Defining and Analyzing the Frequency and Severity of Flood Events to Improve Risk Management from a Reinsurance Standpoint

Elliott P. Morrill[1,2], Joseph F. Becker[2]

[1]Rosensteil School of Marine and Atmospheric Science at the University of Miami
[2]Guy Carpenter and Co. LLC

Correspondence emails:

Elliott.Morrill@guycarp.com
Joseph.F.Becker@guycarp.com

1.0 Abstract.

The National Flood Insurance Program (NFIP) debt has accelerated research into private flood insurance options. Offering this coverage begins with the ability to transfer the risk to the reinsurance market. Within the industry perils such as hurricanes and earthquakes have standard definitions but no such definition exists for flood. An event definition must examine the spatial and temporal aspects of the flood as well as the complexities of individual events. In this paper we were able to apply a data driven methodology to capture and aggregate flood peaks into independent events. To aggregate flood peaks into independent events we needed to define what constituted a basin as our area of aggregation. The USGS utilizes the Hydrological Unit Code (HUC) a 2 – 12 digit code that follows the Pfafsetter Coding System. The HUC code is used to identify varying levels of basin sizes ranging from Region (2 digits) to Sub-Watershed (12 digits). Choosing to analyze both the HUC8 and HUC6 a total of 7,932 HUC8 events and 8,444 HUC6 events were recorded during the 15 water years used in our study. Each event was characterized by duration, magnitude and severity. Focusing on the HUC8, events were unevenly distributed nationally while severity was relatively evenly distributed. The goal for our study was to take a method and be able to apply it to basins of varying characteristics. This framework relied on the ability to analyze the individual processes related to each individual basin.

> **Comment [EM1]:** Included the HUC definition here.

2.0 Introduction:

Throughout the world, flood events are one of the most destructive natural disasters. Floods occur for a variety of reasons, and risk factors such as total rainfall, soil types and land use can contribute to the complexity of events, in particular impacted area and event duration (Uhlemann 2010). Every year, major and minor floods contribute to economic and insured losses (Joyce 2014, FEMA). In the United States, the National Flood Insurance Program (NFIP) is the primary provider for residential flood insurance. Since its inception in 1968, the NFIP premiums have largely covered the amount paid out in losses (NFIP Act of 1968). However, the 2005 Hurricane season, including Hurricane Katrina, which was the costliest storm in the program's history costing more than 16 Billion USD, pushed the NFIP into debt (Fig.C1). The NFIP debt was exacerbated by the significant property damage experienced during Superstorm Sandy in 2012. Currently, the NFIP debt is estimated at $24 Billion as of 2014 (Joyce 2014).

This extreme debt has accelerated research into a number of different private flood insurance options. One necessary issue to address before primary flood insurance can become a more standard offering is the ability to transfer risk to the reinsurance community. A challenge specific to flood is the complexity of individual events. Unlike the perils with an unambiguous event definition, such as hurricanes and earthquakes, there is no standard definition for a flood event, which can range in length from hours to months. The problem for flooding is not specific to the United States. In fact, reinsurers have offered flood risk transfer products in Europe and Asia for a number of years. For example, (re)insurers in Spain have provided flood insurance since 1971 (Barredo et al. 2012). Typically, reinsurance contracts define a flood event using an hour's clause ranging between 168 hours in the UK to 504 hours in Germany. Using the hour's clause insurance companies are able to aggregate claims during this period of time to limit

2

43 cumulative losses from multiple events (Munich Re. 2005). Defining events this way allows for

44 providers to aggregate claims that can be associated with the same temporal event.

45   However, the hour's clause definition lacks the ability to discern between the shorter and

46 longer events. Not all events can fit into a single defined time frame. If there are multiple short

47 duration events occurring in quick succession then the claims from those events may be

48 aggregated together. The hour's clause also lacks the ability to determine spatial aspects of each

49 flood event. If events occur within the same window of time but in two different areas those

50 flood are still attributed to one event. Aggregating these events limits the ability to understand

51 the spatial extent based on impacted areas and the severity of each of the individual flood

52 occurrences.

53   While research into flood event definitions is accelerating, it is not a novel topic.

54 Research into event definitions has primarily focused on single site analysis (Bačová-Mitková &

55 Onderka 2010, Mallakpour & Villarini 2016 and Kahana et al. 2002). However, as flood events

56 are spatially complex, they often impact many locations limiting the use of single site definitions

57 for reinsurance contract definitions. When events impact larger areas, multiple locations or entire

58 basins, there is currently no method that can properly group flood peaks to the same event.

59   Public entities have complied databases of flood occurrences to assist in frequency and

60 severity analyses (NCDC). One goal of this type of analysis is to determine if floods are

61 occurring more often and with increased severity due to climate change or other anthropogenic

62 causes (Himmelsbach et al. 2015). Public databases are comprised of documentary sources and

63 trained spotter observations (NCDC, EM-Dat, and DFO). The major downside of using this type

64 of database to assist with reinsurance contracts is that they are based on subjective measures such

65 as spotter definitions. Definitions follow a series of guidelines but varying flood characteristics

66   between regions can categorize flooding differently between these two regions. Variations in

67   categorization have an impact on event durations and impacted areas. In addition to the

68   definitions themselves, trained spotters respond to citizens reports of the peril. Depending on the

69   area, what is considered abnormal flooding, in terms of standing water or bankfull discharge,

70   may be reported in one area compared to another. For example an area such as Florida

71   experiences significant precipitation year round which may contribute to minor flooding that is

72   considered normal and thus not reported. However in an area like Los Angeles that similar minor

73   flooding may be reported, which affects the frequencies of flooding in each area. Another source

74   of flood occurrence information is using a documentary source, which involves examining media

75   sources as well as government reports to comprise a set of occurrences across a state, country or

76   globe (Himmelsbach et al. 2015 and Doocy et al. 2013). These sources rely heavily on the

77   quality of the reporting, using the reports to assign severity and frequency estimates to cover an

78   expansive region.

79   Relying solely on visual reports can lead to three main areas of inconsistencies in flood

80   observations. Firstly, multiple sources can report statistics about an individual event that

81   drastically vary in the event details and determining the accuracy of conflicting points is

82   challenging without additional information. Secondly, relying on trained spotter reporting to

83   accurately defining an event is problematic. In many cases the reports cover the first instance of

84   flood and damages associated but do not report flood on subsequent days which should logically

85   define the event duration. Finally, determining the size of event requires insight from the entire

86   domain that was flooded. Relying solely on trained spotters may only confirm flooding in areas

87   that contain the most crucial infrastructure or areas of interest leading to underestimating the

88   entire flood extents.

Comment [EM2]: Re-worked sentences with respects to comment 6 from the referee

89   EM-DAT and National Climatic Data Center (NCDC) *Storm Data* databases are the two

90   that are most commonly used datasets for this type of analysis. EM-DAT uses official records of

91   areas affected, persons killed, disaster declarations issued and calls for international assistance

92   made (EM-Dat, Doocy et al. 2013). The NCDC *Storm Data* database is a compiled set of

93   observations from National Oceanic and Atmospheric Administration (NOAA) trained spotters.

94   NCDC events are categorized by county and then separated by dates (Dobour and Noel 2005,

95   Gaffin and Hotz 2000). EM-DAT catalogues events by year with summary statistics detailing

96   frequency and overall event impacts (i.e. deaths and losses) from that year. Such summary

97   statistics include injured, affected, total deaths and total damage. Both methods contain a number

98   of different biases preventing use in reinsurance contracts including population biases, frequency

99   biases and reporting biases. Due to the incomplete and often inconsistent reporting,

100   implementing this method to formulate an event definition for reinsurance contracts presents a

101   challenge. Despite their limitations, these datasets are useful first checks when developing a

102   more robust method to define flood events as historical events can be compared to.

103   Many authors have shifted toward a data driven approach using the peaks over threshold

104   analysis to examine changes in flood event frequency (Mallakpour and Villarini 2016, Bačová-

105   Mitková & Onderka 2010), as well seasonality (Black and Werritty 1997). A data driven

106   approach allows for the definition of an event to encompass a variety of basin characteristics.

107   Authors choose a somewhat arbitrary threshold where if a peak observation exceeds the

108   threshold, it is considered to be a peak over threshold (POT). A subsequent step for this method

109   was to determine a metric for identifying independent peaks. Varying windows of time were

110   used to identify the independence between the individual POT. Mallakpour and Villarini (2016)

111   used an arbitrary window of 15 days, where any peak that occurs within this period is aggregated

Comment [EM3]: Year added: Comment 7

112   to a single event. Black and Werrity (1997) determined their window by calculating the "time to

113   rise" and identifying when the discharge dropped below 2/3[rd's] of the previous peak. Authors

114   using these windows then looked at all individual peaks occurring within these windows to

115   attribute them to the same event.

116   Site specific event identification is the base in developing a consistent method of event

117   identification. However, our method will address the window of independence through an

118   observational approach. Event independence should not be based on a standard window

119   (Mallakpour and Villarini 2016). It must be based on how each site reacts to the flood waves.

120   Implementing a concept similar to time to rise and a drop in discharge (Black and Werritty 1997)

121   was the first of many steps taken toward resolving this. The window must cover the time before

122   and after a peak, as previous peaks have an influence on succeeding peaks. Incorporating this

123   into our definition will reflect the individuality of each site and the flexibility of our definition to

124   cover a wider range of sites.

125   The primary goal of this research is to expand our definition to an entire basin or

126   catchment area. These regionally impacting events are titled basin or "trans-basin" events (Nied

127   et al. 2014, Uhlemann et al. 2010). Both papers used the POT method as well. Starting with a

128   single site, individual events were identified (Uhlemann et al. 2010) and then all mutually

129   dependent events were identified from a moving temporal window. The window defined from

130   previous literature provides a solid structure but categorizes catchments and basins into an all-

131   encompassing time frame. A more basin specific time frame is measurable and would not

132   underestimate the smaller basins or overestimate the larger basins.

133   This paper seeks to define events through a data driven approach aimed at accounting for

134   the individuality of flood waves and the basins they impact. Our main goal is to develop a

135 consistent definition in order to examine how frequency and severity vary regionally. Looking at

136 frequency regionally provided us with a clearer picture of the specific areas that were more at

137 risk for flooding. Severity allowed us to look at how areas with similar frequencies were

138 experiencing events in terms of impacted areas and overall magnitude. Severity will factor into

139 future implementation of risk mitigating factors that can look at two areas and determine the

140 steps needed to protect a certain area. It also allowed us to determine if our method is

141 representing more local or extreme flooding across the various basins.

142 Methods implementing the hour's clause or standard event windows lack the ability to

143 interpret how each individual flood wave progresses. Understanding the individuality of the

144 flood is the basis for how our method will tackle a standard event definition. This paper will be

145 structured as follows: Section 2 will cover the data availability as well as the data selection

146 process along with which tools were used to analyze the data. The concepts that feed into our

147 method as well as our method itself will be discussed in Section 3. Section 4 will provide the

148 results of the analysis from our methodology with comparisons to methodologies exhibited in

149 previous research. Section 5 will provide the discussion and concluding remarks regarding our

150 results within this study.

151 **2.0 Site Selection:**

152 This research focuses on expanding the definition of a flood event from an individual site

153 to river basin. As this research focuses on the United States, USGS daily flow gauges stations

154 were used to identify individual sites and USGS Hydrological Unit Codes (HUC) were used to

155 define river basins. River basins can be defined in a number of ways and determining the

156 appropriate size can be a non-trivial task. For use in reinsurance contracts, river basin should be

157 defined in such a way that flooding events within a portion of the basin show a correlation to

158   events in other portions of the same basin. Basins will also need to be defined in such a way that

159   we can see how flood waves impact the entire basin and not individual sections of that basin. The

160   USGS HUC codes follow the concept of the Pfafstetter Coding System meaning that each unit

161   code is delineated in a hierarchical fashion ranging from larger to smaller. Drainage areas are

162   defined on a continental scale and then divided and subdivided into 6 levels. Each level is

163   associated with number of digits corresponding to size. The USGS utilizes a 2 digit system that

164   defines each basin level by the number of digits each code contains. HUC Codes range from 2 –

165   12 digits, largest to smallest (USGS). For example, each basin defined as a HUC8 (Subbasin) has

166   a unique 8 digit code. Based on this system as well as past research, the 8 & 6 digit HUC's were

167   chosen as the basin levels that we would analyze. A majority of the papers that we referenced in

168   this study have dealt with European or Asian basin definitions and were focused on one or two

169   basins within a finite area. With our broad scope of study, we needed to look at basins across a

170   variety of characteristics so a common basin code was needed for comparisons of frequency.

171   Other research of flood frequency did not yield any references to the HUC basin codes so as

172   authors we developed our own criteria that we felt best represented the size of the basins most

173   applicable for our methodology. Our decision to use these two size HUC's relied on looking for

174   a basin size that allowed us to observe how the events would aggregate to a basin level event

175   rather than being identified as two separate events. We wanted our dataset to contain the largest

176   percentage of HUC's possible after our site selection criteria to get a better nationwide picture of

177   how our method observed basin wide flood events. With the HUC8 we were able to get

178   approximately 20% of coverage across the United States with a basin contained all 20 HUC2's

179   (Fig.1). With any HUC size below the HUC8 such as the HUC10 we were left with a much

180   lower coverage percentage, roughly less than 10% for the HUC10, which would not accurately

8

181 represent the methodology across the country. When we look at the upper end of our HUC size

182 for the HUC6, we when look at how frequency compares with site count above the HUC6 we

183 saw that frequencies were heavily affected by site count. From these two factors we felt that the

184 HUC8 and HUC6 were the most applicable basin sizes. Daily mean discharge as well as Annual

185 peak streamflow was used for all sites, which provided data for those parameters.

186 From all available HUC's, sites and basins were selected based on a number of selection

187 criteria. The first criteria removed sites with less than 5 years of daily discharge data. The second

188 criteria required sites to occur along natural rivers and streams; gauges impacted by reservoirs

189 and other impediments to natural flow were excluded. Following site removal, HUC's with less

190 than 5 sites were excluded. Finally, HUC's were required to have at least 3 sites that overlapped

191 with 70% of the data during each individual year that was examined. Due to the nature of our

192 method seeking to aggregate peaks from multiple sites, the sites needed to overlap or else that

193 method would be looking primarily at individual site events instead of the basin events. Of the

194 2,300 HUC8's and 387 HUC6's available, 462 HUC8's and 276 HUC6's were used (Fig.1) with

195 a total of 3,121 and 4,919 gauge stations within the HUC8 and HUC6 respectively. Both HUC

196 sizes were analyzed for initial frequencies and the most applicable HUC was chosen for

197 subsequent analyses.

198 **3.0 Methodology:**

199 Daily discharge data from 8,084 river gauge stations was obtained from the USGS

200 (*http://nwis.waterdata.usgs.gov/nwis/dv/?referred_module=sw*). A study period of 15 water

201 years between 2000 and 2015 was selected for this analysis. Initial attempts to expand the period

202 of analysis severely reduced the number of basins that met the criteria for analysis. The peak

203 over threshold method outlined in Uhlemann et al. (2010) was conducted on all basins that fit the

204   criteria for analysis. The peak over threshold method consists of identifying individual

205   observations over a specified threshold within a particular time window. The procedure was split

206   into 4 major steps: (1) identifying peaks occurring at each site within each basin and the

207   subsequent peaks over threshold; (2) applying a window of independence at each site to

208   determine independent site specific events; (3) compiling all independent site specific events and

209   applying a secondary window of independence to determine independent basin specific events;

210   (4) applying multiple characteristics to determine a severity score to compare differing events

211   from one another.

212   The first step involved selecting a minimum threshold. The median of annual maximums

213   was chosen as the threshold in which a flood peak must exceed. The median of annual

214   maximums was chosen because it corresponds to the 2-year quantile, or Q2. Uhlemann et al.

215   (2010) states that the "Q2 is a rough estimation for bankfull discharge on naturally occurring

216   streams." For sites with at least 5 years of annual peak streamflow data, their Q2 was calculated

217   by taking the median across the entire time series. As peak discharges are determined by

218   instantaneous measurements, small catchments can exhibit extreme values, which are rarely

219   observed in the daily record. The extreme values may lead to a minimum threshold that may not

220   be a representative measurement of flooding for that catchment area. The discharge at each of the

221   peaks recorded, were then compared to their respective sites Q2 value to determine all of the

222   peaks over threshold.

223   The next step in identifying site specific events is to determine a time criteria that defines

224   independent site events. Two metrics were calculated for all peaks over threshold to determine

225   the duration of each event: base to peak (BtoP) and peak to base (PtoB). Base to peak is the time

226   it takes for the discharge to reach the peak after it has crossed the minimum threshold. Peak to

10

227   base is the amount of time it takes for the discharge to return to the minimum threshold

228   following a peak (Fig.2a). In the case where there are multiple peaks before the discharge returns

229   to base, the peak was selected as the observation that experienced the maximum discharge. Each

230   peak over threshold has a unique BtoP and PtoB that could have a significant range. To

231   standardize the windows of independence for each site the median of both metrics was calculated

232   and then the peaks start and end times were recalculated. Our window of time was aimed at

233   eliminating the extreme events on either end of the temporal distribution to determine a window

234   that reflected the time it would take for a flood wave progress through a site.

235   After the windows were recalculated, combining peaks with overlapping or consecutive

236   windows into a single site specific peak consolidated peaks. All peaks over thresholds with

237   windows that did not overlap were treated as independent events. Each event was characterized

238   by, site number, start time, peak time, end time and peak discharge. For the peaks, which

239   overlapped, the start time was defined as the earliest start day and end time was the latest end

240   date. The peak discharge from each event was then scaled by the Q2 at each site. Scaling each

241   peak discharge reduced the impact of catchment size when comparing magnitude of discharge

242   and made the different sites comparable.

243   A similar methodology of consolidating overlapping observations was applied to define

244   basin specific events from the site specific events (Fig.2b). The basin specific events used the

245   start and end time of each site specific events that occurred within the basin. If the windows of

246   time between the start and end of the site specific events overlapped or were consecutive (i.e.

247   occurred within 1 day of another peak), then these events comprised one basin specific event.

248   The start of the event was the earliest start time recorded at any site and the end of the event was

11

249   the final end time recorded. Each event was defined by start time, end time, peak time, and peak

250   discharge for all events from the desired HUC's.

251   The final step involved determining a severity score for each basin event. Defining

252   severity allowed us to compare areas of like frequency. From these we were able to see the

253   certain areas that are more vulnerable during flooding. Severity scores in future analyses will

254   also factor into pricing of reinsurance contracts. Severity of each event was designed to include

255   elements of the spatial extent as well as the magnitude of the flooding experienced in the basin

256   by the affected sites during each event. The severity score represents a number between 0 and

257   infinity where the high value indicates a more severe event. The affected sites were defined as

258   the number of sites within the desired HUC, which recorded a peak over threshold during the

259   event. Total discharge was the sum of the discharges, scaled by their corresponding minimum

260   threshold, observed at all the affected sites. Severity was calculated by taking the sum of all

261   scaled discharges and dividing by the total number of sites within the basin, *EQ.A1*. If a site was

262   impacted more than once during a basin event, the maximum-scaled discharge was selected to

263   calculate the severity score. Scores less than one are expected when looking at the minimum

264   threshold as it represents small scale and localized flooding, in terms of discharge and the

265   percentage of sites it may impact within the individual HUC.

266   From the analyses, we compared the HUC6 and the HUC8 frequencies, event duration

267   and severity distributions. With our goal of a basin wide definition, it is imperative to compare

268   these two basin sizes and determine the most appropriate basin level for our methodology. To

269   compare, we looked at the differences between the statistics listed previously as well as the

270   distribution of the percentage of impacted sites by event for each HUC. The distribution of the

12

271 <mark>percentag of impacted sites was used to determine whether events in each basin level are being</mark>

272 <mark>aggregated to a basin event or if they are being segmented due to the size of the basin.</mark>

> **Comment [EM5]:** Included a paragraph discussing the selection process in the Methodology section in reference to comment 2

273   Two comparisons were made to the NCDC *Storm Data*. The first method looks at all

274 reports of flooding and aggregates them by county. The second method used a standard 13-day

275 independence window, 3 days pre-peak and 10 days post-peak (Uhlemann et al. 2010). A

276 standard window was used because the NCDC observations are unable to provide a site specific

277 window of independence.

278 **4.0 Results:**

279   A total of 7,932 and 8,444 events were calculated for basins defined by the HUC8 and

280 HUC6 respectively. Table B1 provides the frequency summary statistics for both the HUC8 and

281 HUC6 basins. Comparing the frequency distribution of events between the two selected basins

282 sizes suggests that frequencies within basins defined by the HUC6 are higher than frequencies

283 defined by the HUC8 (Fig.3 & Fig.4). We can see that from Figure 3, the frequencies in each

284 HUC8 are typically lower than the frequencies found each HUC6. This is highlighted in Figure

285 4, where we focus on 6 HUC8's that make up 1 HUC6 (Outlined in Blue). From here the

286 individual basins in the HUC8 indicate a lower basin level frequency than at the HUC6. This

287 comparison is important because the aim of this paper is to define events at a basin level by

288 aggregating individual events into basin wide events. To explore this concept more we wanted to

289 look at the impacted sites during the events compared to the total number of sites within the

290 basin to get a sense of how many events are being determined as local when they should be

291 aggregated. While there will be a number of small local floods that this methodology captures,

292 we looked at this to provide us with an indication of whether the HUC8 is too small of a basin to

293 use or the HUC6 is too large.

294   We looked at the distribution of the percentage of impacted sites by event for each HUC

295   (Fig.5). We took each event within our catalog and identified how many sites were impacted.

296   The percent impacted was calculated by taking the number of sites impacted and dividing by the

297   total number of sites within the basin. For the events within the HUC8 on average 36% of the

298   sites were impacted compared to 21% for the HUC6. When you look at the CDF of the events of

299   HUC6 and HUC8 (Fig.5), we can clearly see that the HUC6 events impact a fewer percentage of

300   sites. While HUC6 does have more sites, due to our methods intended aggregation of events we

301   would expect a similar percentage of sites impacted between the two. However, because the

302   HUC6 is showing a lower percentage of sites impacted during the events in their catalog this is

303   an indication that the HUC6 does not aggregate individual events as well as the HUC8. 80% of

304   the events within the HUC6 have % impacted <40% compared to the the HUC8 where

305   approximately 50% of their events are impacting 50% of the sites. Due to the size of the HUC6,

306   the basin is being segmented during our method and is not capturing events that should be attributed

307   to the same event. The segmentation of the events within the basin will lead to an overstating of the

308   frequencies. Overall, the HUC8 is showing a higher percentage of events in the higher percentages of

309   sites impacted meaning that our method is aggregating more individual events into basin events at

310   this basin level when compared to the HUC6. From both the CDF and the average we have

311   concluded that the HUC8 is a more applicable basin size due to its ability to aggregate the events

312   within the basin rather than segmenting them.

313   Nationally, the median frequency of events HUC8 basins was 1.00 events per year while

314   the mean was 1.14 events per year (Fig.6). This frequency varied regionally with some areas

315   experiencing higher frequencies (Fig.1 Left Panel). Notable population centers that experience

316   elevated frequencies include the Upper Midwest (south of Lake Michigan), Southern California

317   and Southern Florida. For the HUC6 basins, the median frequency of events was 1.87 events per

14

318   year with a mean of 2.03 events per year (Fig.7). Similarly to the HUC8 basins, the frequencies

319   varied regionally with some areas of elevated frequencies (Fig.1 Right Panel).

320   To investigate how event duration varies nationally, we calculate the mean event duration

321   for each basin. Nationally, the mean event duration ranged from two to 79 days for the basins

322   defined by the HUC8 and two to 73 days for the basins defined by the HUC6. The mean event

323   duration for 95% of HUC8 and HUC6 basins is less than 14 and 17 days respectively (Fig.8 and

324   Fig.9). The minimum event duration was two days and was observed at 336 HUC8's and 227

325   HUC6's. The maximum event duration for HUC8's was 232 days and occurred in the 10160003

326   basin. For HUC6 basins that maximum event duration was 237 days occurring in the 101600

327   basin. When we look at the shape of both curves, we can see that there is a higher percentage of

328   HUC8's that have shorter mean and maximum durations, as the curves approach the lower event

329   durations more rapidly leading to a steep curve when compared to the curves for the HUC6.

330   However, when we look at the minimum duration, a larger percentage of the HUC6's have a

331   minimum duration of 2 days when compared to the HUC8 which is an indication that there is a

332   larger number of events that are impacting only one site. While both the HUC6 and the HUC8

333   tapper of towards the higher event durations, there is a lower percentage of the HUC8's that have

334   event durations greater than 20 days. With those two factors we can see that durations within

335   each HUC6 have a wider range than those compared to the HUC8.

336   Figure 10 represents two sites that reflect longer recession periods following their peaks.

337   With a data driven approach identifying the generation and recession of the events, certain

338   extreme events may show increased event durations. The extreme durations are a reflection of

339   the minimum threshold as well as the hydrological processes at hand. Looking at the two sites,

340   the left is located in South Dakota and the right is located in Florida; both of the extreme events

15

341   that are observed have certain factors that impacted their recessions. The site in South Dakota

342   experienced an event that was impacted by the melting of an ice jam represented by the quick

343   generation. Following the melt there was a significant rain event as well as a release of water

344   from a dam further upstream. The site on the right is located on a natural tourist spring. These

345   springs contain a significant amount of ground water. Following an intense rain event the

346   buildup of water caused the increased recession. When we define an events' duration as the first

347   occurrence of discharge above the Q2 to the final occurrence of discharge below the Q2, if our

348   site is impacted by a natural occurrence, events will reflect longer than expected durations. These

349   durations are longer than we would expect and further analysis will be conducted to examine

350   changes to the minimum threshold to examine the influence of these natural processes. While a

351   majority of the durations reflect reasonable time frames for flooding events that exceed the Q2 it

352   is important to note that the method might not be appropriate for all streams.

353   When looking at the distribution of severity scores there is a slight skew towards the

354   extreme events. Severity scores ranged from the least severe, 0.032 to the most severe, 26.9

355   (Fig.11) with a median severity score of 0.32 and a mean of 0.57. While the range in severity

356   scores is quite large, a majority of the events received a score less than 1. Regionally the severity

357   scores are generally distributed evenly throughout the country (Fig.12). There appear to be

358   pockets of higher severities but across the country there does not appear to be a pattern within

359   the regional distribution. While it is evenly distributed regionally, within the regions we can see

360   the wide range in severity that was observed in the distribution of frequency.

361   Finally, comparisons were made to other methodologies applied to the same dataset as

362   well as other publically accessible datasets. The first comparison examined a method used by

363   FEMA to estimate floods using NCDC Storm Events Database (Fig.13). The distribution of

16

364   events was broken down into total event frequency by county ranging from one to 4,114. While

365   the trained spotters follow guidelines in identifying events, the method lacks a way to group

366   events. The inability to group events that would otherwise be considered a single event, leads to

367   an overestimation of events. This overestimation is evident when it is noted that the maximum

368   frequency of events for a specific county was 4,114.

369   The final comparison was made to the NCDC applying a 13-day standard window. While

370   the NCDC map provides a more complete national coverage two patterns occur (Fig14). Within

371   the 5-boxed areas, either the NCDC frequency is far greater or the daily discharge frequency was

372   far greater. For example, in Florida, we see frequency range from 6 to 25 events for NCDC

373   observations but events observed through daily discharge range from 26 to 45. The opposite

374   occurs in Missouri with NCDC estimates ranging from 16 to 85 events with events observed

375   through daily discharge ranging from 6 to 15.

376   From these estimates there is no obvious reason for the discrepancies in frequencies but

377   we can speculate. For example Florida experiences significantly fewer events using NCDC data

378   than the daily discharge data. A possible explanation could be how trained spotters define events.

379   An area in Florida may experience a peak over the threshold triggering our event definition, yet

380   that peak may not be recorded as an NCDC observation based on the spotters perspective.

381   Another reason could be due to the fact that these trained spotters respond to citizen's reports

382   and, due to the frequency of flooding in an area like Florida, the citizen may not call and the

383   peak may not be recorded.

384   However a similar thought process can be applied to our threshold selection. As stated

385   the minimum threshold was selected as a representation of bankfull discharge. While this

386   assumption was the basis for our method, in certain areas it is conceivable that the threshold may

17

387 be lower than bankfull discharge which could possibly lead to an over estimation of flooding

388 events in certain areas. There is no certain explanation for the discrepancies in the results. With

389 no certain explanation for the results from this comparison, the assumptions that define the

390 compared methodologies will be explored in future analyses.

391 **5.0 Discussion:**

392   This study was able to provide a data driven approach in attempts to solve the issues of

393 inconsistent event definitions within the (re)insurance industry. We derived a methodology based

394 on a peak over threshold analysis that was able to capture and aggregate multiple occurrences of

395 flooding at various locations. Using physical assumptions, our minimum threshold and window

396 of independence were able to capture each individual sites reaction to passing flood waves. An

397 approach identifying windows based on the impacted site allows for each site to represent their

398 individual characteristics of flooding rather than applying standard metrics throughout. Each

399 event was defined through their duration, impacted area and magnitude. The development of a

400 severity index examines overall impacted areas as well as individual flood magnitudes.

401   Analyses were conducted on both HUC8 and HUC6 to determine which size of

402 Hydrological Unit Code was more applicable for further analysis. 7,932 HUC8 and 8,444 HUC6

403 events were identified during our study. Understanding the applicability of different basin sizes

404 is important because it aids in our main goal of applying a consistent definition to reinsurance

405 contracts. From our definition our goal was to understand the frequency that represents an entire

406 basin or area. We also hope to use the definition to define a parametric trigger or an alternative

407 form of defining the event. All of this is possible when we know what basin size is the most

408 applicable. The HUC8 was chosen as a more applicable basin size as it was a better

409 representation of site interaction during flooding events.

> **Comment [EM6]:** Split the discussion and the Conclusions into 2 sections – Comment 5

410 Nationally, there are areas with large discrepancies between the HUC6 and HUC 8

411 frequencies. One explaination of this discrepency is represented by HUC6: 071200 (Fig. 4). The

412 area of this HUC6 is 28,309.78km^2 and contains 6 HUC8s. The annual frequency of events of

413 the HUC8 ranges between 1 and 2.33, while the HUC6 produces 5 events per year. Although it is

414 expected that the larger basin will have a slightly higher frequency due to some events occuring

415 in one part of the basin and not impacting the other, a more than doubling of events per year

416 indicates that a large number of events do not interact with other sites in the basin. This lack of

417 interaction is inconsistent with the goal of this research to identify basinwide event frequencies.

418 The inconsistencies and lack of interaction are represented by the relationship between site count

419 on frequency.

420 Based on our analysis of two levels of HUC's, determining which basin size was the most

421 appropriate was a crucial portion of our analysis. To determine which was more appropriate

422 distributions of the percentage sites impacted were analyzed in order to see how sites were

423 interacting during events. When examining the cdf of percentage of sites impacted (fig. 5), we

424 can see that the HUC 8 is the more applicable basin level to use for our analysis. HUC8's showed

425 a higher percentage of events had a higher percentage of sites impacted that were impacted during

426 each event when compared to the HUC6.With this comparison we can see that there are more

427 individual events that are being aggregated to basin events rather than those events being segmented

428 into multiple events. With this aggregation we are seeing a more complete picture of frequencies at

429 the HUC8 than the HUC6.

Comment [EM7]: Discussion about which HUC size is best for our analysis. Comment 3.

430 We found that HUC8 frequencies are relatively normally distributed but are unevenly

431 distributed regionally. For all HUC8's a median of 15 events (1 event per year) and mean of

432 17.21 events (1.14 events per year) were recorded. In a number of areas there were pockets of

433 elevated frequencies. Durations for all events ranged from 2 – 232 days with a mean duration of

434   6.34 days. The wide range of event durations prompts further investigation into events with

435   durations in the positive tail of the distribution. For example, we considered two HUC8's, one in

436   South Dakota (10160003) and another in Florida (03100207), that are impacted by natural events

437   leading to longer durations. Some sites within these two basins were affected by ice jams as well

438   as natural springs, which have contributed to significant recessions of their events. While these

439   events are natural, the resulting event durations should prompt examination into the selection of

440   thresholds for the sites, as an assumption of bankfull discharge might be slightly lower than a

441   threshold that produces flooding. Investigation into the bankfull discharge assumption is

442   necessary when determining appropriate level of flooding to conduct our methodology. Analyses

443   will be conducted testing our methodology using varying levels of flooding, comparing our

444   estimates using the Q2 to Q5 and Q10. In addition to testing the various levels of flooding based

445   on return period, we will examine the impact of the estimated bankfull discharge to the observed.

446   Severity scores calculated for all events in the dataset showed a slight skew toward the

447   more extreme events. The smaller and local events are represented by the median of 0.32 and

448   mean of 0.57, as we can expect events slightly above the threshold to not necessarily affect all

449   the sites in the basin, producing a score less than 1. Regionally severity is relatively evenly

450   distributed nationally.

451   **6.0 Conclusion:**

452   With a data driven approach to our methodology, a focus on the individual site

453   parameters shifts the focus from generalities about events to site specific understanding leading

454   to an applicable method regionally. A fundamental aspect of this research is to understand spatial

455   extent of flooding and we were able to expand from single gauge stations to entire basins. The

456   data driven approach allowed us to apply the methodology to a number of basins with varying

Comment [EM8]: Added additional commentary on the selection of the Q2 and what steps will be taken to address the possible change in threshold based on Characteristics of the basin or a standard return period approach – Comment 1

20

457   characteristics. The final advantage to our method is that when looking at flood severity we do

458   not look at exclusively magnitude but the addition of spatial extent adds an element to

459   differences in severity regionally.

460   While there are a number of advantages that come from this method, relying on public

461   data have revealed drawbacks in its application. Being a data driven method limits our ability to

462   estimate frequencies in areas that do not have data. Across all USGS gauges there is no

463   uniformity in data availability for number of years or number of stations within a basin. Through

464   our site selection process we were only able to use 20% of all available HUC8's, which limits

465   national coverage in our estimates.

466   The minimum threshold for flooding is based on the assumption that it is a representation

467   of bankfull discharge; in certain areas this may not be accurate. Riverbanks are not uniform so

468   how bankfull discharge is recognized at each site is dependent on that location, which may lead

469   to underestimation or overestimation of flood stage at that site. The final drawback we observed

470   was that when taking the median of the BtoP and PtoB slight variations in the event window

471   occurred on the more extreme events. Instead of median other statistics will be tested to

472   determine the most applicable way to represent the basin flood generation and recession.

473   For further research a comparative analysis will be conducted altering the threshold to

474   examine how that might affect frequency as well as severity. Increasing the time frame will also

475   provide insight as to whether or not this 15-year period is representative of the entire time frame

476   of data or if we see a significant increase in events during certain subsections. Seasonality tests

477   will be run to observe areas more frequent and more severe times of year which may also

478   provide insight for risk managers. The final test that will need to be conducted is a sensitivity

479   analysis on the threshold selected to prove which threshold is the most reasonable for an analysis

480   such as this.

481

522 **Appendices:**
523
524 **Appendix A.**
525

$$\text{Severity} = \frac{\sum Q(i)_{\text{Scaled}}}{\text{\# of Sites (HUC)}}$$

526
527 *EQ.A1. Severity Score*
528 **Appendix B.**

HUC	Total HUCS	Selected HUCS	Minimum Freq.	1st Quantile	Median Freq.	Mean Freq.	3rd Quantile	Maximum Freq.
08	2300	462	0	10	15	17.17	21	63
06	387	276	0	19	27	30.59	38	145

529 *Table.B1. HUC8 and HUC6 Frequency Summary Statistics*
530 **Appendix C.**
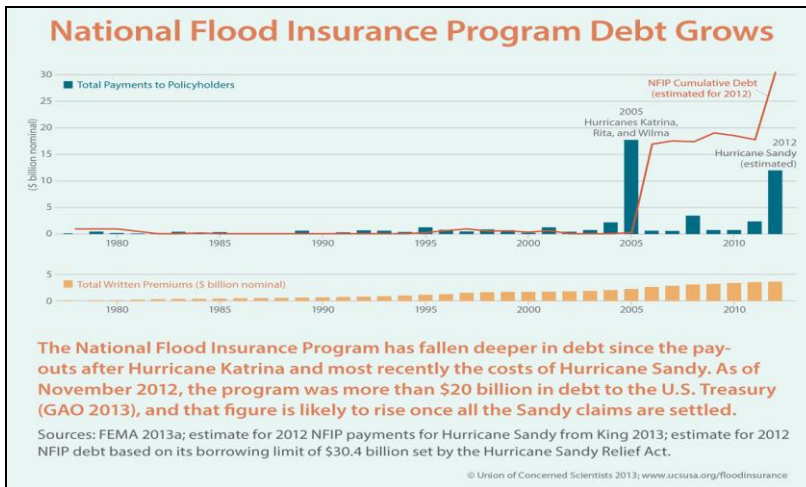
531
532 *Fig.C1 NFIP Cumulative Debt, Total Payments and Total Premiums, 1978-2012*
533

534

535

536

537

538

539

540

541

542 **Figures:**
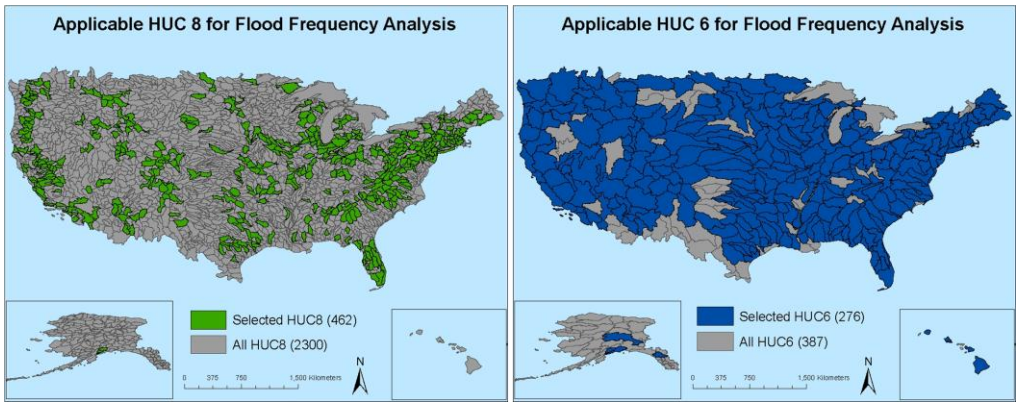
Fig.1. A map of the selected HUC8 and HUC6

543
544
545
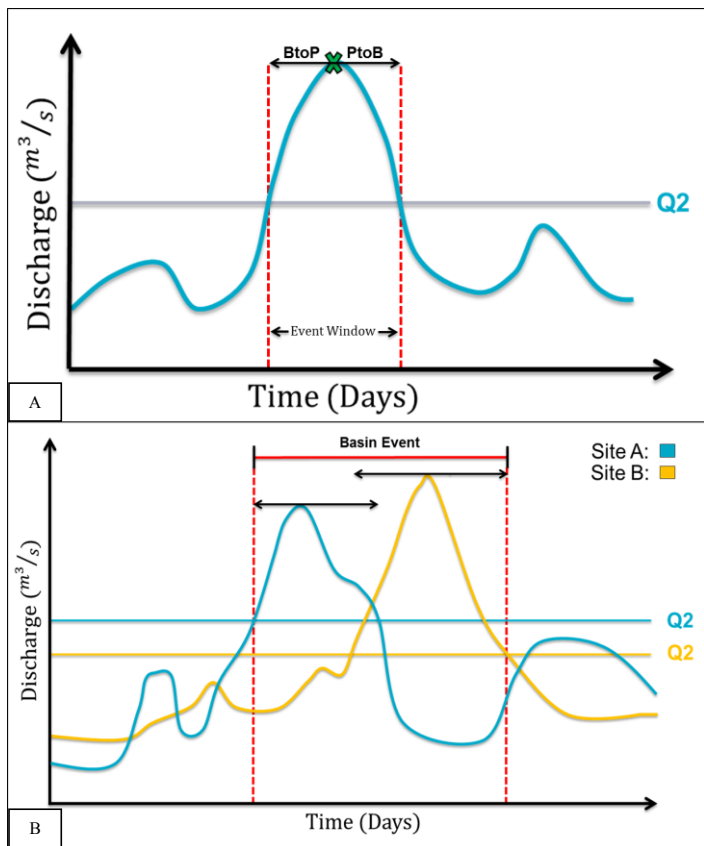
546
547 ***Fig.2a. Site Event Identification (Top Panel) and Fig.2b. Basin Event Identification (Bottom Panel)***
548

25
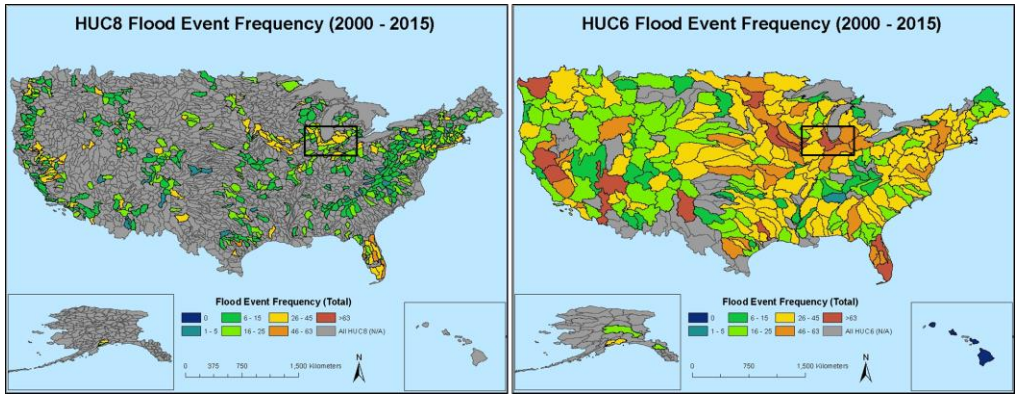
Fig3. HUC 8 and HUC6 Frequency Comparison, National
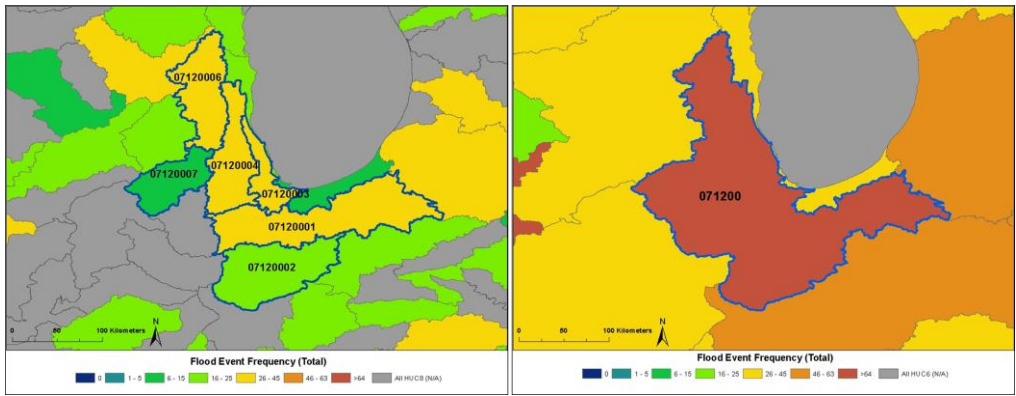
Fig.4. HUC 8 and HUC 6 Frequency Comparison, Upper Midwest
Blue Outline (HUC6: 071200, HUC8: 07120001---07120007)

Distribution of % of Sites Effected
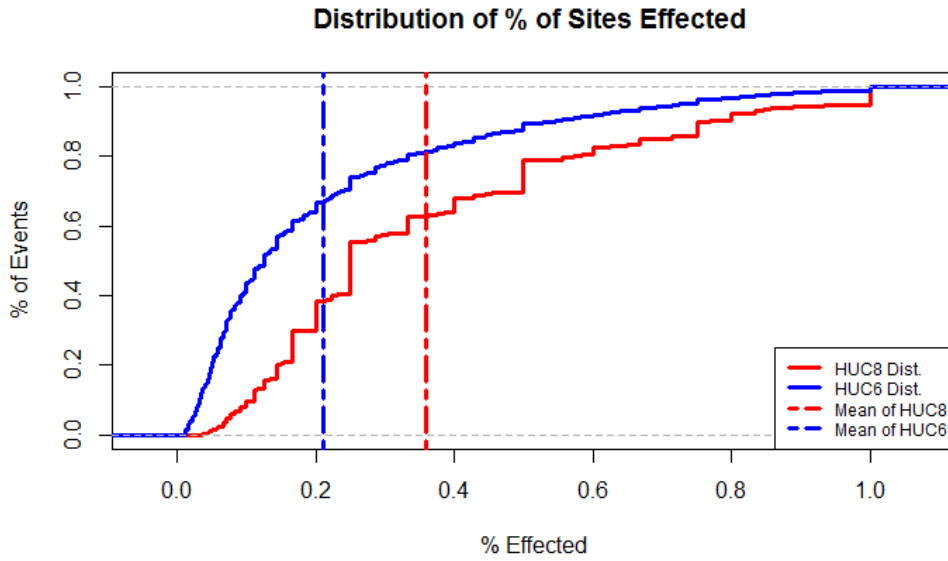
Fig.5. CDF of the percentage of sites impacted during each event within our catalog. Mean % of the entire distribution is noted and split by HUC.

HUC8 Event Frequency Histogram
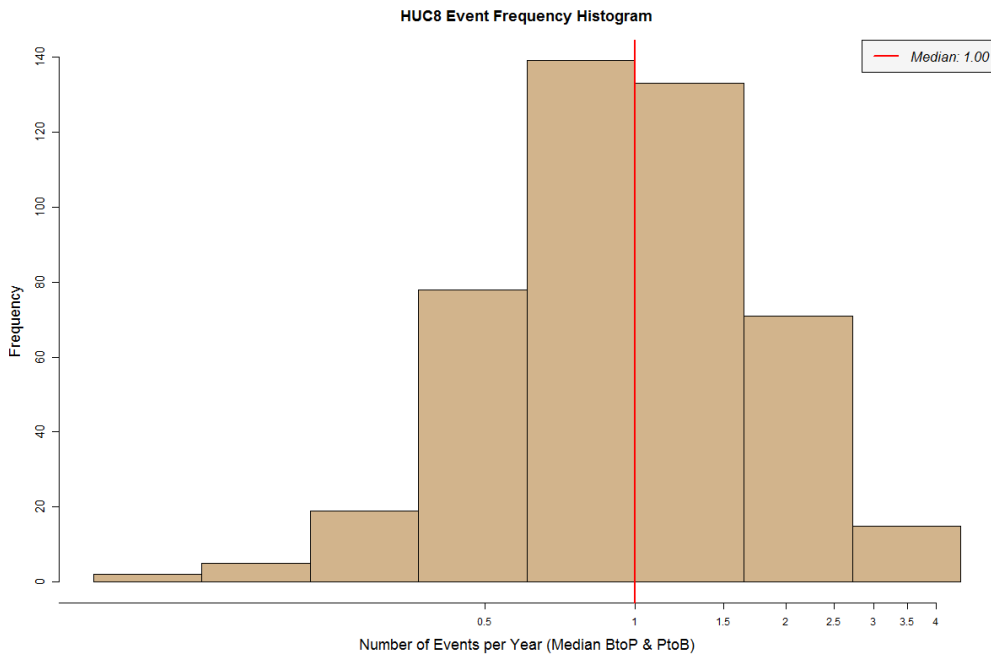
Fig6. HUC8 Frequency Distribution

563
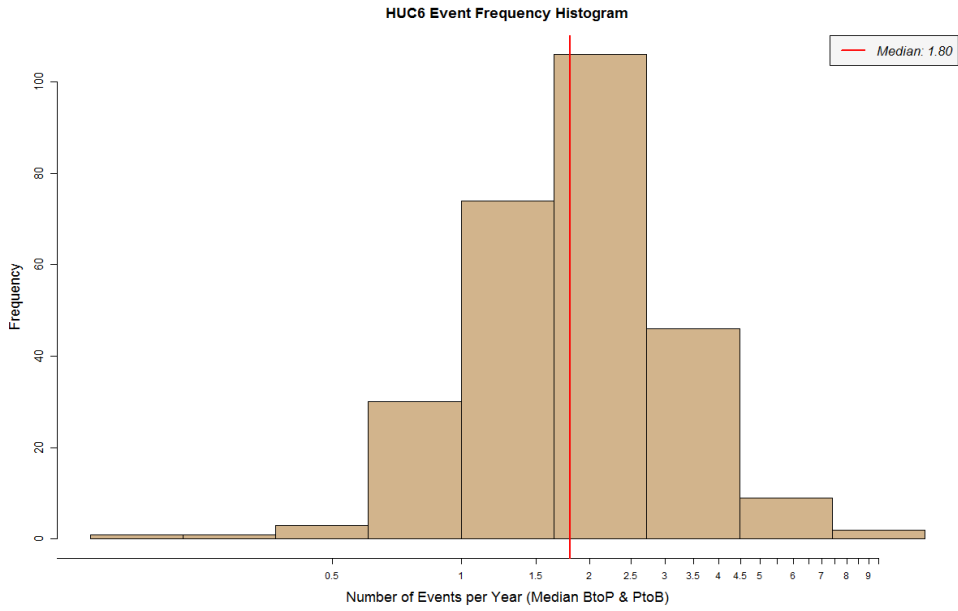
Fig.7. HUC6 Frequency Distribution

564
565
566
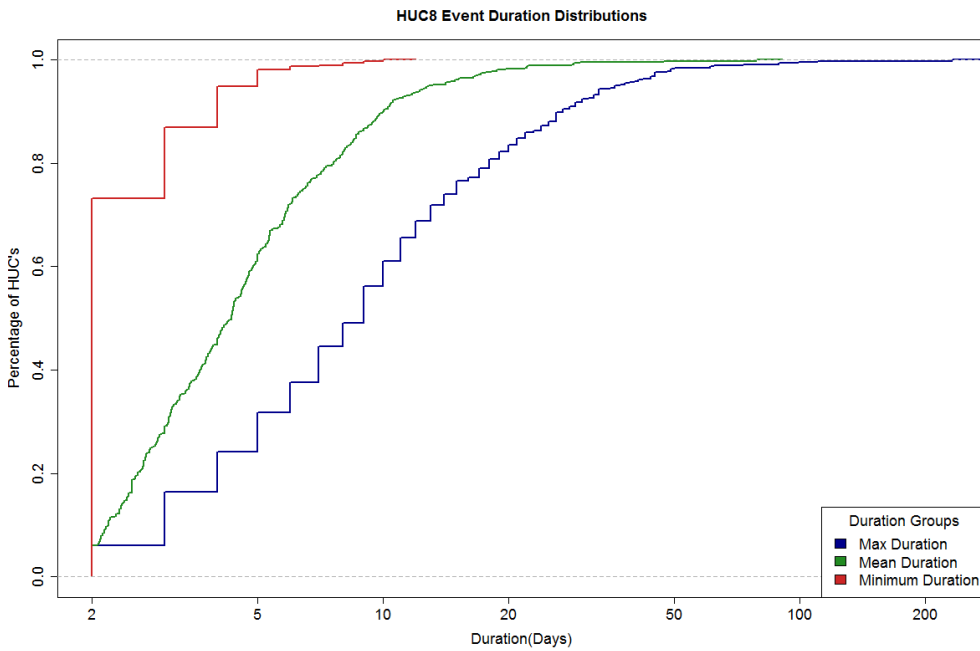
Fig.8. HUC8 Event Duration CDF

567
568

28

HUC6 Event Duration Distributions
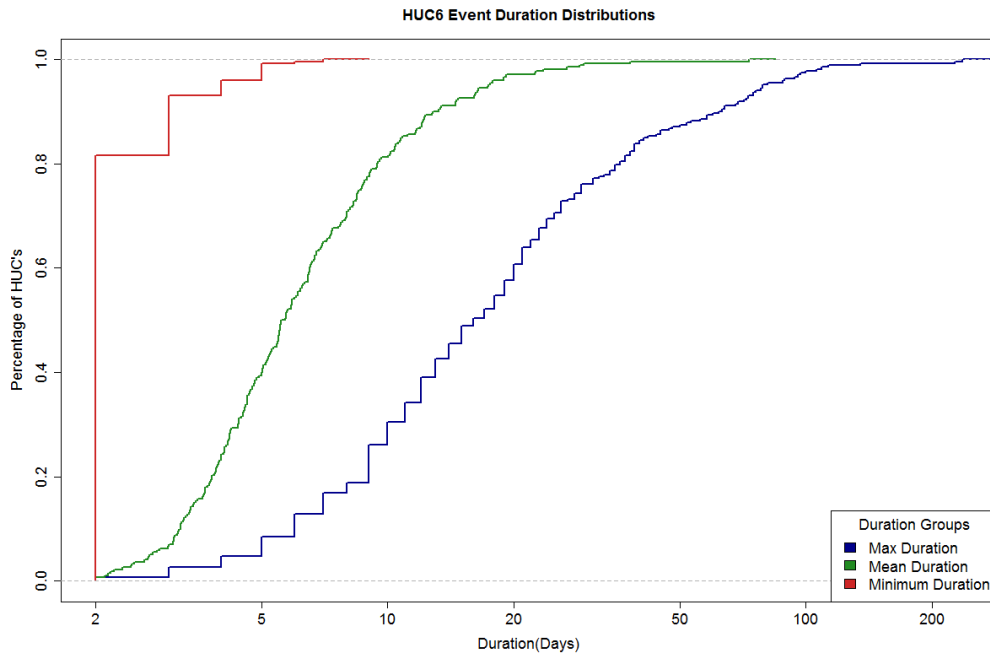
569
570   ***Fig.9. HUC6 Event Duration CDF***
571
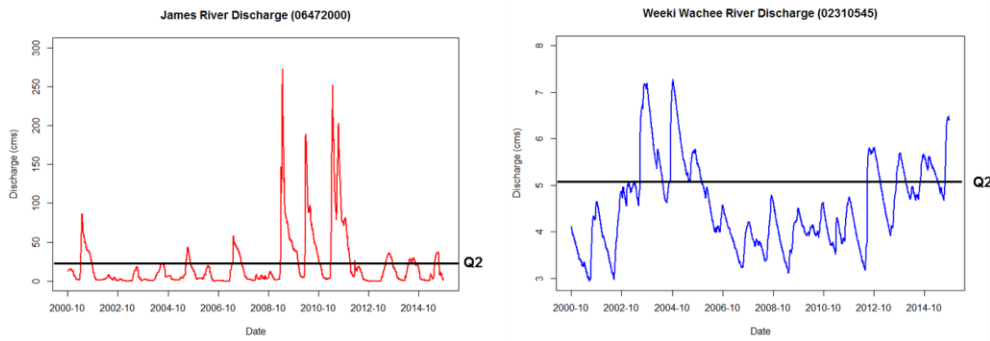
572
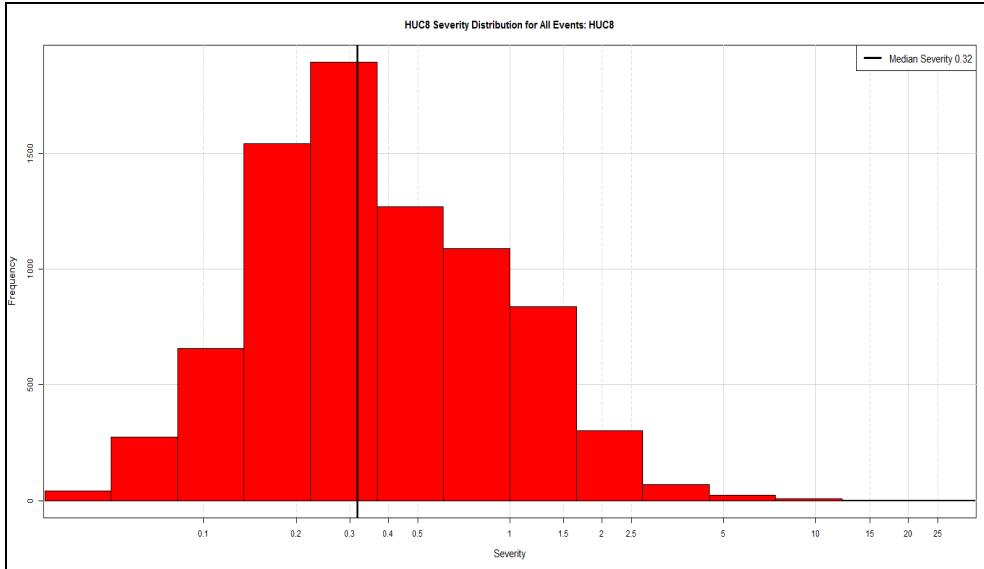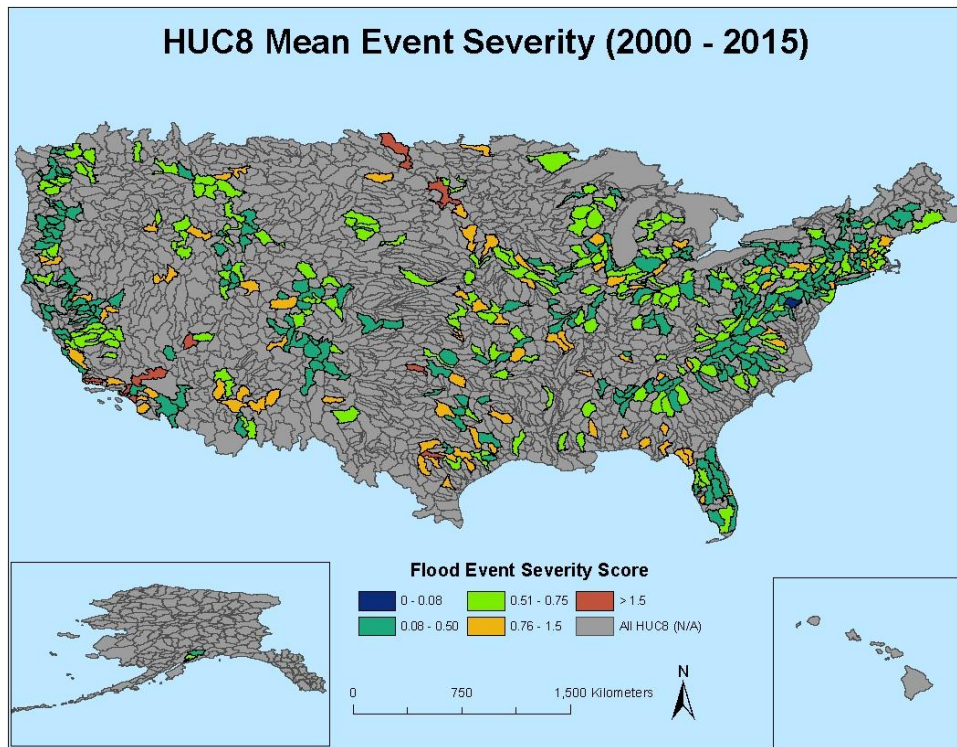573   ***Fig.10. Example Sites for Event Duration Concerns***
574

Fig.11. Severity Score Distribution

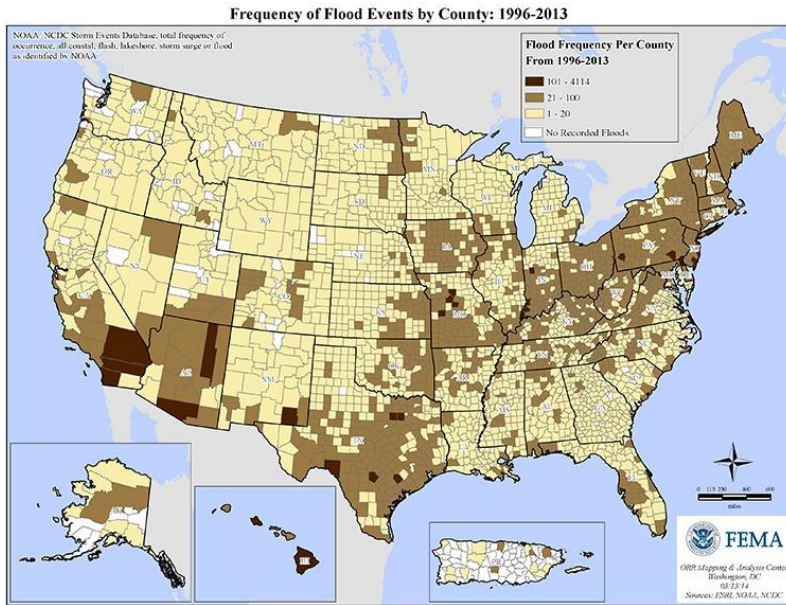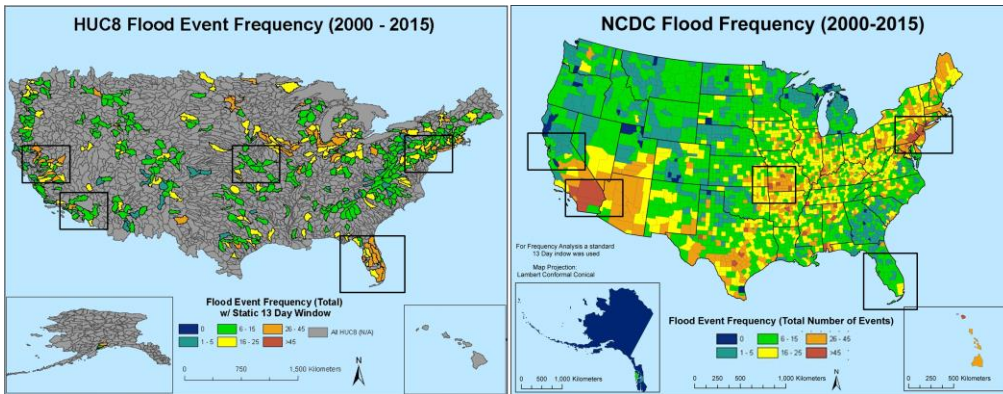*Fig.12. Regional Distribution of Severity*

Fig.13 FEMA Flood Frequency Estimates

Fig.14. Frequency Comparisons with a 13 Day Window (NCDC & Daily Discharge)

580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595

References:

596
597
598 Bacova-Mitkova V, Onderka M. 2010. "ANALYSIS OF EXTREME HYDROLOGICAL
599 EVENTS ON THE DANUBE USING THE PEAK OVER THRESHOLD METHOD." Journal
600 of Hydrology and Hydromechanics, 58: 88-101. DOI: 10.2478/v10098-010-0009-x.
601
602 Barredo, J. I., Saurí, D., and Llasat, M. C.: "Assessing trends in insured losses from floods in
603 Spain 1971–2008", Nat. Hazards Earth Syst. Sci., 12, 1723-1729, doi:10.5194/nhess-12-1723-
604 2012, 2012.
605
606 Black A. R. and Werritty A. "Seasonality of flooding: a case study of North Britain." Journal of
607 Hydrology, Volume 195, Issue 1, 1997, Pages 1-25, ISSN 0022-1694,
608 http://dx.doi.org/10.1016/S0022-1694(96)03264-7.
609 Dobour, J. C. and J. Noel. "A CLIMATOLOGICAL ASSESSMENT OF FLOOD EVENTS IN
610 GEORGIA." *National Weather Service* (2005): n. pag. Web.
611 Doocy S., Daniels A., Murray S., Kirsch TD.: "The Human Impact of Floods: a Historical
612 Review of Events 1980-2009 and Systematic Literature Review." PLOS Currents Disasters.
613 2013 Apr 16, Edition 1. doi: 10.1371/currents.dis.f4deb457904936b07c09daa98ee8171a.
614 Gaffin, D. M. and D. G. Hotz. "PRECIPITATION AND FLOOD CLIMATOLOGY WITH
615 SYNOPTIC FEATURES OF HEAVY RAINFALL ACROSS THE SOUTHERN
616 APPALACHIAN MOUNTAINS." *NOAA/National Weather Service* 24.3 (2000): n. pag. Web.
617 Himmelsbach, I., Glaser, R., Schoenbein, J., Riemann, D., and Martin, B.: "Reconstruction of
618 flood events based on documentary data and transnational flood risk analysis of the Upper Rhine
619 and its French and German tributaries since AD 1480." Hydrol. Earth Syst. Sci., 19, 4149-4164,
620 doi:10.5194/hess-19-4149-2015, 2015.
621 Joyce, Christopher. "Federal Flood Insurance Program Drowning In Debt. Who Will Pay?" *NPR*.
622 NPR, 1 Jan. 2014. Web. 05 Aug. 2016.
623 Kahana, R., Ziv, B., Enzel, Y., and Dayan, U.: "Synoptic climatology of major floods in the
624 Negev Desert, Israel." Int. J. Climatol., 22, 867–882, 2002.
625 Karnbaum K. and Wolfgang Kron, Dr. "What Is a Flood? Defining Flood Loss Occurrences for
626 Reinsurance Purposes." *Munich Re* (2005): n. pag. Web.
627 Mallakpour, I., and G. Villarini. "The Changing Nature of Flooding across the Central United
628 States." *Nature Climate Change* 5.3 (2015): 250-54. Web.
629
630 Mallakpour, I., and G. Villarini, Investigating the relationship between the frequency of flooding
631 over the central United States and large-scale climate, *Advances in Water Resources*, 92, 159-
632 171, 2016.

633   "NATIONAL WEATHER SERVICE INSTRUCTION 10-1605." *National Climatic Data Center*
634   *(2007)*: n. pag. Web.

635   Nied M., Pardowitz T., Nissen K., Ulbrich U., Hundecha Y., Merz B., "On the relationship
636   between hydro-meteorological patterns and flood types." Journal of Hydrology, Volume 519,
637   Part D, 27 November 2014, Pages 3249-3262, ISSN 0022-1694,
638   http://dx.doi.org/10.1016/j.jhydrol.2014.09.089.
639
640   ORR Mapping and Analysis Center, Washington, D.C., March 13, 2014. Sources: ESRI, USGS.

641   "Overwhelming Risk: Rethinking Flood Insurance in a World of Rising Seas (2013)." *Union of*
642   *Concerned Scientists*. N.p., Aug. 2013. Web. 05 Nov. 2016.

643   "The National Flood Insurance Program." *FEMA.gov*. FEMA, 20 May 2015. Web. 10 Aug.
644   2016.
645
646   Uhlemann, S., Thieken, A. H., and Merz, B.: "A consistent set of trans-basin floods in Germany
647   between 1952–2002." Hydrol. Earth Syst. Sci., 14, 1277–1295, doi:10.5194/hess-14-1277-2010,
648   2010.
649
650