

## **Response to John England:**

### **Major Comments**

This paper is an interesting case study using non-stationary concepts to illustrate a changing flood risk profile. The authors suggest that the paper “provides the first end to- end analysis using non-stationary GEV methods coupled with contemporary downscaled climate projections to demonstrate how the risk profile of existing infrastructure evolves with time over its design life.”

In my opinion, this topic is very interesting, and certainly timely and relevant to HESS. Given ongoing activities within the United States to revise flood frequency guidelines (England and Cohn, 2008; HFAWG, 2013; England et al., 2013) this paper could potentially serve as a reference or example methodology to estimate flood frequency and flood risk in a non-stationary environment dominated by future trends in climate (precipitation and temperature), at least within the U.S. It is a very nice example of illustrating time-varying distribution parameters, relevant for flood frequency applications.

***Response: We thank the reviewer for acknowledging the contribution of our work. We appreciate his assessment and hope that this paper can serve as a real world example of non-stationary flood techniques for others to follow.***

However, in my opinion, there are some technical issues with the methodology and application, as presented in this paper, that need further investigation and/or simply explanation, prior to publication in HESS.

***Response: We appreciate the thoughtful suggestions and have addressed every comment below.***

1. Functional form of flood non-stationarity (equations 2 and 3), covariates and nonstationary parameter selection.

One key issue in non-stationary flood frequency analysis is the identification of the functional form of non-stationary behavior. For example, Salas and Obeysekera (2013) highlight three situations: increasing floods; decreasing sea levels; and shifting floods. Other examples are in Olsen et al. (1998) and Stedinger and Griffis (2011).

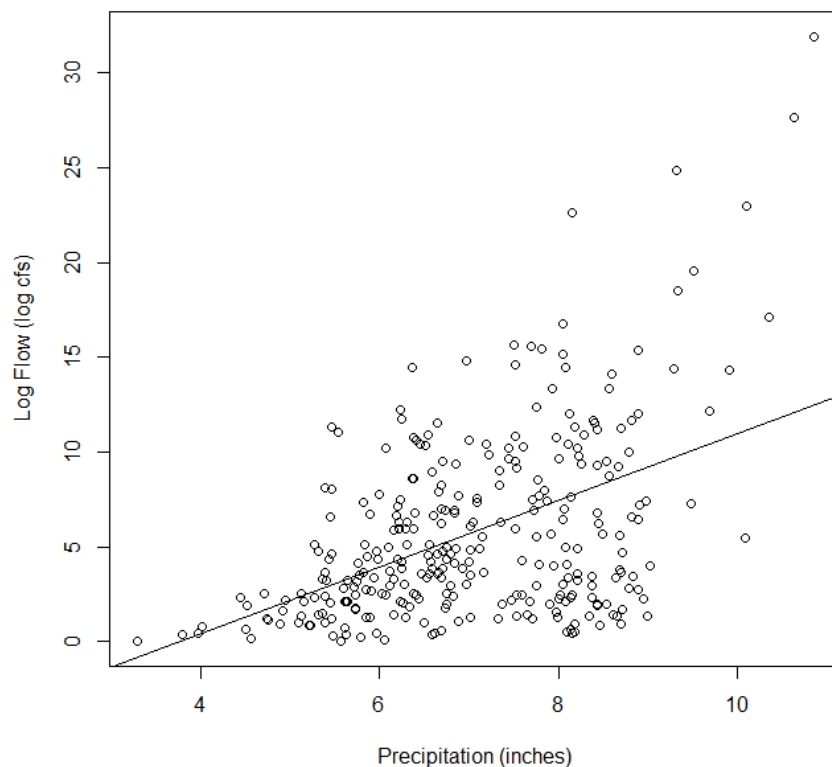
With all the information behind Table 2 on covariates (page 5105), and presumed functional form (equations 2 and 3), the authors need to clearly demonstrate the evolution of aggregated winter monthly precipitation and monthly average temperature trends with time. Are trends in location and scale needed? Are there any shifts in these time series? Why are these linear models adequate?

***Response: We do not observe statistically significant trends in precipitation or temperature over the historical period of the study. However, there are changes in both variables looking into the future. Salas and Obeysekera (2013) and Stedinger and Graffis (2011) both develop model parameters that depend explicitly on time and therefore they must both derive a trend from historical data and make assumptions about how this trend will continue into the future. In our approach, we vary the location and scale parameters only as a function of the covariates precipitation and temperature and not explicitly as a function of time. In other words, a temporal trend will only be introduced if there is a trend in the relevant climate variables. Thus even though we didn't observe a strong temperature trend in the historical period of record, we can still capture the impact of future trends as long as we have adequately characterized the connection between temperature and flooding. Given this***

*approach, it is not important to demonstrate historical trends in variables but rather connections between our covariates and flooding. To address this point we've plotted precipitation and log flow within the flood months at Farad. As shown here, there is a positive connection between flooding and precipitation. However it should also be noted that the connections we are investigating in our model go beyond simple correlation because we use precipitation as a covariate for both the location and scale of the distribution.*

*We agree with the reviewer that this this is an important distinction to make clear and therefore we have added the following text the revised manuscript.*

*“Some previous studies [e.g. Salas and Obeysekera, 2013; Stedinger and Griffis, 2011] developed non-stationary location and scale parameters that are explicitly dependent on time. This approach requires first, the derivation of temporal flooding trends and second, the projection of this trend into the future. Here we derive location and scale parameters based on time varying meteorological variables (i.e. temperature and precipitation). With the approach used here, temporal trends in flooding are introduced as a function of temporal variability in precipitation and temperature but no explicit trend is specified apriori.” (Revised Manuscript, Page 12, Lines 7-13)*



Describe model performance between nonstationary location, and nonstationary location and scale. Do model diagnostics clearly show that the scale parameter needs to be included (adding 3 more coefficients), or would a simpler model be sufficient just based on location (e.g., Towler et al., 2010)?

Figure 2 isn't compelling to distinguish between competing models, or discriminate between tail behavior at the 2 sites.

**Response: Table two compares model performance for models with nonstationary location, nonstationary scale and both. As discussed on page 5088 lines 10-17 we compare model performance using the Akaike Information Criterion which weighs model complexity against the goodness of fit. As shown in table two the models with non-stationary location and scale perform significantly better than the non-stationary location models even when the added complexity is taken into account.**

**To make this point more clearly we have modified Table 2 to report AIC scores as well as deviance statistics and the number of model parameters. Also, we have expanded to the discussion of metrics as follows:**

**“Here  $nllh$  is the negative log likelihood estimated for a model fitted with  $K$  parameters. In this formulation, higher ranked models have lower AIC scores. For this analysis the best model is selected using pairwise comparisons of NLLH scores following the methods of Salas and Obeysekera [2014] and others. Models are compared using the deviance statistic ( $D$ ) which is equal to twice the difference in NLLH scores. Deviance statistics are then tested for significance tests based on a chi-squared distribution with the degrees of freedom set equal to the difference in the number of parameters ( $K$ ) between models.  $P$ -values less than 0.05 indicate a statistically significant (alpha of 0.05) improvement in model performance.” (Revised Manuscript, Page 12, Lines 19-25)**

Supplemental material could be used here to build your case for the functional form and the choice to use two time-varying parameters. You could also show the simple plots of precipitation and temperature evolution over time, for historical and future periods. This would help partly explain the non-overlapping risks, depending on period (Figure 5). You could also show cdfs from the 2 sites, enhancing and further explaining the discrepancies on the shape parameters between these two nested watershed sites.

**Response: As noted above our method does not rely on the demonstration of historical trends. However we thank the reviewer for their comment and we agree that this point was not made clearly in the original manuscript. In response to the questions posed here we have expanded the text (as detailed above) with respect to the model form for the covariates as well as the model selection.**

2. Discussion on quantiles of interest, record length used in analysis, and risk

The authors need to clearly articulate what they mean by infrastructure (levees? Flood protection?), quantiles of interest (Q50? Q100?), and how their key measure is risk (of failure, in the binomial sense).

**Response: We appreciate the suggestion and have added the following clarifications to the text:**

**Page 5- The first instance of the term ‘infrastructure’ in the manuscript now reads:**

**“Historically, most flood infrastructure that is vulnerable to flooding (e.g. dams, levees, sewers and bridges) has been designed to withstand flooding of specified return period (e.g. the 100 year flood)” (Revised Manuscript, Page 6, Lines 3-5)**

**Also we have modified the reference in the abstract to infrastructure to read:**

**“This paper provides the first end-to-end analysis using non-stationary GEV methods coupled with contemporary downscaled climate projections to demonstrate the evolution of flood risk**

*profile over typical design life periods of existing infrastructure that is vulnerable to flooding (e.g. dams, levees, bridges, and sewers).” (Revised Manuscript, Page 2, Lines 12-16)*

*We do not see the term “quantiles of interest” in the manuscript. As discussed on page 5093 lines 1-5, we focus our analysis on the historically derived ‘design flood’ of 37,600 cfs.*

*We have further clarified our use of the term risk in the abstract and within the main text as follows:  
“The resulting exceedance probabilities are combined to calculate the probability of a flood of a given magnitude occurring over a specific time period (referred to as flood risk) using recent developments in design life risk methodologies.” (Revised Manuscript, Page 2, Lines 10-12)*

*“Note that following the convention of Rootzén and Katz [2013] we use the term flood risk in the non-technical sense to refer to the probability of an extreme event occurring and not as a quantification of expected losses.” (Revised Manuscript, Page 6, Lines 18-20)*

*“This concept is easily extended to flood risk (here defined as the probability of a flood of a given magnitude occurring, not expected losses).” (Revised Manuscript, Page 13, Lines 30-31)*

How does the record length chosen (50 year windows) affect the quantile estimates, particularly the precision of the shape parameters at these two sites, for the flows of interest (37600 cfs)?

*Response: We decided to use a 50 year window because this is the maximum time span we could use based on the length of our historical record. While we didn’t quantify the impacts of using a shorter window, we understand that working with the longest record available provides us with the most points to fit our model with. Using shorter window will only add to sampling variability and hence add additional uncertainty to the estimated model parameters.*

Could you better explain (page 5088), that using block maxima (6 per year), your main output is risk (and changes) for a particular flow level (Fig 5), rather than a “traditional” annual flood frequency curve? Some minor adjustments would be needed to estimate annual one-day maximum flood probabilities. You could cite (for example) Towler et al. (2010, p. 3).

*Response: As detailed on page 5089 lines 14-20 The reason why we can’t make the traditional annual flood frequency curves is because our flood frequency changes over time and therefore the probability of a flood occurring depends both on the length of time considered and the period of interest. For this reason we adopted the design life methodology. The reviewer is correct in pointing out that the use of block maxima does impact the probabilities that are estimated. However, we have taken this into account in our calculations. To make this point more clearly we have added the following text to page 11 of the revised manuscript:*

*“However, as noted by Towler et al. [2010], when multiple values are used per year the calculated probabilities must be adjusted appropriately to derive annual values.” (Revised Manuscript, Page 13, Lines 6-8)*

3. Uncertainty estimates for quantiles

Uncertainty estimates are required (NRC, 2000; USACE, 2011) to evaluate flood risk and alternatives in a stationary environment. In a non-stationarity framework, there are existing tools (Gilleland and Katz, 2011; Obeysekera and Salas, 2014) to estimate uncertainty. The authors need to address or add a discussion of uncertainty of the estimates provided. Would confidence intervals for the future period(s) overlap the stationary model estimates?

***Response: For this analysis we have focused on uncertainty with respect to future climate as opposed to model parameter uncertainty. In response to the comments of other reviewers we have added discussion of uncertainty with respect to the VIC model on page 17 of the revised manuscript. In response to this comment we reviewed Gilleland and Katz (2011) as well as Obeysekera and Salas (2014). While Gilleland and Katz (2011) do not provide any specific details on uncertainty estimation Obeysekera and Salas (2014) outline several approaches. However as noted on p.1439 of Obeysekera and Salas (2014), the methods they discuss are designed to assess uncertainty in flow quantiles not the uncertainty in return periods, which is what our analysis is. We acknowledge that there are methods for handling parameter uncertainty, however following the approach of Towler et al. (2010) we are focusing our analysis on shifts in the distributions as a function of climate change.***

4. Use of the term “Data”, consideration and use of additional data relevant for floods and flood frequency

The authors need to add some caveats in Sections 2.2. and 2.3 on the use of the term “data” that include VIC unregulated flow simulations and CMIP-5 downscaled projections. In both situations these are model results, not data. I suggest a simple title change of each section to “Streamflow data and models” and “Climate data and models”.

***Response: We appreciate the suggestion and agree with the reviewer that our current terminology could be confusing. We have changed the title of section 2.2 to “Streamflow data and simulations” and 2.3 to “Climate data and models”. In addition we have gone through the manuscript and revised any sentences that referred to model results as data.***

Why were not other streamflow data sources considered, particularly for calibration of VIC soil moisture parameters? There are monthly unregulated streamflow estimates available in the Truckee River Basin (Rieker and Cowan, 2005) that may be useful. How would one extend the historical period, with evidence of several very large floods, for the period prior to 1950? There are at least 3 large floods prior to 1950 that are equal to or exceed those used in the analysis.

***Response: We made every effort to obtain additional unregulated flow data for the study area. In the end we limited our analysis to the documented unregulated flow values we found. We appreciate the suggestion of an additional dataset and will consider this data if we do additional analysis in the future. Our historical period is limited by the meteorological dataset we used. In order to extend our analysis we would need a longer historical forcing dataset. Furthermore, we need to be consistent with the historical period as this has implications in the way future climate projections were developed.***

On streamflow, please explain how VIC was calibrated (or not) and how parameter estimates were made for the historic period (Section 2.2). Why are these stationary streamflow model parameters representative of future periods? Validation is mentioned (page 5084, line 17), but only in reference to six one-day flows (Figure 3). Were these six events used in any way for model calibration?

**Response: The model calibration was completed as part of the larger West Wide Climate Risk Assessment project (Reclamation, 2011). We have added a paragraph summarizing this and provided the appropriate references for calibration as follows:**

***“The VIC model used for this analysis was developed and calibrated as part of the Bureau of Reclamation’s West Wide Climate Risk Assessment (WWCRA). The WWCRA VIC model encompasses the western US. Streamflows were evaluated at 152 locations from the USGS Hydroclimatic Data Network [Slack et al., 1993] and 43 additional locations of importance to Reclamations water management activities. Among the evaluated locations are several in the Truckee basin including the Truckee River at Farad. For details on model calibration and development we refer the reader to Reclamation [2011] and Gangopadhyay [2011]. While we do not discuss model calibration further here, in the subsequent sections we provide additional model verification for flood simulation in the UTRB.” (Revised Manuscript, Page 9, Lines 7-16)***

**We have validated our GEV model using the six observed flood events that we have unregulated flow values for as well as the time series of simulated flows from the Calibrated VIC model.**

**As discussed in our response to comment 1, our approach does not require any assumptions about current and future trends. Rather the model parameters that we derive relate flood occurrence to meteorological variables. Therefore, we assume only that the physical mechanism between precipitation and temperature and flooding will remain the same in the future. This approach is well documented in the other studies that we cite.**

#### **Some Specific Comments**

page 5078, line 14: Clarify that existing infrastructure – might be appropriate for floodplain management activities.

**Response: We have modified the sentence in question to read:**

***“This paper provides the first end-to-end analysis using non-stationary GEV methods coupled with contemporary downscaled climate projections to demonstrate the evolution of flood risk profile over typical design life periods of existing infrastructure that is vulnerable to flooding (e.g. dams, levees bridges and sewers)” (Revised Manuscript, Page 2, Lines 12-16)***

page 5079, lines 28-29. Notably missing in the citations to US flood trends is Hirsch (2011).

**Response: Thank you for the suggestion. We have added the following reference to Hirsh in the paragraph subsequent to the one you reference.**

***“Hirsch [2011] noted both increasing and decreasing trends in annual flood magnitudes in different regions of the US” (Revised Manuscript, Page 4, Lines 6-7)***

page 5082, lines 21-22. May want to expand this crucial definition of risk.

**Response: We have added the following definition to this section:**

***“Note that following the convention of Rootzén and Katz [2013] we use the term flood risk to refer to the probability of an extreme event occurring and not as a quantification of expected losses.” (Revised Manuscript, Page 6, Lines 18-20)***

page 5083, lines 20-22. This statement is erroneous. Snowmelt flood peaks from April to July are nearly always within the channel; peaks are not substantially reduced by upstream storage as reservoirs are usually full. Forecasting plays little to no role in reducing flood peaks using reservoir storage. Peaks and max daily flows in these months are much, much smaller than winter floods.

***Response: While the reviewer is correct that the reservoirs are usually full at this time of the year, our understanding from water managers in the basin is that snowmelt floods are still managed by reservoirs because predictions allow managers to evacuate storage space in advance.***

page 5084, line 17. You mention validation; can you describe calibration here?

***Response: Please refer to comment 4 above. We have added text to describe model calibration.***

page 5085, lines 15-16. Can you demonstrate that cumulative monthly precipitation in winter months is adequate, rather than pairing individual monthly precipitation and flow estimates?

***Response: We did not experiment with total winter precipitation because we wanted to use multiple flood values per year. If we used just one precipitation value per year then we would have less basis for simulating multiple floods. We appreciate the suggestion though and freely acknowledge that there are many potential predictors that could be experimented with. However, for this assessment we chose to stick with published approaches.***

page 5089, line 22-23. Highlight a bit more here what the “risk of flood” means. Can you also clarify here (Table 1) how you obtain an annual “ $p$ ” (in year  $x$ ) from the GEV for use in these binomial risk models, or that annual probabilities are not needed?

***Response: We have appreciate the suggestion and have added the following definition:***

***“This concept is easily extended to flood risk (here defined as the probability of a flood of a given magnitude occurring, not expected losses).” (Revised Manuscript, Page 13, Lines 30-31)***

***We have adjusted all of our analysis to account for the fact that we have 6 values per year not 1. In response to comment 4 we added the following clarification:***

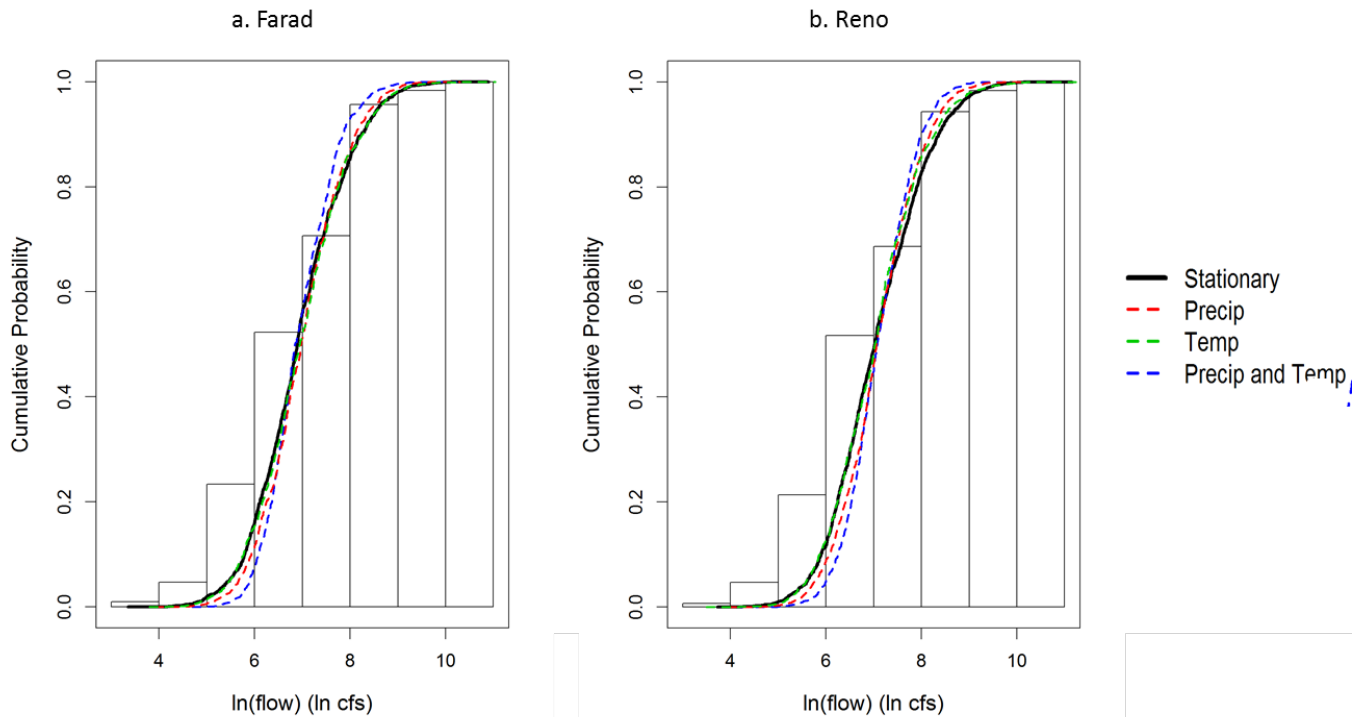
***“However, as noted by Towler et al. [2010], when multiple values are used per year the calculated probabilities must be adjusted appropriately to derive annual values.” (Revised Manuscript, Page 13, Lines 6-8)***

page 5090, discussion on Table 2. Here is where more explanation and justification is needed on the use of time-varying location and scale.

***Response: Please refer to the response to general comment 1.***

page 5090, line 24. Recommend using supplemental material to show cdfs of these distributions. The pdfs do not adequately show the right-hand tail to demonstrate what you are saying here.

**Response: We have provided the corresponding cdfs below. While this is of course a matter of personal preference, we do not feel that the cdfs provide any improvements in displaying the right-hand tails. Because the cdf is a monotonically increasing function, all of the curves converge to 1 on the right-hand tail.**



page 5091, line 1. Did you fix the shape parameter in any way across future months? Why would it vary in time?

**Response: Yes the shape parameter remains fixed. To clarify this point we have modified the sentence in question to read:**

***“Using the coefficients determined above, the location and scale and shape parameters are calculated for every climate projection (i.e. 234) and flood season month (i.e. November to April 1950 to 2099) based on the downscaled precipitation and temperature values detailed in Section 2 (Note that the scale parameter remains fixed).” (Revised Manuscript, Page 15, Lines 4-5)***

page 5091, Section 4.2. The term “ungaged” is in error throughout this paragraph. These flows are “unregulated”. Gages exist at both locations. You could have compared the USACE unregulated flows with those from the gage, and examined the differences.

**Response: The reviewer is correct. We have replaced all instances of “ungaged” with “unregulated” in this section and throughout the manuscript. While we appreciate the suggestion the purpose of this analysis is not to understand the adjustments that were made to generate the unregulated flow**



**estimates. These adjustments account mainly for the reservoir operations and diversions within the basin. We refer the reader to the USACE documentation for these details.**

page 5092, lines 10-15. Citations are needed to support these numbers. Can you explain the precipitation gradient from upstream to downstream, and where the snow resides in this basin? It may help explain the differences in shape parameters at Farad (relatively fatter tail) and Reno, and why these distributions are bounded.

**Response: We have added a citation to the sentence in question. Also, we appreciate the suggestion and have added the following discussion to the basin description in section 2.1:**

***“Most of the available water supply is generated upstream of the Farad Gauge [USACE, 2013a]. The Reno gauge is located downstream of Farad in the heart of Reno and is a good reference point for analyzing urban flooding. The intervening area between the Farad and Reno gauges is small, roughly 350 square kilometers and there are only two small tributaries that enter the main stem between Farad (Reno Dog Creek and Hunter Creek).” (Revised Manuscript, Page 7, Lines 15-20)***

page 5093, lines 15-16. Clarify here the uncertainty of the GEV model parameters and quantile estimates for the stationary case (as well as other cases) is not included. How might its inclusion change the perception shown in Figure 5?

**Response: Please refer to major comment 3 above.**

page 5094, line 10. Clarify that you use MLEs to estimate the LP3 parameters (same as GEV), so that the differences you mention truly are the models, and not mixed with parameter estimation. Moments and MLEs of the LP3 can sometimes give very different results for the shape parameter.

**Response: Following the methodology outlined in Bulletin 17B we fit the LP3 distribution using L-moments. However, the reviewer is correct that this does introduce additional uncertainty because the stationary GEV model was fit using MLEs. We have decided to maintain the current LP3 fit because this is an accurate representation of standard practice. In response to this comment though, we have modified the text as follows to make this point more clearly:**

***“Also, the risk calculated using a stationary GEV model and a stationary LP3 model (i.e. the distribution prescribed by Bulletin 17B and fit using the L-moments methodology [IACWD, 1982]) fit to the historical flow data are plotted for reference (blue and red dashed lines respectively). Comparing between these three approaches (non-stationary GEV, stationary GEV and stationary LP3) provides information on the sensitivity of results to model approach and non-stationary parameters. For instance, both stationary models are fit to the same historical simulated flows (one using MLE and the other using L-moments) so differences between the stationary lines reflect the impact of model choice and fitting approach on estimated risk. Conversely the stationary GEV model (blue line) and the historical non-stationary models (grey boxplot) have the same model form and cover the same time period; the only difference is the addition of covariates to estimate model parameters. Thus differences between these two show the effect of model parameter changes from the non-stationary approach.” (Revised Manuscript, Page 18, Lines 5-17)***

Fig 3. Correct caption to read “unregulated” not “ungaged”.

**Response: This has been corrected.**

If you had used the shape parameter from Farad, or some weighted combination of Farad and Reno, how much would the tails of these GEV distributions widen?

***Response: This is an interesting question that we didn't explicitly consider. We have not seen any research that looks at spatial combinations of GEV distributions but this could be an interesting topic for future research.***

**References Cited and Other Relevant Information** (with hyperlinks to pdf copies, as available)

England, J.F. Jr. and Cohn, T.A. (2008) Bulletin 17B Flood Frequency Revisions: Practical Software and Test Comparison Results, American Society of Civil Engineers, EWRI World Water Environmental Resources Congress, May 13-16, 2008, Honolulu, HI, 11 p.

[https://sites.google.com/a/alumni.colostate.edu/jengland/file-upload/england\\_cohn\\_b17b\\_asce2008\\_final.pdf?attredirects=0&d=1](https://sites.google.com/a/alumni.colostate.edu/jengland/file-upload/england_cohn_b17b_asce2008_final.pdf?attredirects=0&d=1)

England, J.F. Jr., Cohn, T.A., Faber, B.A., Stedinger, J.R., Thomas, W.O., and Mason, R.R. (2013) Revisions Recommended to Bulletin 17B – US National Flood Frequency Guidelines, Abstract H51Q-03, presented at 2013 Fall Meeting, AGU, San Francisco, Calif., 9-13 Dec.

Gilleland, E., and Katz, R. W. (2011) New software to analyze how extremes change over time." Eos, 92(2), 13-14.

Hirsch, R.M. (2011) A Perspective on Nonstationarity and Water Management, Journal of the American Water Resources Association, 47, 436-446.

Hydrologic Frequency Analysis Work Group (HFAWG) (2013) Recommended Revisions to Bulletin 17B. Memorandum to Subcommittee on Hydrology, Advisory Committee on Water Information, 10 p., <http://acwi.gov/hydrology/Frequency/HFAWG-B71B-Recommended-Rev-SOH-12June2013-revised-final.pdf>

National Research Council (2000). Risk analysis and uncertainty in flood damage reduction studies. Technical report, National Academy Press.

Obeysekera, J. and Salas, J.D. (2014) Quantifying the Uncertainty of Design Floods under Nonstationary Conditions, J. Hydrolog. Eng., ASCE, 19(7), 1438-1446, doi: 10.1061/(ASCE)HE.1943-5584.0000931

Olsen, J.R., J.R. Stedinger, N.C. Matalas, and E.Z. Stakhiv (1999) Climate Variability and Flood Frequency Estimation for the Upper Mississippi and Lower Missouri Rivers. Journal of the American Water Resources Association 35(6):1509-1524.

Reiker J.D., and D.M. Cowan (2005), Documentation for Hydrologic Data: Truckee RiverWare Model, Technical Memorandum, U.S. Bureau of Reclamation.

Stedinger, J. R. and Griffis, V. W. (2011). Getting from here to where? Flood frequency analysis and climate. JAWRA Journal of the American Water Resources Association, 47(3):506–513.

Towler, E., Rajagopalan, B., Gilleland, E., Summers, R. S., Yates, D., and Katz, R. W (2010) Modeling hydrologic and water quality extremes in a changing climate: a statistical approach based on extreme value theory, Water Resour. Res., 46, W11504, doi:10.1029/2009WR008876

U.S. Army Corps of Engineers (2006). Risk analysis for flood damage reduction studies. Technical Report ER 1105-2-101, U.S. Army Corps of Engineers.