hydrologic Engineering, Volume 8 Issue 5, pp. 294-296.

3.4 These types of studies tend to get heavily used over a very long period. With the web service and ARC files, are there plans at HDSC to ensure that this information will be available for the next 15 years given the advances in computers. Other than ASCII files, I can't handle many of the media or file formats of 15 years ago.

Response: HDSC recognizes the use and value of the estimates over the long term and intends to maintain the estimates in usable formats.

3.5 Page 43 - Section 4.7. Although most engineers will ordinarily not use confidence limits in studies that require precipitation frequency estimates, there are occasions in which quantification of uncertainty is important. I assume that in your Monte Carlo simulations you used some type of random number generator that not only generated random numbers, but also entered the generator at equally random points. About 35 years ago, at Texas A&M, I had a student study random location of precipitation stations of various densities to reproduce known rainfall patterns. We soon found that random number generators are not always random.

Response: We’re of the opinion that as a profession we need to better account for the uncertainty in the precipitation frequency estimates. Not just from the point of view of use of the estimates to design in accordance with regulations, but also from the point of view of being prepared for failure of the design. In structural design we use safety factors as a catch all for a range of uncertainty in things from yield strength to load estimation. A storm water drain designed to regulation will fail with much higher probability than a building. When you look at the uncertainty in the estimates, it can be quite large in relative terms. For example the upper and lower 90% confidence limits on the 100 year 60 minute estimate at Tucson Airport are 2.71" and 2.09". Changing the way the profession accounts for this when rational method and SCS curve numbers are so ingrained would be a big deal. On the other hand, Pima County has chosen to regulate based on our upper 90% numbers and so there's some progress. We'd also like to see the profession move ahead from the simple methods of computation developed for pre-computer days towards some sort of Monte Carlo computation that accounts for uncertainty. There have been a number of papers on this in the Australian literature and there's more work to be done. But the concept isn't so difficult. We do a similar thing in ESP - compute the possibilities and examine the ensemble of results. If common tools such as the Corp's HEC-RAS and Hastead had this sort of sophistication built in then we don't think it would be so difficult to move the profession.

With respect to the random number generator, there are several types of random number generator. The random number generator that is used in our L-moments computation software is *Multiplicative Congruential Generator with Base 2**31-1 and Multiplier 7**5* (P.A.W. Lewis et al., 1969, *IBM Systems Journal*), which is a pseudo random number generator. This is an effective and common type of random number generator. Hosking adopted it in his original program for generating random numbers. The random number generator program was tested by HDSC twice in early 1990's and in about 1997 before it was coupled into the confidence limits computation program. It was tested for the effectiveness of the generated random numbers in representing natural observed precipitation data. During the test, time was spent consulting other references. No discrepancies were apparent during the tests and through the research. To further enhance the effectiveness of the generator, the seed, a major

parameter of the random generator, was changed from a lower number to a higher number as recommended in a reference. Hosking made the same change in his 1997 version.

3.6 In the introduction and appendix A3 are statements made about climate change. The treatment of climate change is the only scientific shortfall of this study. The scope of the trend analysis was fairly limited, confined to 1-day events. There is a whole body of scientific literature on precipitation changes past, present, and future. To make statements like climate change was negligible based on this study and that it is assumed "that there is no effect of climate change in future years on precipitation frequency" really aren't defensible. Also implied in the analysis is that any sub-regional trends are not important (e.g., the upward trends around Chicago). Anyway, I do not have the time to wordsmith the document but I strongly urge that it be made clear that the analysis is very limited and acknowledge that there is considerable research showing changes in precipitation over time and caveat emptor w.r.t. future applicability.

Response: You make a very good point. We have added text in the documentation to clarify our comments and conclusions regarding climate change. Our implication was meant to be that it was necessary to make the assumption that there is no effect of climate change in future years on precipitation frequency estimates due to limitations in the quantification of climate change in future years and based on our trend results of the 1-day annual maximum series. Admittedly, time and resources precluded us from taking our trend analysis further.

3.7 Page A.3-1, 2.1 Methods - Linear regression models and t-test were used in the trend analysis and the reference is to Maidment (1993, p. 17.30). In this context, it seems the t-test is really the use of the Student's t distribution for testing the significance of the regression coefficient and not the t-test for equality of means. Linear regression (trend) analysis on the observed rainfall amounts is effected by outliers so a regression on ranks or some other non-parametric test (like Kendall's tau) is better for detecting trends. This comment is intended for future analyses because it would be too time consuming to repeat the trend analysis with another approach for this study.

It is not clear how the linear trend in variance was tested. How was the change in variance with time evaluated? How was linear regression analysis used to detect a change in variance?

Response: Yes, the test used here for detection of a linear trend is based on the Student's t distribution – a simple linear trend model (*Handbook of Hydrology*, Maidment, 1993, p. 19.17), in which the mathematical expression for testing the hypothesis is similar to the Kendall's τ for independence test (*Handbook of Hydrology*, Maidment, 1993, p. 17.30) based on Student's t distribution.

We use a minimum of 50 data years for the analysis to ensure statistical integrity of the linear regression, however, your suggestion to use a non-parametric trend model is a good one. We may use it in our future work in addition to the linear trend model. In addition to the parametric linear trend model used here, several parametric and non-parametric tests of trend have been suggested in the literature, such as the Mann-Kendal Test for Trend as a non-parametric test (*Handbook of Hydrology*, Maidment, 1993, p. 19.17), which is based on number of signs (+ or -) by comparing the sequential magnitudes of a series.

The test for variance was created by HDSC as an in-house tool to provide additional insight into the behavior and quality control of the data. Linear trend in variance cannot be directly tested through a simple t-test model. Instead, it was indirectly tested by constructing a variance-related variable, an index of the square of deviation, or $v_i = (x_i - \bar{x})^2$ where, x_i is the AMS data for i = 1, 2, …, n - the data year at a station, and $\bar{x}$ is the mean of the AMS data. The index was then applied as a simple variable in the linear trend model of $v_t = a + bt$. Hence, v_t is a variable of time (year), or an indication of the variation of the deviation from the mean with time t (year t = 1, 2, …, n). Rejection of the hypothesis of b = 0 indicates a detection of a linear trend in the variance-related variable. From this an inference that there may be a linear trend in "variance" can be drawn.

3.8 Page A.3-1, Introduction – It seems the purpose of the trend analysis should be to evaluate the homogeneity of the data, not the long-term effects of climate change. Apparently the entire period of record was used for all stations even though a high percentage of significant trends were detected. Several states had over 20 percent of the stations with a significant trend (Table A.3.1). Some of the trends may be due to beginning data collection in a dry period and ending in a wet period or vice versa. For example, the 1930s were dry and the early 1980s were wet in the certain areas of the eastern USA. Several 50-year records covering the period 1930 to 1980 may show a trend due to the climatic variability noted above. Are the trends related to changes in data collection procedures or the gages being moved?

Bottom line on trend testing – The trend analysis did not have any impact on the period of record or the stations used in the analysis. It seems like more stations had trends (16.4%) than would be expected by chance (10%). Perhaps some reason(s) should be provided for the high incidence of trends and why this did not have any impact on the analysis. For example, does the change in data-collection procedures over time contribute to the number of significant trends?

Response: The focus of the study is on frequency analysis not climate change. Climate change is a long-term impact on precipitation and the representativeness of data is very important in precipitation frequency estimation. Because the stations that exhibited trends were scattered geographically in different regions and were in both the positive and negative directions, we did not consider the 16.4% as having a high impact on our analyses.

3.9 Page 67: "Climate change" – second paragraph. I have recently read a paper that suggested, at least in Albuquerque, New Mexico, that there was a trend towards more storms from 1948 to 2000 from about 30 to more than 40 storms. Please contact Pierce Runnels (prunnels@ascg.com or 505-247-0294) for more on this study. He was looking at very frequent, common storms while you are making statements about the bigger, less frequent storms. I assume it's possible that both your findings and his are simultaneously possible.

Response: Exactly. While such research points to an increasing trend in the occurrence of common storms leading to an increasing trend in mean annual precipitation, our data has not shown that there is a change in the statistics associated with estimating precipitation frequency.

3.10 Page 43, 4.8.1 Mean annual maximum – Regression analysis is a powerful tool but how could PRISM account for user knowledge? Isn't this an overstatement of PRISM's capabilities?

Response: PRISM is a knowledge-based system (KBS), a well-known discipline in computer science, that provides a useful framework for an effective way to combine the strengths of both human-expert and statistical methods. The PRISM KBS allows user knowledge/experience to be incorporated into the predictive linear regressions PRISM makes for each grid cell. Please see the paper listed below which describes the thinking and rationale behind the development of a spatial climate KBS, and presents working examples in the form of algorithms from PRISM.

http://www.ocs.oregonstate.edu/pub/prism/docs/climres02-kb_approach_statistical_mapping-daly.pdf

3.11 Page 54, Derivation of upper/lower limit precipitation frequency grids – In Section 4.7, Estimation of confidence limits, I thought the confidence limits were being computed independently of the mapping process and now I am not sure. The paragraph on page 54 indicates that the upper and lower limit precipitation frequency grids were also derived by the CRAB procedure. Sometimes the upper (lower) limit grid cell values were larger (smaller) than the mean values. Is this because the confidence limit computations do not take into consideration the mapping standard error?

Response: The confidence limits (i.e. upper and lower limit precipitation frequency estimates) are in fact computed independently of the mapping process, but the grids of the confidence limits are derived with CRAB. By

definition, the "mean values" which in this context refers to the precipitation frequency estimates, must lie between the upper and lower precipitation frequency estimates. So it is mandatory that the upper (lower) limit grid cell value be larger (lower) than the mean values.

3.12 Spatial Interpolation Methods: This is my largest concern with the authors' methods. My concern is the projection/coordinate system used by the authors in performing their spatial interpolations. With the exception of two figures in Appendix 4 (A.4-23, and A.4-24) all of the Ohio River Basin images I've seen and all the text I've seen seems to show the exclusive use of geographic coordinates (latitude, longitude). My concern is that the spatial interpolation actually took place in the geographic space. If this is the case, then there is a bias that is introduced. True distances measured in the North-South direction are under-estimated relative to distances in the East-West direction if distances are measured in degrees or arc-seconds. In other words, space is being distorted. Look at page 51 of the report for example. This image is in geographic coordinates and you can see the effect of distortion (though subtle) on the shape of Washington DC. DC is bounded by right angles with the State of Maryland but you can see that the shape of DC in this figure does not have right angles for the DC border. Looking closer at this figure you should note that east-west distances are greater than north-south distances than if you compare to a figure like the one on page A.4-23. The extent of this distortion grows with increasing latitude. At mid-latitudes the true distances are about 4 units north-south to 3 units east-west for an equivalent amount of degrees or arc-seconds.

My concern is that if the spatial interpolation measured distances (see the r's in the equation by Neteler and Mitasova on page 48 in the IDW scheme) in units of degrees or arc-seconds that this introduces a bias in the interpolations. If this is the case, a gage that is located a distance, x (measured in geographic coordinates), due south of an arbitrary location will have an equal impact on estimates at that location as a gage located the same distance, x, due east of the location. In reality the gage to the east is closer than the gage to the south and therefore should be given more weight. I'm afraid that if the interpolations were done in geographic coordinates I'd have to argue that all interpolations would need to be redone in a projection that preserves length and area measures such as an Albers or Lambert projection – there are several that could reasonably be chosen. I could work up an example of how this distortion would affect answers if that would be helpful, but I think Geoff or others at NWS should know or be sure to find out how the spatial interpolations were done.

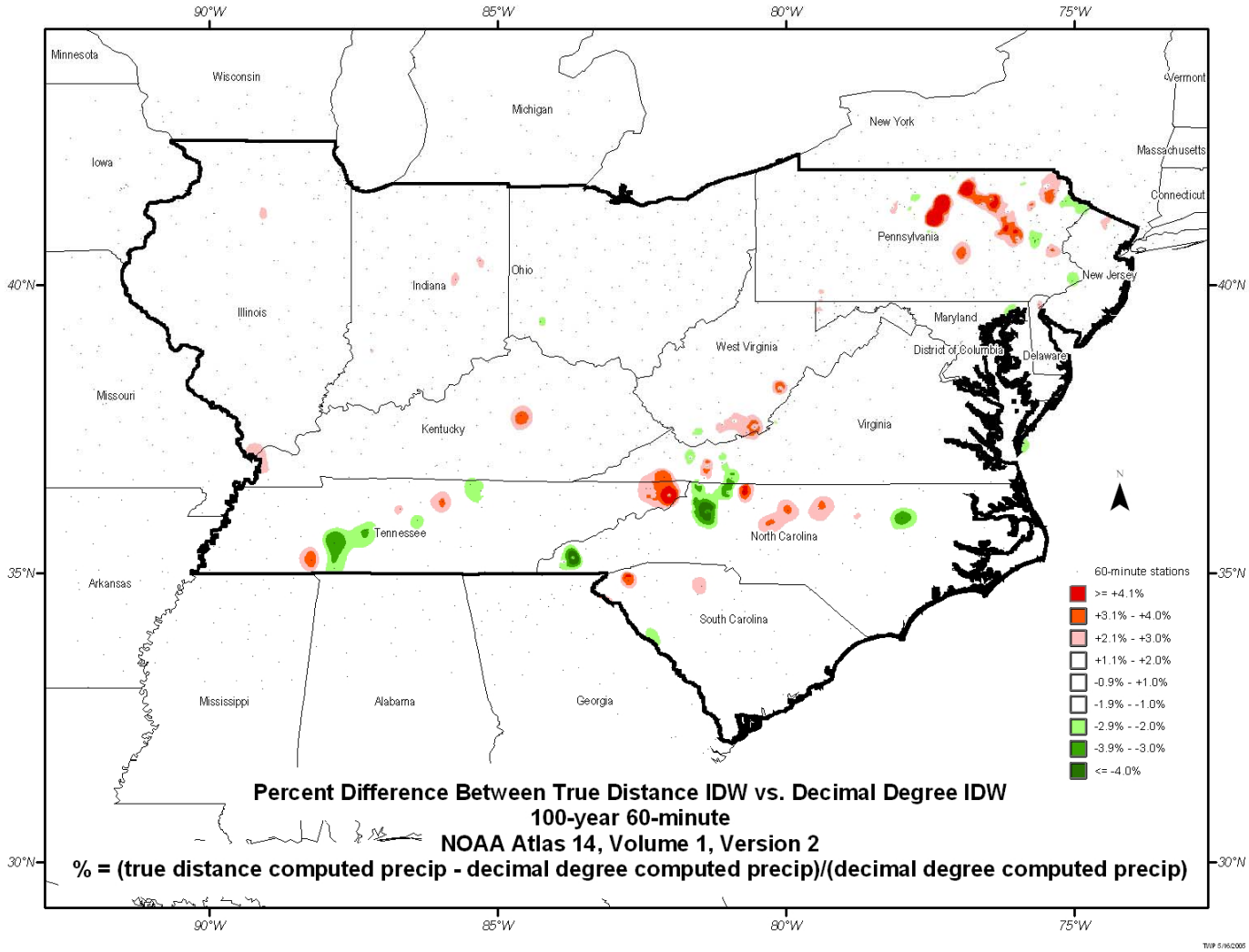
Response: PRISM is free of geographic distance distortion issues because whenever decimal degree distances are computed, they are corrected for convergence of the meridians (e.g. longitude). The Spatial Climate Analysis Service (who ran PRISM on the "index flood" grids) made a conscious and justified decision to use a geographic projection and we adopted that convention in the derivation of the precipitation frequency grids for the following reasons:

(1) It is the universal projection from which all other projections are derived. There is significant error associated with reprojecting from one grid to another. The fewer steps needed to get to the desired projection, the better for users reprojecting the precipitation frequency grids.

(2) Because CRAB's (Cascade Residual Add-Back) predictive PRISM grids of mean annual maximum ("index flood") are given in a geographic projection, it makes sense to produce CRAB output in the same projection as the input, predictive grids as well.

(3) Geographic projections do introduce unequal cell sizes (i.e. the grid cells are taller than they are wide), but this really a pretty small problem. We actually used the unequal cell size to our advantage when we recognized that the spatial scale at which westerly-produced precipitation operates seems to be smaller in the east-west direction than in the north-south direction. This is because terrain features that interact with these westerlies produce more accentuated rain shadows in the east-west direction. This essentially means that we attain a slightly higher resolution in the east-west dimension than in the north-south dimension.

CRAB uses GRASS's inverse distance weighting (IDW) function (s.surf.idw) to distribute the point normalized residuals to a grid. This function does in fact use distances computed in unadjusted decimal degrees which introduces some error into the CRAB-derived derived precipitation frequency grids. In order to assess the inaccuracy associated with this in the precipitation frequency grids, we ran the derivation of the 100-year 60-minute grid using a different function, r.surf.idw. This particular function calculates its distances along a geodesic (a.k.a. a great circle), however each of the points must first be converted to a grid cell, a 30-second by 30-second grid cell in our case. In other words, each point is converted into a grid cell while the surrounding grid cells are set to missing, then r.surf.idw interpolates values for all of the missing cells. Depending on where the station resides with relation to the grid cell boundaries, some distance error is introduced at the onset; to what degree is unknown, but it is known that GRASS georeferences each grid cell with the upper left hand coordinates of the cell, not the center. Although r.surf.idw assumes the earth to be a perfect sphere, its distance calculations better reflect true ground distances than unadjusted decimal degree distances. The map below indicates the difference (in percent) between the decimal degree and true distance weighting of the normalized residual for NOAA Atlas 14 Volume 2 100-year 60-minute precipitation. Keep in mind however some of these differences are associated with distance error associated with grid to grid distances, instead of point to point.



The differences are relatively small and all within about 5% of the already published 100-year 60-minute precipitation estimates and also for the 100-year 24-hour (not pictured). The percent change equates to precipitation values on the order of a few hundredths up to .20 inches. The red/orange areas indicate places were the 100-year 60-minute estimates would go up with a true distance IDW, while the green areas would decrease.

As an alternative approach to using r.surf.idw, is to change the s.surf.idw source code to use the same geodesic distance logic as r.surf.idw. Although there was not time, this approach would provide an "apples to apples" comparison. Consideration could also be given to using the latest version of GRASS, version 6.0, which has more power and functionality than the older version of GRASS used for NOAA Atlas 14 Volumes 1 and 2. It is possible some of these issues could be resolved in already built-in functions.

In summary, the distance calculations in the current version of CRAB do not correctly represent true distances therefore the weighting of the normalized residual is imprecise. Initial indications suggest that this could cause up to +/- 5% change in a very small number of the currently posted 100-year 60-minute grid cell values. Considering all the factors involved, we will thoroughly investigate the implications of this and make a judgment on how to proceed. Regardless, some level of spatial error will exist.

3.13 Page 52: top paragraph. The authors write of smoothing between northern and southern regions. Why is this necessary? I'm concerned that this is an artifact of the use of geographic (lat/long) coordinates.

Response: This issue is completely independent of the projection. Although subtle, the northern and southern regions have different ratios which are applied to the 60-minute precipitation frequency estimates to compute the n-min (5-, 10-, 15- and 30-minute) precipitation frequency estimates. The delination between north and south is not a distinct line, but rather a transitional zone. In order to create this zone, smoothing of the values was done to provide a climatologically sensible transition from one region to the next.

3.14 Page 54 – Figure 4.8.5 – it is good that the majority of stations show very small percent difference but the asymmetry of this figure suggests that there is a greater tendency to under-represent the information at a location rather than over-estimate this information. Is this the correct interpretation of this figure? This should be stated. Also, smoothing tends to reduce local peaks which may be what this figure is saying. Does this suggest that too much smoothing was applied? The authors should at least discuss this. I expect there is a trade-off between over smoothing the data and accepting local anomalies and it's very likely the authors have struck a favorable balance between these two, but this should be stated.

Response: This figure tells us several things, with one being that there is a greater tendency to under-predict the precipitation frequency value at a location in a station's absence. However, the primary message we are trying to convey is the fact that overall, CRAB does a good job at reproducing the values in the absence of station data. We have added text to reflect this point. Although smoothing does contribute to some of the differences this figure represents, it does not suggest too much smoothing was applied. And yes, we've tried to maintain important local anomalies, but yet create spatially smooth and consistent results. In an effort to do this, we have applied less smoothing where local anomalies are expected and observed (e.g., complex terrain) and more smoothing where the climate is more uniform (e.g., flat, open terrain). See section 4.8.2.

Page 50: Section 5.3. The authors talk of calculating areal estimates. This is precisely where the spatial interpolation methods matter. I assume and hope that the areal estimates that the PFDS provides are done in a coordinate system other than geographic (even if the user input is a series of lat/long coordinates). At a very small scale, I doubt this matters much, but there will be differences at larger scales (e.g. the 500 square miles figure provided in the documentation on this page) and where the gradients in the data are steep.

Response: The PFDS areal functionality is still under development, but the area and distance issues will be critically evaluated and tested before anything is publicly released.

3.15 Page 63: Figure 5.4.4.: Two things:

a. Large comment: Did I miss it, how were month-based estimates developed? It seems like the authors needed to

apply all the same techniques as described elsewhere, but now as filtered on a monthly basis. This can't be the case, but I don't know how the values presented in the sample figure were determined. Please clarify.

b. Small comment. The graph shows some sort of spline function fitting data between months. It would be helpful if the horizontal axis were 13 months long so that January appears at beginning and end. Otherwise, it's impossible to interpolate values in the month of December.

Response: a. The seasonal exceedance graphs indicate the percentage of observed events exceeding the corresponding annual exceedance probability for the specified duration in each region. They are not seasonal precipitation frequency estimates. We did not compute seasonal precipitation frequency estimates. The exceedances are presented as a tool for the user to estimate during which month an exceedance is more likely to occur. The explanation of the method is presented in Appendix A.2.

b. We have added the month of January to the graphs in the document and on the PFDS. However, keep in mind that the seasonal exceedance graphs were based on monthly data and thus the interpolation down to a resolution less than a month is not recommended.

4 General comments

4.1 Page 19 - Although I have never used the regional approach based on L-moments, the concept of utilizing regional groups that result in dimensionless frequency distributions common to n stations in a region is appealing. I found the discussion under 4.2, 4.3, and 4.4 well written and of interest.

Response: Thank you for your comment. We feel that the benefits of the regional approach based on L-moments are many. The approach produces much more accurate estimates as well as providing the functionality to generate confidence limits.

4.2 I like what you have done with the update, and appreciate how much work went into the document, but I wish that you had continued it into Georgia, since you used Georgia data to develop it and earlier versions included Georgia in the maps. Is this a funding issue or was there another reason Georgia was not included?

Response: It is indeed a funding issue. Our work is not directly funded from the NOAA budget. Rather it is done at the request of and with funds provided by users. NOAA Atlas 12 Volume 2 was funded through the Ohio River Basin Commission by its member states along with contributions from USACE, TVA, FEMA, and NRCS. We've presented the need to update the full U.S. in a number of forums. In particular, the Federal Subcommittee on Hydrology who agreed that (a) it should be done, (b) it should continue to be done by the Federal Government, and (c) it should continue to be done by NOAA's NWS. However, funds have not been made available by the agencies who use the data for completing a national update. Some months ago we spoke with folks in Atlanta about that fact that NA14 Volume 2 non-core area included Atlanta where large amounts of money are to be spent on the storm sewer separation program. Our preliminary results indicate the TP-40 precipitation frequency estimates in that area would be reduced by between 10-20%. That should save quite a bit of money, orders of magnitude more than what extending the study would cost. We didn't get any interest. Anything you could do to promote interest in funding a national update would be helpful.

4.3 I'm getting the same data retrieval for the IDF and annual maxima for the Sellersville and Perkasie gages in PA. Shouldn't these yield two different sets of data. Also, the text version of the PIE tables is not working.

Response: The high frequencies, low duration events at these two locations are similar, but there are some differences. These stations are located near each other, so it is not surprising that they are similar. This is the first time we've been told of trouble with the text version function. Apparently, depending on the web-browser that you are using you will either be prompted to save a file with the name "download.perl" or another name with a .perl

extension, but either way the file is a comma-delimited ASCII text file. We recognize this is confusing and we are working on making this functionality more user-friendly and allow users to provide their own save-as filename.

4.4 We understand the PRISM was used to interpolate the rainfall data from the various stations and develop the isohyetals for the various recurrence intervals, the Atlas. We also understand that not all the available data were used in the analysis. It would be interesting to know if there were checks of the resulting isohyetals at the locations where actual observed data were available, and comparisons were made to see if there were substantial differences between the data developed by PRISM at those locations and the station statistics. We offered a similar comment when we reviewed Volume 1. In that case, when we compared the actual data statistics from a station which was included in the analysis by PRISM, we found that the rainfall frequency data from the Atlas were lower than those analyzed using the actual observed data.

Response: Thank you for your comments. As we understand your comment, you have precipitation frequency estimates derived solely from data at a particular site and you've compared them with estimates at that site published in NOAA Atlas 14 Volume 2. If that is the case, we're not surprised you see differences.

The results published in NOAA Atlas 14 Volume 2 (as discussed in the documentation) are derived initially using a technique described in Hosking, J.R.M. and J.R. Wallis, 1997: Regional frequency analysis, an approach based on L-moments. Cambridge University Press, 224 pp. Those point estimates are then spatially interpolated using a combination of PRISM and our own CRAB technique and then subject to a variety of spatial, temporal and frequency domain consistency tests. The primary difference you see may well be the result of the Regional Frequency Analysis based on L-moments. This approach selects the appropriate probability distribution function and parameterizes it based on the assumption (which we specifically validate) that groups of observing stations are similar in their climatologies. More particularly they share higher order L-moments. In other words, the shape (not scale) of the probability distribution function for the specific group is common to the group and derived from all the data from the group members. This approach is generally significantly more accurate in estimating precipitation frequency estimates at a specific site than making the estimates based on the site's data alone. In our process we test to ensure that this is the case and in those extremely small numbers of cases where it isn't we use at-site estimates. Those specific cases are discussed in the documentation.

Therefore, the differences may be due to the technique you used to arrive at your site-specific estimates (plotting on semilog paper, conventional moments, L-moments, or some other). There may also be differences in the selected probability distribution function. We've also put a lot of effort into quality control of the historical record. You might like to check the data used in your estimates against our final QC'd time series published on our web site. After we compute point estimates, there is still spatial interpolation, and consistency checks and adjustments that we do that, again, may lead to difference from a site-specific estimate.

However, the approach we use is currently regarded as state-of-the-art and we're quite confident in the result. The statistical methods are being used in all countries that we know of who are currently updating estimates and in most academic studies. The approach is far more accurate than previous approaches, including those applied in NOAA Atlas 2 and Technical Paper 40. For the first time we've been able to estimate the uncertainty associated with the estimates and we've published an indication of the uncertainty through the upper and lower 90% confidence intervals about the estimates. It's interesting to note that in most cases the previous NOAA Atlas 2 or TP-40 estimates were within these confidence limits. Have you checked whether the at-site results you have are within these confidence limits?

If your question is specifically about any impact that PRISM may have on the results, it should be noted that little difference exists between the point estimates as calculated from the regional analysis and the estimates at that point after the spatial interpolation. There may be minor adjustments, usually within 5%, due to spatial smoothing, but those differences are within the relative noise of the data.

4.5 The reader would benefit by a listing or outline of the major steps in the frequency analysis, possibly at the beginning of Section 4. Method. Something like the following may be helpful:

- Perform at-site frequency analysis and define the L moments.
- Perform discordancy and heterogeneity tests and define homogeneous regions.
- Perform goodness-of-fit tests, select the best frequency distribution and estimate the quantiles.
- Develop regional growth factors by determining the ratio of a given quantile to the mean annual maximum flood.
- Map or grid the mean annual maximum flood with PRISM.
- Convert the mean annual maximum grid to a grid of frequency estimates.

The following specific comments are identified by page number and section.

Response: In the Introduction, Section 3.3 Approach, we provide a general outline of the major steps as you describe here, although not in a bulleted list. There were many opportunities to be unnecessarily redundant in the writing of this document, so we tried to minimize repeating text already included in other sections.

4.6 Page 4, 3.2 Terminology – The statement is made that estimates based on the annual maximum series (AMS) differ significantly from those based on the partial duration series (PDS) at the shorter average recurrence intervals (below about 20 years). However, Table 4.6.2 indicates that the PDS to AMS ration is only 1.010 for the 10-year event. The former statement should be modified to say (below about 10 years).

Response: Good point. We have deleted the word "significantly" from this reference in Section 3.2.

4.7 Page 4, 3.3 Approach – The statement is made that "Probability distributions selected for annual maximum series were not necessarily the same as those selected for the partial duration series." Later in Section 4.5 Choice of frequency distribution, it should be clearly stated that the probability distributions given in Tables 4.5.1 and 4.5.2 are for the annual maximum series.

Response: Good point. We have added text to the captions of those tables.

4.8 Page 46, Figure 4.8.1 – Why not show an example from the Ohio River Basin study?

Response: We have changed the examples in the flow chart depicting the CRAB process to reflect the Ohio River basin Project.

4.9 General comment – The CRAB procedures is very innovative and a good approach for integrating or smoothing the results across the many regions. I really need to study the procedure further in order to fully understand it.

Response: Thank you. Much work went into developing the procedure and it is somewhat complex.

4.10 The 1-yr, 24-hour precip amount is conspicuously absent. If it's not published, we will probably have to continue to rely on current estimates based TP-40. That's because some regulating agencies are naturally suspicious of lower precip amounts. We're hopeful you can include it, but if not, it would seem appropriate to mention its absence in the write-up since TP-40 did include it.

Page 3: claim is made that NOAA Atlas 14 replaces TP-40 and return periods from "1 to 100 years" is written. NOAA Atlas 14 does not include the 1-year event so how can this supercede TP-40 for this frequency?

Response: HDSC has been approached by the State of Maryland State Highway Administration (MDSHA) to calculate and include the 1-year average recurrence interval (ARI) precipitation frequency estimates in NOAA Atlas 14 Volume 2. Discussions are being held with MDSHA on funding and contractual mechanisms and the

areas to be covered. It is likely that the 1-year ARI will be included.

4.11 The comparison of NA 14 with TP-40 (p.12-13) is important to those of us who deal with regulators that may want to favor the older study. Any additional discussion and a concluding statement on the superiority of NA 14 would be of value.

Response: As stated in Section 3.1, the information in NOAA Atlas 14 Volume 2 supercedes information contained in Technical Paper No. 40. The new estimates are based on improvements in three primary areas: denser data networks with a greater period of record, the application of regional frequency analysis using L-moments for selecting and parameterizing probability distributions and new techniques for spatial interpolation and mapping. The new techniques for spatial interpolation and mapping account for topography and have allowed significant improvements in areas of complex terrain. Various other sections discuss the advantages of NOAA Atlas 14 over previous publications: Section 3, Section 4.1.1 - Technical Paper 40 data comparison, Section 4.2.1 and 4.2.2 regarding the regional approach based on L-moments, Section 7 - Comparison with Technical Paper 40.

4.12 There are occasions when we want to simulate historic events, but acquiring the rainfall data can be difficult. Would you please make available the gauge data you've compiled for public use?

Response: The complete time series – both annual maximum and partial duration – are available for download via the Precipitation Frequency Data Server at http://hdsc.nws.noaa.gov/hdsc/pfds/pfds_series.html. We are in the process of making these datafiles more user friendly by adding the date of each observed amount to the files.

4.13 Page 1: Under identified funding sources, should the Maryland State Highway Administration be identified?

Response: We've identified the resources by states in the Ohio River Basin Commission, rather than by specific state agencies.

4.14 Page 20: Quality control first paragraph: Where the erroneous data logged in any way? Do you have a record of all changes made to data?

Response: We have kept various records of changes made to the different datasets (daily, hourly, n-minute) during our quality control procedures. All quality-controlled annual maximum and partial duration series for all durations (1-hour through 60-day) are available on-line.

4.15 Page 68: One series of maps that might be nice to produce would identify in a binary (yes/no) fashion regions where the TP-40 estimates are outside of the confidence limits of the Atlas 14 estimates. I like the figure on page 69 but the map I'm suggesting would give a definitive, "Here's where the difference is big and we're confident that the difference is real."

Response: This would certainly be an interesting series of maps but it might take considerable resources to produce because of the number of durations involved and the lack of explicit estimates at those durations in the previous documents. Our purpose in providing figure 7.1 was to give a general indication of the changes rather than an explicit description.