Response to Referee 1:

In this paper, a non-stationary approach is applied for the estimation of the probability of failure of infrastructures in two locations in the Upper Truckee River Basin (US). The approach uses climate scenarios as input to determine the expected (and range of) changes in precipitation and temperature. The results show that, based on the assumptions made, the probability of failure of infrastructures increases considerably with time, from now to the end of the century.

The paper is well written and interesting. I like the fact that the evolution of probability of failure is investigated, instead of the change in return period. However I have the following concerns, which I believe should be addressed/discussed before publication in HESS:

Response: We thank the reviewer for their thoughtful review. We have addressed all of the specific comments below.

1) What is the novelty of the paper? Non-stationary models for flood hazard are not new and nor is the use of the probability of failure in climate change studies (i.e., the "design life level" of Rootz and Katz, 2013).

Response: We appreciate the reviewers comment and would like to note that we are not claiming that this is a new method. Rather the purpose of this work is to demonstrate how this approach can be applied. In their both their 2013 and 2014 papers, Salas and Obeysekera comment that, although non-stationary flood methods have been well documented there has been 'insufficient' attention paid in the water resources engineering literature. Our goal with this paper was to demonstrate the application of these methods for regional decision making. We have modified the end of the introduction to make this point more clearly as follows:

> ***"To address this issue, Rootzén and Katz [2013] introduced the concept of design life level to calculate the risk of a given flood magnitude occurring over a specified time period. Salas and Obeysekera [2014] further demonstrated the relevance of this technique to hydrologic community using flood frequency examples. However, this methodology has yet to receive widespread attention within the hydrologic community. Here, we present a non-stationary flood frequency assessment for the Upper Truckee River Basin (UTRB) using contemporary downscaled climate projections and the non-stationary design life level technique introduced by Rootzén and Katz [2013] to quantify flood risk (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). While the methodology used for this analysis is previously established, this paper provides the first end-to-end demonstration of non-stationary GEV analysis coupled with contemporary downscaled climate projections (specifically, downscaled climate projections from the Coupled Model Intercomparison Project Phase-5 (CMIP-5)), to quantify how the flood risk profiles may evolve in the Truckee river basin over the next century. The intent of this work is 1) to investigate potential flood risk changes over time in the Truckee basin and 2) to demonstrate the applicability of non-stationary techniques in a regional flood analysis to make these tools more accessible to the hydrologic community." (Revised Manuscript, Page 6, Lines 11-28)***

2) The results are conditioned to strong assumptions and there is no explicit uncertainty analysis in the paper (e.g., Steinschneider et al., 2012, provide a framework for that). Prediction bounds are plotted in the figures, but they just show the range of variability of climate model inputs once propagated through the hydrologic models (VIC + non-stationary GEV). In my opinion, it would have been more interesting to analyse if, based on the observations in the last decades, the use of non stationary flood-frequency models gives results significantly different from those obtained with stationary models. To do so, the uncertainty associated to the use of both approaches should be quantified: Fig. 6 could contain the stationary models results with confidence bounds + the non-stationary model results for the observation period with confidence bounds that account not only for the variability of the climate models, but also for the uncertainty in the estimation of (VIC+ non-stationary GEV) model parameters. It would be very interesting to see how the two ranges of estimates differ.

Response: We appreciate the reviewer's suggestion and agree that our approach only encompasses uncertainty in the climate model inputs. Recent analysis by Elsner et al. (2014) did investigate uncertainties in historical forcings and their impact on VIC simulations. In response to this comment, we have added the following clarification to section 4.3 of the manuscript:

> ***"Here we use the interquartile range, as opposed to the 5th and 95th percentile, to focus on the central tendencies of each time period and not the variability between projections. Note that the ranges presented here express the variability between climate models. Uncertainty in the historical data sources used for calibration and in the parameters of the VIC model are not investigated directly here. For more detailed analysis on uncertainty in VIC simulations the reader is referred to Elsner et al. [2014]." (Revised Manuscript, Page 18, Lines 11-15)***

In response to this comment we also explored uncertainty in our results based on the historical variability of the GCM simulations. Based on this analysis we added uncertainty bounds to the parameter estimates in table three and added the following discussion to the manuscript:

> ***"To address uncertainty, models of the same form (i.e. non-stationary location and scale with precipitation and temperature as covariates) were also fit to the historical simulation period (1950-1999) using downscaled precipitation and temperature from all 234 climate projections. Because each climate projection seeks to reproduce historical behaviour over the historical period, the variability between projections in this time frame is a measure of uncertainty in model coefficients given the same physical system. This differs from the variability between climate projections in future periods (i.e. after 1999) which is a measure of uncertainty in future forcing conditions. Table 3 shows the interquartile range of model coefficients calculated from the 234 historical GCM simulations.***
>
> ***Using these parameters the return period of the design flood at Reno (37,600cfs) was calculated for every set of model parameters using observed historical precipitation and temperature. The observed model estimates a return period of 45 years while the interquartile range (IQR) using the simulated model parameters (i.e., the model parameters estimated from each of the 234 historical GCMs) with observed precipitation and temperature varies from 28 to 247 years. Note that the return period of 45-years estimated from observed meteorology is within the IQR of 28 to 247 years. Although the IQR is large it should be kept in mind that some of the uncertainty in this range is a result of the downscaling methodology. The monthly BCSD algorithm used for downscaling GCM climate only constrains the monthly precipitation***

and temperature statistics (total precipitation and mean monthly temperature) over the historical 1950-1999 period. Furthermore, uncertainty is introduced when monthly total precipitation and mean temperature are translated to daily values. Thus the estimated IQR implicitly captures downscaling uncertainties, in addition to explicitly representing parameter uncertainty. The need to consider uncertainties at each and every step of the process starting with, for example, downscaling methods (statistical, dynamical or some combination of statistical and dynamical methods) is a topic of ongoing research." (Revised Manuscript, Page 15 lines 7-31)

3) The Authors use the wording "flood risk" to refer to the probability of failure. Even though in engineering books "risk" and "probability of failure" are used interchangeably, "flood risk" is widely accepted in the literature as product of hazard (probability of flooding) and consequences (see e.g. Plate, 2002, among many). Since this paper looks at hazard only, I would strongly suggest to change the wording in it (including the title).

Response: We use the term 'risk' to be consistent with the design life literature we are citing (e.g. Salas and Obeysekara, 2014). However we acknowledge that the term risk can be used several ways. In response to this comment we have added the following clarifications to the manuscript:

"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)

"Here, we present a non-stationary flood frequency assessment for the Upper Truckee River Basin (UTRB) using contemporary downscaled climate projections and the non-stationary design life level technique introduced by Rootzén and Katz [2013] to quantify flood risk (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 15-20)

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

4) The references in the paper are biased toward US, while relevant literature exists abroad. As a suggestion, since I am European, the Authors could refer to some of the many studies cited in Hall et al. (2013) about flood changes in Europe (and scenario approaches).

Response: We thank the reviewer for pointing out this reference. We focused our introduction on studies in the Western US, as this is where our study area is. However, we acknowledge that there is also much to be learned from studies of other regions of the world. In response to this comment we have added the following reference to the introduction:

"Difficulty in diagnosing flood trends is not unique to the Western US; a literature review of historical flood studies across Europe also found spatial variability in flood trends [Hall et al., 2014]." ***(Revised Manuscript, Page 4, Lines 9-11)***

Also, through Hall et al. we found Merz et al. [2012] which provides a useful discussion on the drivers of change. We also added the following later in the introduction:

"Merz et al. [2012] note that attributing changes in flood hazard is complicated by the complex array of drivers that can include; land cover change and infrastructure development as well as natural climate variability and change. Here we set aside the impacts of development and management practices and focus on the role of climate change." (Revised Manuscript, Page 4, Lines 24-26)

<u>Specific comments:</u>

Page 5079, line 11: I do not agree with the wording "additional non-stationarity". It does not make sense unless stationarity is defined (see e.g., Koutsoyiannis, 2006; Montanari and Koutsoyiannis 2012). Under my understanding, stationary models can cope with long-term climate oscillations (see e.g., Koutsoyiannis, 2011).

Response: We agree with the reviewer that this sentence was unclear and have changed it to read:

> ***"Antropogenic climate change has the potential to amplify natural climatic variability throughout the interconnected climate and hydrologic systems" (Revised Manuscript, Page 13, Lines 13-15)***

Page 5083, line 7: being HESS international (and European) international unit system (e.g., km instead of miles) should be used. This comment applies to the all paper.

Response: We have changed all units to SI.

Page 5088, Eqs. (2) and (3): one line could be added to motivate why the shape parameter xi is considered stationary.

Response: We have added the following clarification following equations 2 and 3:

> ***"In keeping with previous studies the shape parameter, which is the most difficult to estimate, is assumed constant [e.g.*** **Obeysekara and Salas, 2014; Salas and Obeysekera, 2013; Towler et al., 2010*****]." (Revised Manuscript, Page 12, Lines 3-6)***

Page 5088, line 18: are the GEV distributions fitted to simulated streamflows only? The Authors should add a line here to motivate why the observed streamflows are not used here. I see that Fig. 2 and 3 include observed flows and provide a kind of validation of the procedure.

Response: Unfortunately we only have unregulated flow estimates for six flood events which would not be enough to fit the GEV distributions alone. We decided not to mix simulated and 'observed' flow datasets in order to be consistent. The unregulated flows are estimated by adjusting observed gauge flows for water resources management, i.e. water deliveries to meet agricultural water demands and reservoir storage changes. There are fundamental differences in the underlying processes (and assumptions) to estimate unregulated flows from gauge data and to simulate flows using a hydrologic model. Hence, it is not appropriate to combine flows from these two disparate sources. In response to this comment we have added the following clarification to the text.

“Although, there are some unregulated historical flow estimates, the available dataset only covers six storms. Therefore, to be consistent we fit our model only to the simulated flows.” (Revised Manuscript, Page 12-13, Lines 28 & 1-2)

Page 5089, Section 3.2: the section discusses “flood hazard”, not “flood risk”. The same applies to the rest of the paper, specially to Section 4.3 and related figures (e.g., y-axis in Figs. 5-8 should not be “risk” but “probability”)

Response: Please refer to the response to comment 3. We have decided to use the term ‘risk’ to be consistent with the terminology of the design life level we apply. However, to avoid confusion we have added clarification throughout the text. Also in response to this comment we have changed the captions of the figures in question.

Additional References:

Hall, J., et al. (2013) Understanding flood regime changes in Europe: a state of the art assessment, Hydrol. Earth Syst. Sci. Discuss., 10, 15525-15624, doi:10.5194/hessd-10-15525-2013.

Koutsoyiannis, D. (2006), Nonstationarity versus scaling in hydrology, J. Hydrol., 324, 239-254.

Koutsoyiannis, D. (2011), Hurst-Kolmogorov dynamics and uncertainty, J. Am. Water Resour. Assoc., 47, 481-495.

Montanari, A., and D. Koutsoyiannis (2012), A blueprint for process-based modeling of uncertain hydrological systems, Water Resour. Res., 48, W09555, doi:10.1029/2011WR011412.

Plate, E.J. (2002) "Flood risk and flood management." Journal of Hydrology 267.1: 2-11.

Steinschneider, S., A. Polebitski, C. Brown, and B. H. Letcher (2012), Toward a statistical framework to quantify the uncertainties of hydrologic response under climate change, Water Resour. Res., 48, W11525, doi:10.1029/2011WR011318.

<u>Response to Referee 2:</u>
In summary I found this an interesting paper and a useful contribution to the incorporation of non-stationarity in water infrastructure planning. As noted in the comments below, I think a restructuring of the paper is needed to improve the contribution and make it more useful to water planners and managers.

Response: We thank the reviewer for acknowledging the contribution of our work and we appreciate their careful review. We have provided detailed responses to every comment below.

Most importantly, the thrust of the paper should be the demonstration of a method – with essentially no calibration or quantitative validation of results for the Truckee river, this case study is not useful in itself.

Response: We agree with the reviewer that one of the primary purposes of this manuscript is to demonstrate non-stationary flood frequency analysis and to make these tools more accessible to the hydrologic community. However, we respectfully disagree with the assertion that this study is not useful in itself. While it is true that we are working in a data limited basin, we would argue that this is true of many if not most basins across the globe and it does not change the fact that water managers need to make decisions. Despite data limitations, we still made every effort to both calibrate and validate our model. As noted in the comments below we have expanded the manuscript to make these aspects of our work more quantitative and transparent.

1) The title "Climate change and non-stationary flood risk for the Upper Truckee River Basin" does not reflect the contribution of the paper, which is really a demonstration of a methodology. This is noted in p. 5082, line 18: "This paper provides an end-to end demonstration of nonstationary GEV analysis coupled with contemporary down-scaled climate projections (specifically, downscaled climate projections from the Coupled Model Intercomparison Project Phase-5 (CMIP-5)), to quantify how the risk profile of existing infrastructure, designed on the basis of a specified flood event, evolves with time over its design life." As noted in my comments below, the Truckee seems to be more of a demonstration data set. This should not be interpreted as a paper providing significant planning information for managers of the Truckee system.

Response: As noted above, we are of the opinion that this paper has two primary contributions. The intent of this work is to both demonstrate the method and provide site specific results for the Truckee basin. To make this point more clearly we have revised the sentence in question to read:

> ***"While the methodology used for this analysis is previously established, this paper provides the first end-to-end demonstration of non-stationary GEV analysis coupled with contemporary downscaled climate projections (specifically, downscaled climate projections from the Coupled Model Intercomparison Project Phase-5 (CMIP-5)), to quantify how the flood risk profiles may evolve in the Truckee river basin over the next century. The intent of this work is 1) to investigate potential flood risk changes over time in the Truckee basin and 2) to demonstrate the applicability of non-stationary techniques in a regional flood analysis to make these tools more accessible to the hydrologic community." (Revised Manuscript, Page 6, Lines 20-28)***

Also we would like to note that this study was commissioned by water managers and our results have been communicated directly with decision makers. While we acknowledge that there is limited data, we respectfully disagree that this work does not provide planning information. We hope the changes

we have made in response to the Reviewer comments will help make this point more clearly. Also to clarify this point we have added the following text to the introduction:

> ***"The flood analysis presented here is part of a larger study on climate change impacts in the Truckee River basin (Reclamation, 2010). This project is supported by local water managers and conducted by the Bureau of Reclamation through the Water Smart Basin Studies Program authorized under U.S. Public Law 111-11, Subtitle F (SECURE Water Act)." (Revised Manuscript, Page 6, Lines 26-30)***

2) p. 5084, line 15, it states "we simulate unregulated flows from 1950 to 1999 using the Variable Infiltration Capacity (VIC) model and validate results using the available unregulated flow estimates." There does not appear to have been any calibration done as part of this effort. Is the validation done on an uncalibrated model? Some basic hydrology validation statistics would be helpful (NS, RMSE, ...) is assessing the streamflow simulation. The qualitative interpretation like "in close agreement..." (p. 5092, l 9) and "in good agreement" (p. 5092, l 14 and l 19) needs to be quantitative. That would provide support for the claim "This demonstrates that the model behavior is a reasonable match to the natural system."

Response: The VIC model we used for this analysis was developed and calibrated as part of the Bureau of Reclamation West Wide Risk assessment (Gangopadhyay et al., 2011). We agree with the reviewer that calibration was not adequately discussed in the original manuscript and we have added the following text to the revised version:

> ***"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 primarily 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 et al. [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 10-19)***

Also we added the following clarification to section 4.1:

> ***"A suite of models were fit to the logarithms of block (cool season, November-April) maxima flows (simulated by the calibrated VIC model)" (Revised Manuscript, Page 14, Lines 11-12)***

We agree with the reviewer that quantitative validation statistics would be ideal. However, given our small sample size (six floods) we do not have enough points to calculate metrics which would be statistically significant. We have added additional specifics throughout this section in the following places:

- ***"generally good agreement" (p. 5091 l17) Replaced with "*****the maximum percent difference between the natural logarithm of simulated and observed flows is 12%."** ***(Revised Manuscript, Page 16, Lines 13-14)***
- ***"matches very closely" (p. 5091) added "(percent difference in the natural logarithm of flows are 0.5% and 1.2% respectively)" (Revised Manuscript, Page 16, Lines 19-20)***

- ***"in close agreement..." (p. 5092, l 9) added "(the difference between the natural logarithm of simulated and observed flows is the smallest of any event at 0.5%)" (Revised Manuscript, Page 17, Lines 2-3)***
- ***"in good agreement" (p. 5092, l 14) replaced with "the VIC simulated flow falls within the interquartile range of the GEV model" (Revised Manuscript, Page 17, Lines 8-9)***
- ***"we note good agreement" (p. 5092, l 19) we rewrote this paragraph to read: "In general, Figures 3 and 4 show that the VIC simulated flows match closely with the observed floods (based on percent difference in the natural logarithm of flows) and that the interquartile range of the GEV distributions encompass the observed and simulated flows in most instances. Figure 3 does illustrate some of the complications in matching individual events, however based on analysis of the driving forces behind each individual event we are able to document the sources of these discrepancies. Based on this analysis we conclude that the model behaviour is a reasonable match with the natural system." (Revised Manuscript, Page 17, Lines 20-26)***
- ***"This demonstrates that the model behavior is a reasonable match to the natural system." See the revision above.***

3) p. 5085, line 20, 234 projections are analyzed, which lumps together extremely aggressive mitigation futures (like RCP 2.6) and more business as usual scenarios (RCP 8.5). It would seem that, for planning purposes, these should be separated. Only one pathway into the future will actually be experienced, and the variability among GCM projections should reflect that. It would make more sense to present each RCP separately, as this allows a consideration of the variation due to following different pathways from the variation in how the atmosphere might respond to the changed atmospheric conditions. These are very different sources of variability.

Response: This is an interesting question and we agree with the reviewer that between the GCM projections there are multiple sources of uncertainty. In our original analysis we did compare results based on emissions scenarios. The figure below plots the risk of a 1 day flow exceeding the design flood of 37,600 cfs in 10 years for three time periods with results grouped by the emissions pathway. As you can see here in all cases the differences between emissions pathways are smaller than the changes over time. Based on this result we decided not to group our results by emissions pathways in the final paper. Still we agree that this is an important point and we discussed it on page 5096 lines 5-11:

> ***"Although the figures are not shown here, results were also grouped by RCPs to analyze connections between greenhouse gas emission rates and changes in flood risk. We observed no clear trend in flood risk based on the different RCPs. This indicates that the variability between GCM model form and initial conditions likely overwhelms the influence of greenhouse gas emissions when comparing between scenarios. In other words, the variability between projections within any RCP scenario is larger than the difference between RCP scenarios."***

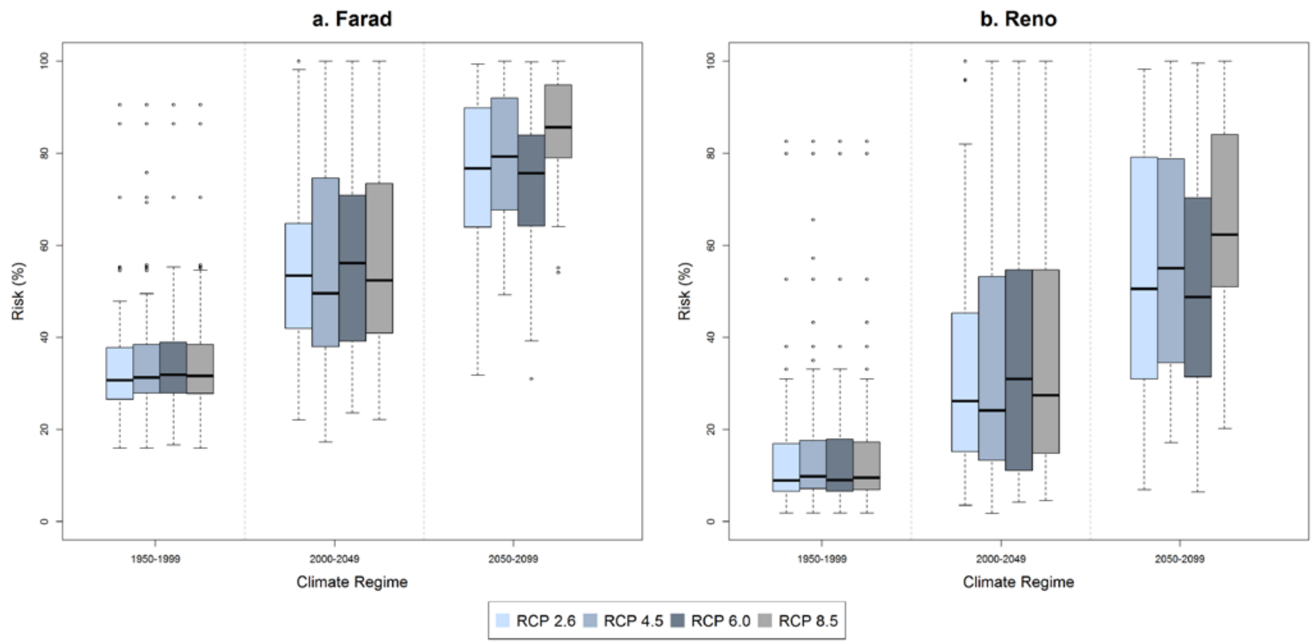


Figure 1: Risk of 1 day flood exceeding 37,600 cfs in a ten year period grouped by emissions pathways.

A second point is that this includes multiple contributions from some GCMs and single contributions from other GCMs. Plenty of research demonstrates that different runs of a single GCM are less independent than runs of different GCMs, and lumping them all together inappropriately weights models that happen to have submitted many runs as part of CMIP5. To demonstrate the method for this paper, there is no need to use 234 projections – a more carefully selected set of a dozen or two would seem to suffice, and also provide a better demonstration of appropriate use of climate model output.

Response: While we understand the reviewers point, we would like to note that there has been much scientific debate on this topic in recent years. As of yet, there is no consensus on the "best practice" for selecting climate projections and likely there may not be one given the fundamental non-linear dynamics of the earth system. The IPCC presents every GCM scenario as equally likely and there is significant debate about whether it is appropriate to select projections based on historical simulation skill because this approach is based on the assumption that historical simulation skill is correlated to realistic simulation of sensitivity to increased greenhouse gasses. As noted in the West Wide Climate Risk Assessment (Reclamation, 2011):

> ***"To date, there is still limited evidence to support such a philosophical bridge (Reichler and Kim 2008; Santer et al. 2009; Pierce et al. 2009). It also has been shown that when such skill assessments are based on many climate metrics (e.g., Tebaldi et al. 2005; Mote and Salathé 2010), the clarity of "better" versus "worse" climate models is less obvious than when the assessment is based on few metrics (Brekke et al. 2008; Reichler and Kim 2008; Gleckler et al. 2008). Even when the historical skill assessment results have been used to rank and cull climate models, thereby conditioning the assessments of future climate uncertainty (Brekke et al. 2008) or detection and attribution of causes for trends in historical atmospheric water vapor over large spatial scales (Santer et al. 2009), the effect of model culling on***

> ***assessments has been minor. These latter results suggest that other factors, beyond historical skill, are driving impact assessments from projected climate conditions within an ensemble, including emissions pathway and a GCM's —natural variability."***

Therefore, we decided to use all 234 projections because we do not have a rational way to "carefully select" a smaller subset of projections. Again, there is presently no one objective way to subset projections.

4) Section 2.3, last paragraph, it is mentioned that the Bureau of Reclamation has developed an archive of downscaled data, but then the downscaling is described as if it were done again for this effort. Were the projections obtained from a published archive? If so, state that, provide an appropriate citation and acknowledgement (I see on the BoR site there is a standard citation and acknowledgement).

Response: The BCSD CMIP-5 projections are currently available on the BoR archive. However, at the time of this analysis the complete set of BCSD projections was not developed yet and so we extended the CMIP-5 analysis for our domain as part of this project. Still, we agree with the reviewer that it would be helpful to point readers to the most recent references for this data source. Therefore we have added the following clarification to the revised manuscript:

> ***"For this analysis we extended the existing hydrology archive to cover the UTRB domain for all 234 BCSD CMIP-5 climate projections following the steps detailed below. A subset of the CMIP-5 hydrology projections is publically accessible through the "Downscaled CMIP3 and CMIP5 Climate and Hydrology Projections" archive at http://gdo-dcp.ucllnl.org/downscaled_cmip_projections/. Additional documentation on the archive and the methodology is provided in Reclamation [2014]" (Revised Manuscript, Page 10, Lines 6-11)***

5) p. 5092, the lack of qualitative model validation appears again here, such as "the VIC simulated and observed floods are in close agreement and the discrepancy with the GEV model is explained by the flood timing described above." What constitutes 'close agreement' and at what point would they be considered not in agreement? And the discrepancy is not explained by the timing, but is apparently consistent with it, which is much weaker. Later instances in this section show things like "the GEV model is in good agreement with the VIC simulated flow", and ultimately "This demonstrates that the model behavior is a reasonable match to the natural system." These general observations are not helpful in determining significant correspondence of modeled and simulated values.

Response: We appreciate the comment and have made changes throughout section 4.2. to make our analysis more precise (please refer to the changes listed in comment 2). Also, in response to this comment we have replaced the phrase 'is explained by' with 'is consistent with'. We hope that the review will find our revised section 4.2 to be more clear.

6) p. 5094, Figure 6 is presented, which is interesting. another way to cast this would be in a manner similar to that of Mailhot and Duchesne (J. Wat. Res. Plann. Mgmt., 2010, doi 10.1061/_ASCE_WR.1943-5452.0000023) Figure 3, which aims to provide planners with a design return period for today that would be needed to provide protection at the level of a historic return period (in a stationary climate).

Response: We thank the reviewer for pointing out this paper. We focused our analysis on the risk to existing infrastructure. Hence we present our results in terms of the probability of exceeding the flow rate that was used in previous design. However we agree that Mailhot and Duchesne provide an interesting approach for taking non-stationarity into account with new design. We have added a reference to their work in the beginning of the Introduction as follows:

> ***"This discrepancy has not gone unnoticed within the scientific community and there is a growing body of research investigating, (1) trends in observed floods [e.g. Franks, 2002; Vogel et al., 2011], (2) ways to incorporate non-stationarity into frequency distributions [e.g. Katz and Neveau, 2002; Raff et al., 2005] and (3) methodologies to interpret risk and approach design within a non-stationary framework [e.g. Mailhot and Duchesne, 2010; Rootzen and Katz, 2013; Salas and Obeysekara, 2014]." (Revised Manuscript, Page 3, Lines 3-8)***

Minor comments:

SI units should be used throughout, not square miles, feet , etc.

Response: We have changed all of the units to SI.

p. 5088, line 16, is the 0.05 alpha? And what significance test is being referred to here?

Response: Yes the threshold is referring to the alpha. For the significance test we are using a chi-squared distribution. These points have been clarified in the revised manuscript as follows:

> ***"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 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)***

References cited in Response:

Brekke, L.D., M.D. Dettinger, E.P. Maurer, and M. Anderson. 2008. "Significance of model credibility in estimating climate projection distributions for regional hydroclimatological risk assessments," *Climatic Change*, 89(3–4), 371–394.

Gleckler, P.J., K.E. Taylor, and C. Doutriaux. 2008. "Performance metrics for climate models," *Journal of Geophysical Research* 113(D06104).

Mote, P.W., and E.P. Salathé. 2010. "Future climate in the Pacific Northwest," *Climatic Change*, DOI: 10.1007/s10584-010-9848.

Pierce, D.W., T.P. Barnett, B.D. Santer, and P.J. Gleckler. 2009. "Selecting global climate models for regional climate change studies," *Proc National Academy Sciences*, v106 (21), 8441–8446.

Reclamation. 2011. “West-Wide Climate Risk Assessments: Bias-Corrected and Spatially Downscaled Surface Water Projections,” prepared by Bureau of Reclamation. U.S. Department of the Interior, Draft Technical Memorandum 86-68210-2011-01, March 2011, 138 pp.

Reichler, T., and J. Kim. 2008. “How Well Do Coupled Models Simulate Today’s Climate?” *Bulletin of the American Meteorological Society*, 89(3) pp. 303–311.

Santer, B.D., K.E. Taylor, P.J. Gleckler, C. Bonfils, T.P. Barnett, D.W. Pierce, T.M.L. Wigley, C. Mears, F.J. Wentz, W. Bruggemann, N.P. Gillett, S.A. Klein, S. Solomon, P.A. Stott, and M.F. Wehner. 2009. Incorporating model quality information in climate change detection and attribution studies, *Proc National Academy Sciences*, v 106 (35), 14778–14783.

Tebaldi, C., R.L. Smith, D. Nychka, and L.O. Mearns. 2005. —Quantifying Uncertainty in Projections of Regional Climate Change: A Bayesian Approach to the Analysis of Multi-model Ensembles, *Journal of Climate*, 18 pp. 1524–1540.

<u>Response to John England:</u>

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).

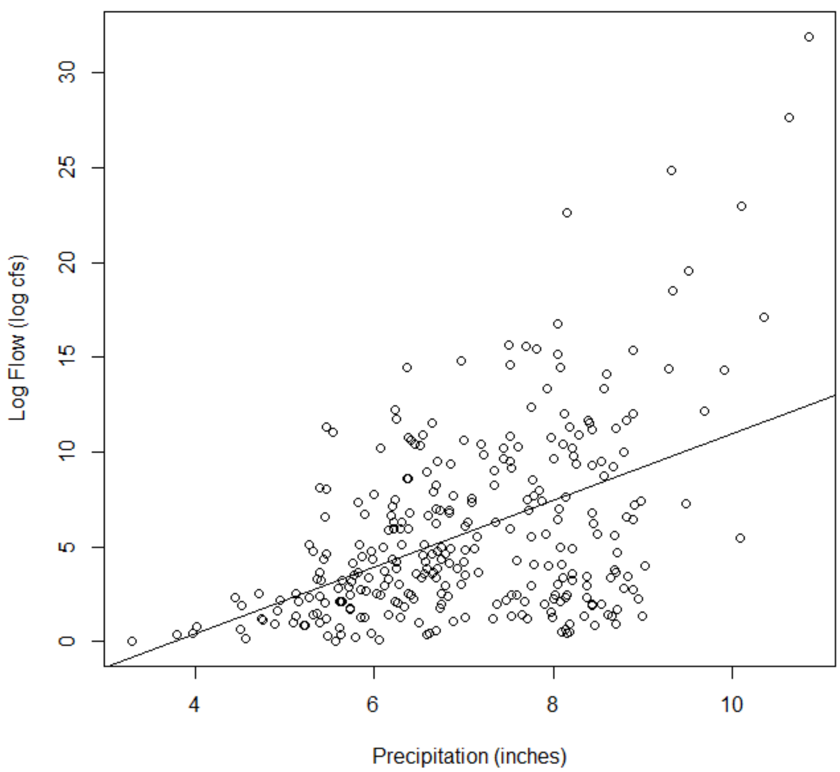
With all the information behind Table 2 on covariates (page 5105), and presumed functional form (equations 2 and 3), the authors need to clearly demonstrate the evolution of aggregated winter monthly precipitation and monthly average temperature trends with time. Are trends in location and scale needed? 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 Obeysekara (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 18-25)***

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 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 loses)." (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. Still, we agree that this is an important topic and have added uncertainty bounds to our parameter estimations along with the following discussion:

> ***"To address uncertainty, models of the same form (i.e. non-stationary location and scale with precipitation and temperature as covariates) were also fit to the historical simulation period (1950-1999) using downscaled precipitation and temperature from all 234 climate projections. Because each climate projection seeks to reproduce historical behaviour over the historical period, the variability between projections in this time frame is a measure of uncertainty in model coefficients given the same physical system. This differs from the variability between climate projections in future periods (i.e. after 1999) which is a measure of uncertainty in future forcing conditions. Table 3 shows the interquartile range of model coefficients calculated from the 234 historical GCM simulations.***
>
> ***Using these parameters the return period of the design flood at Reno (37,600cfs) was calculated for every set of model parameters using observed historical precipitation and temperature. The observed model estimates a return period of 45 years while the interquartile range (IQR) using the simulated model parameters (i.e., the model parameters estimated from each of the 234 historical GCMs) with observed precipitation and temperature varies from 28 to 247 years. Note that the return period of 45-years estimated from observed meteorology is within the IQR of 28 to 247 years. Although the IQR is large it should be kept in mind that some of the uncertainty in this range is a result of the downscaling methodology. The monthly BCSD algorithm used for downscaling GCM climate only constrains the monthly precipitation and temperature statistics (total precipitation and mean monthly temperature) over the historical 1950-1999 period. Furthermore, uncertainty is introduced when monthly total precipitation and mean temperature are translated to daily values. Thus the estimated IQR implicitly captures downscaling uncertainties, in addition to explicitly representing parameter uncertainty. The need to consider uncertainties at each and every step of the process starting with, for example, downscaling methods (statistical, dynamical or some combination of statistical and dynamical methods) is a topic of ongoing research." (Revised Manuscript, Page 15 lines 7-31)***

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 Hydrocliamtic 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 10-19)***

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.

<u>Some Specific Comments</u>
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 loses)." (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:

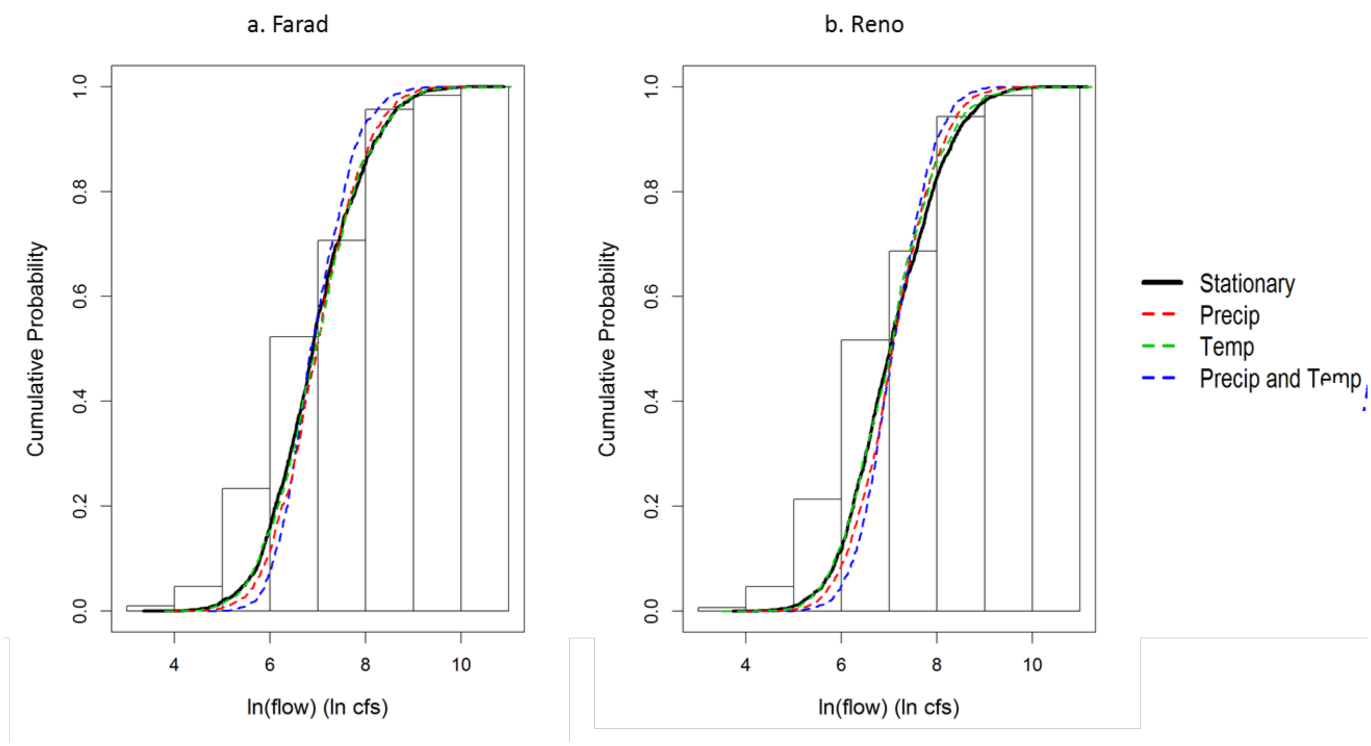
> ***"However, as noted by Towler et al. [2010], when multiple values are used per year the calculated probabilities must be adjusted appropriately to derive annual values." (Revised Manuscript, Page 13, Lines 6-8)***

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 2-5)***

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 "ungagged" 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 precipitaiton 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 Line 28- Page 19 Line 8)***

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

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.

Climate change and non-stationary flood risk for the Upper Truckee River Basin

L. E. Condon[1,2], S. Gangopadhyay[1] and T. Pruitt[1]

[1]{Bureau of Reclamation Technical Service Center, Denver, Colorado}

[2]{Hydrologic Science and Engineering Program and Department of Geology and Geological Engineering, Colorado School of Mines, Golden, Colorado}

Correspondence to: L. E. Condon (lcondon@mymail.mines.edu)

Abstract

Future flood frequency for the Upper Truckee River Basin (UTRB) is assessed using non-stationary extreme value models and design life risk methodology. Historical floods are simulated at two UTRB gauge locations, Farad and Reno using the Variable Infiltration Capacity (VIC) model and non-stationary Generalized Extreme Value (GEV) models. The non-stationary GEV models are fit to the cool season (November-April) monthly maximum flows using historical monthly precipitation totals and average temperature. Future cool season flood distributions are subsequently calculated using downscaled projections of precipitation and temperature from the Coupled Model Intercomparison Project Phase-5 (CMIP-5) archive. The resulting exceedance probabilities are combined ~~into a single risk metric~~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. This paper provides the first end-to-end analysis using non-stationary GEV methods coupled with contemporary downscaled climate projections to demonstrate ~~how the~~the evolution of flood risk profile ~~of existing infrastructure evolves with time~~ over ~~its typical~~ design life periods of existing infrastructure that is vulnerable to flooding (e.g. dams, levees, bridges, and sewers). Results show that flood risk increases significantly over the analysis period (from 1950 through 2099). This highlights the potential to underestimate flood risk using traditional methodologies that don't account for time varying risk. Although model parameters, for the non-stationary method are sensitive to small changes in input parameters, analysis shows that the changes in risk over time are robust. Overall, flood risk at both locations (Farad and Reno) is projected to increase 10-20% between the historical period1950-1999 and the future period 2000-2050 and 30-50% between the same historical period and 2050-2099.

1 Introduction

"Stationarity is Dead" [Milly et al., 2008], yet the standard practice for flood frequency analysis is predicated on this very assumption. This discrepancy has not gone unnoticed within the scientific community and there is a growing body of research investigating, (1) trends in observed floods [e.g. Franks, 2002; Vogel et al., 2011], (2) ways to incorporate non-stationarity into frequency distributions [e.g. Katz and Neveau, 2002; Raff et al., 2005] and (3) methodologies to interpret risk and approach design within a non-stationary framework [e.g. Mailhot and Duchesne, 2010; Rootzen and Katz, 2013; Salas and Obeysekara, 2014]. Both the frequency and intensity of extreme events are particularly susceptible to change because small shifts in the center of a distribution can potentially have much larger impacts on the tails [Meehl et al., 2000]. Regardless of climate change, naturally occurring long-term climate oscillations, such as ENSO, have been linked to low frequency variability in flood frequency [e.g. Cayan et al., 1999; Jain and Lall, 2001]. Anthropogenic climate change has the potential to amplify natural climatic variability throughout the ~~this variability by introducing additional non-stationarity to~~ interconnected climate and hydrologic systems.

Already trends in many hydrologic variables have been observed across the Western United states (as well as around the world). For example, clear increases in temperature have been measured across the west [e.g. Cayan et al., 2001; Dettinger and Cayan, 1995]. Precipitation trends are more variable. Regonda et al. [2005] found increased total winter precipitation (rain and snow) from 1950 to 1999 in many sites across the western United States, although springtime snow water equivalent (SWE) was shown to decline over the same period. Similarly, Mote et al. [2005] analyzed snowpack trends in western North America, and reported widespread declines in springtime SWE ~~in much of the North American West~~ over the period 1925–2000, especially since the middle of the 20th century. They attribute this decline predominantly to climatic factors such as El Niño–Southern Oscillation (ENSO), Pacific Decadal Oscillation (PDO), and positive trends in regional temperature. Easterling et al. [2000] summarized previous studies on precipitation trends. They note that trends vary from region to region, but in general, increases in precipitation have occurred disproportionately in the extremes. Several subsequent studies have observed increasing trends in extreme precipitation events, although the changes are relatively small [Gutowski et al., 2008; Kunkel, 2003; Madsen and Figdor, 2007].

~~Similarly, research~~Research has also demonstrated increasing trends in flood frequency in some regions. Walter and Vogel [2010] and Vogel et al. [2011] observed increasing flood magnitudes across the United States using stream gauge records, and Franks [2002] showed statistically significant increases in flood frequency since the 1940s. Still, non-stationary flood behaviour ~~flooding trends have~~has been historically difficult to quantify and there has been some debate on the significance of flood frequency trends. For example, Hirsch [2011] noted both increasing and decreasing trends in annual flood magnitudes in different regions of the US. Also, Douglas et al. [2000] found that if one takes into account spatial correlation, many previous findings of flood trends are not statistically significant. Difficulty in diagnosing flood trends is not unique to the Western US; ~~–~~a literature review of historical flood studies across Europe also found spatial variability in flood trends [Hall et al., 2014].

Even when significant trends are found, the complexity of flooding mechanisms that depend on many variables and can vary regionally and seasonally makes it difficult to attribute trends to specific causes. ~~-~~Illustrating the importance of seasonality, Small et al. [2006] showed that if a high precipitation event occurs in the fall, as opposed to the spring, it will contribute to baseflow rather than inducing flooding. Also, urbanization can drastically increase the impervious area of a basin, thus amplifying floods by decreasing infiltration and speeding runoff. The largest flood magnitude increases observed by both Walter and Vogel [2010] and Vogel et al. [2011] were in basins with urban development. The influence of development trends on flood behavior can be difficult to separate from other variables. For example, Villarini et al. [2009] could not conclusively tie reduced stationarity (i.e. changes in mean and/or variance) in peak discharge records to climate change because of variability in the other factors that influence runoff.

Merz et al. [2012] note that attributing changes in flood hazard is complicated by the complex array of drivers that can include; land cover change and infrastructure development as well as natural climate variability and change. Here we ~~S~~set~~ting~~ aside the impacts of development and management practices and focus on the role of climate change. ~~,~~ However, even with this simplification, future extremes can still be influenced by a number of interrelated variables such as changes in temperature, precipitation efficiency, and vertical wind velocity [Mullet et al., 2011; O'Gorman and Schneider, 2009]. Analyzing global circulation model (GCM) outputs Pierce et al. [2012] found total changes in precipitation to be small relative to the existing variability but noted larger seasonal changes in storm intensity and frequency.

Despite uncertainty, many studies agree that warming will increase the potential for intense rainfall [Allan, 2011; Gutowski et al., 2008; Pall et al., 2011; Sun et al., 2007]. Furthermore, Min et al. [2011] found that some GCM simulations may underestimate extreme precipitation events in the northern hemisphere. Indicating that projections of extreme precipitation based on GCM outputs may be conservative.

Studies have also predicted increases in flood frequency and magnitude with a warmer climate especially in snowmelt dominated basins [e.g. Das et al., 2011]. As with historical flooding trends, translating forecasted climate variables to flood frequency is a complex process and several methodologies have been used. Downscaled GCM climate forcings can be used to drive hydrologic models and simulate future floods directly [e.g. Das et al., 2011; Vogel et al., 2011; Raff et al., 2009]. With this approach traditional stationary flood frequency distributions can be fit to the simulated floods to calculated return periods of interest [e.g. Raff et al., 2009; Vogel et al., 2011]. This allows for return periods and flood magnitudes that change over time, as with the flood magnification and recurrence reduction factors calculated by Walter and Vogel [2010] and Vogel et al. [2011]. However, these approaches still assume flood mechanisms are stationary over the time period that the distribution is fit to.

This limitation can be overcome using non-stationary generalized extreme value (GEV) distributions where the model parameters like mean (i.e. location) and spread (i.e. scale) are allowed to vary as a function of time [e.g. Gilroy and McCuen, 2012] or with relevant covariates [e.g. Griffis and Stedinger, 2007; Richard W Katz et al., 2002; Towler et al., 2010]. This approach has been gaining popularity for flood frequency estimation. Using this technique it is not necessary to simulate future floods directly by forcing a hydrologic model with projected hydroclimate fields (e.g. precipitation and temperature). The parameters of the GEV model, like mean and spread change with time (i.e. non-stationary) based on a linear combination of covariates like precipitation and temperature. Historical relationships between extreme events and hydroclimate fields are used to identify the weighting of covariates. These weights are then used to estimate parameters for future time periods using precipitation and temperature outputs from hydroclimate projections. For example, Gilroy and McCuen [2012] used non-stationary GEV models of flood frequency that incorporated a linear trend in the location parameter. Similarly, Griffis and Stedinger [2007] and Towler et al. [2010] used climate variables as covariates for the distribution parameters.

While, non-stationary flood forecasting methods provide flexibility to analyze flood variability, they are also incongruent with many of the traditional ~~risk~~ metrics used in water resources planning. 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). However, these calculations rely on a flood distribution which is assumed to remain stationary with time, and hence the return period design metric is also assumed to be stationary. When non-stationary methods are used, the underlying flood distributions, and associated return periods, vary with time. Thus, under a non-stationary climate, the notion of static return period flood event (e.g., 100-year flood, 200-year flood, etc.) is no longer a valid concept.

To address this issue, Rootzén and Katz [2013] introduced the concept of design life level to calculate the risk of a given flood magnitude occurring over a specified time period. Salas and Obeysekera [2014] further demonstrated the relevance of this technique to hydrologic community using flood frequency examples. However, this methodology ~~still~~ has yet to receive widespread attention within the hydrologic community. Here, we present a non-stationary flood frequency assessment for the Upper Truckee River Basin (UTRB) using contemporary downscaled climate projections and the non-stationary design life level technique introduced by Rootzén and Katz [2013] to quantify flood risk~~.~~ (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). While the methodology used for this analysis is previously established, ~~T~~this paper provides ~~an~~ the first end-to-end demonstration of non-stationary GEV analysis coupled with contemporary downscaled climate projections (specifically, downscaled climate projections from the Coupled Model Intercomparison Project Phase-5 (CMIP-5)), to quantify how the ~~risk profile of existing infrastructure, designed on the basis of a specified flood event,~~flood risk profiles may evolve in the Truckee river basin over the next century. The flood analysis presented here is part of a larger study on climate change impacts in the Truckee River basin (Reclamation, 2010). This project is supported by local water managers and conducted by the Bureau of Reclamation through the Water Smart Basin Studies Program authorized under U.S. Public Law 111-11, Subtitle F (SECURE Water Act). The intent of this work is 1) to investigate potential flood risk changes over time in the Truckee basin and 2) to demonstrate the applicability of non-stationary techniques in a regional flood analysis to make these tools more accessible to the hydrologic community. ~~s with time over its design life.~~

The paper is organized as follows. Section 2 provides background on the study area along with data sets and models used. The methodologies of using non-stationary spatial GEV analysis in conjunction with climate projections and time evolving risk assessment are described in section 3. Results and discussions of findings are given in section 4. Summary and conclusions from the analysis are presented in section 5.

2 Background

This section provides background on the study area (2.1), streamflow data and simulations ~~and validation~~ (2.2) and climate data and models (2.3).

2.1 Upper Truckee River Basin

The Truckee River originates in northern Sierra Nevada Mountains in California (above Lake Tahoe) and flows northeast to Nevada where it ends in the Pyramid Lake (Figure 1). The total basin area is roughly ~~3,060~~ 7,900 square ~~miles~~ kilometres, however the area upstream of Reno (~~1,067~~2,763 square ~~miles~~kilometres) provides the majority of the basin's precipitation through snowpack. The focus of this analysis is on the Farad and Reno gauge locations shown in Figure 1, henceforth referred to as Farad and Reno. The Farad gauge is located roughly ~~one mile~~1.5 kilometers downstream of the Farad hydropower plant and provides a cumulative measure of all of the upper basin tributaries [Stokes, 2002]. 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).

Flooding in the upper Truckee generally takes one of three forms. Some of the most severe floods have resulted from heavy rain events covering most of the basin and lasting one to six days. These storms generally occur from November to April and may be linked to Atmospheric Rivers [Ralph and Dettinger, 2012]. Snowmelt floods are also common from April to July. Although, snowmelt floods transmit large volumes of water for longer durations, they generally don't cause damage because they are well predicted and can be regulated with upstream reservoirs. Finally, in late summer (July – August) local cloudbursts can generate high intensity precipitation over small areas. These storms can cause local damage to tributaries but generally don't have a large impact on the main stem of the Truckee.

In the twentieth century, nine major floods have been recorded on the Truckee River, all of which occurred from November to April [USACE, 2013b]. The flood of record occurred in January of 1997 and was caused by warm rain falling on a large snowpack (~180% of normal) and melting nearly all of the snowpack below 7,000 feet [USACE, 2013b]. The floods of 1950, 1955 and 1963 were some of the most damaging due to the development of Reno along the river during this time period [USACE, 2013b]. Subsequent flood damages have been, at least partially, mitigated by the implementation of flood infrastructure starting in the 1960s.

2.2. Streamflow data and simulations~~data~~

Streamflow has been measured at both the Farad and Reno USGS gauges. However, gauge flows are not readily applicable to flood frequency analysis due to the presence and development of water supply and flood control structures upstream. For example, upstream of Reno there are four dams with flood control capabilities (i.e. Martis Creek Dam, Prosser Creek Dam, Stampede Dam and Boca Dam) in addition to Tahoe, Donner and Independence Lakes which provide incidental flood regulation. Unregulated flow estimates were developed by the US Army Corps of Engineers (USACE) but are only available for historical flood periods [USACE, 2013b]. Therefore, we simulate unregulated flows from 1950 to 1999 using the Variable Infiltration Capacity (VIC) model and validate results using the available unregulated flow estimates.

A brief summary of the VIC model is provided here, and for additional technical specifications the reader is referred to Liang et al. [1994], Liang et al. [1996] and Nijssen et al. [1997]. VIC is a gridded hydrologic model designed to simulate macro scale (spatial resolution in greater than 1mile) water balances using parameterized sub-grid infiltration and vegetation processes. In the VIC model, surface water infiltrates to the subsurface based on conductivity, and soil moisture is distributed vertically through three model layers extending up to about 2 meters below the land surface. At the surface, potential evapotranspiration (PET) is simulated using the Penman Monteith PET model [Maidment et al., 1993]. Surface flows are determined in a two-step process. First, the water balance for each grid cell is calculated independently to determine surface runoff and baseflow, and subsequently runoff from each cell is routed to river channels and outlets using a predefined routing network. Here we drive VIC with daily weather forcings including precipitation, maximum and minimum temperature and wind speed. Additional climate variables such as short and long wave radiation, relative humidity and vapor pressure are calculated within the model. The VIC model is well documented and has already been used in a number of hydrologic and climate change studies [e.g. Christensen and Lettenmaier, 2007; Christensen et al., 2004; Gangopadhyay et al., 2011; Maurer et al., 2007; Payne et al., 2004; Reclamation, 2011; Van Rheenen et al., 2004]. Recently VIC has also been applied for real time flood estimation [Wu et al., in press].

The VIC model used for this analysis was developed and calibrated as part of the Bureau of Reclamation's (Reclamation) West Wide Climate Risk Assessment (WWCRA). The WWCRA VIC model encompasses the western US. Streamflows were evaluated at 152 locations primarily 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 et al. [2011]. While we do not discuss model calibration further here, in the subsequent sections ~~W~~we provide additional model verification for flood simulation in the UTRB.~~ in subsequent sections.~~

2.3 Climate data and models

As noted in the previous section, the VIC model requires daily climate inputs to drive water balance simulations. We use the national 1/8° (roughly 7 miles) gridded dataset from Maurer et al. [2002] for historical (i.e. 1950-1999) climate observations. Additionally, monthly total precipitation and average temperature were aggregated for the upstream area of each gauge for every month of the flood season (i.e. November through April). These values are used as covariates for fitting non-stationary GEV models as discussed in Section 3.

Future gridded precipitation and temperature values from 2000 to 2099 were generated from Global Circulation Model (GCM) outputs. We analyzed 234 projections generated by 37 different climate models from the CMIP-5 (Coupled Model Intercomparison Project Phase 5) archive [Taylor et al., 2012]. Projections span four Representative Concentration Pathways (RCPs) for greenhouse gas emissions. Each GCM projection includes monthly gridded precipitation and temperature from 1950 to 2099 at a coarse grid resolution ranging between ~~~40~~65-~~160~~ 250 ~~miles~~kilometers.

~~The Bureau of~~ Reclamation in collaboration with other federal and non-federal partners has developed a monthly archive of downscaled CMIP-5 projections at the finer 1/8th degree resolution using the two-step BCSD (Bias Correction and Spatial Disaggregation) algorithm described in Wood et al. [2004]. For this analysis we extended the existing hydrology archive to cover the UTRB domain for all 234 BCSD CMIP-5 climate projections following the steps detailed below. A subset of the CMIP-5 hydrology projections is publically accessible through the "Downscaled CMIP3 and CMIP5 Climate and Hydrology Projections" archive at http://gdo-dcp.ucllnl.org/downscaled_cmip_projections/. Additional documentation on the archive and the methodology is provided in Reclamation [2014].

The downscaled climate variables include monthly total precipitation, monthly maximum and minimum temperatures and monthly average temperature. Before applying the BCSD algorithm all the 234 climate projections were first gridded from their respective native GCM scale to a common grid of 1° latitude by 1° longitude. Similarly, the observed 1/8° degree gridded dataset [Maurer et al., 2002] was aggregated to the coarser 1° latitude by 1° longitude grid. Next, for a given climate variable, GCM, and location (1° latitude by 1° longitude grid cell), the bias correction (BC) step uses quantile mapping between monthly CDFs (Cumulative Distribution Functions) of historical simulated and historical observed values to identify biases over a common climatological period – in this case, 1950-1999. The projected future climate variables from the same GCM at the same location are then bias corrected using the identified bias. The result of bias-correction is an adjusted GCM dataset (20th century and 21st century, linked together) that is statistically consistent with the observed data during the bias-correction overlap period (i.e., 1950-1999 in this application). Note that the BC step happens at the coarse 1° latitude by 1° longitude grid. Next, adjustment factors that are multiplicative (ratio of bias-corrected GCM to observed) for precipitation and an offset (bias-corrected GCM minus observed) for temperature are calculated for each of the 1° latitude by 1° longitude grid cell [Reclamation, 2013]. These adjustments are then spatially disaggregated (SD) to a 1/8° latitude by 1/8° longitude grid. Finally, the adjustments are applied (multiplicative for precipitation; additive for temperature) to the finer resolution, 1/8° degree gridded observed precipitation and temperature fields [Maurer et al., 2002] to derive the 1/8° degree gridded BCSD climate projections.

3 Methodology

This section describes the methodology used for flood frequency analysis in the UTRB. Discussion is divided into two sections. First, we describe the process of extreme value modeling using non-stationary GEV distributions (Section 3.1). Second, the methodology for design life level risk assessment is detailed (Section 3.2)

3.1 Extreme value ~~modeling~~modelling

Extreme values analysis (EVA) deals with the examination of the tail (i.e. extreme) values of a distribution (as opposed to standard approaches which are generally more concerned with the average system behaviour). EVA methods are standard practice for flood frequency analysis because they are designed to capture the behaviour of low frequency high impact events. Furthermore, in climate change studies Katz [2010] points out that traditional approaches are not sufficient and extreme value statistics are needed. For this analysis, we use the Generalized Extreme Value (GEV), which is commonly applied to flood frequency analysis to model block maxima from streamflow time series [e.g. Katz et al., 2002; Towler et al., 2010]. The cumulative distribution function (CDF) for the GEV is as follows:

$$F(z;\theta) = \exp\left\{ -\left[1+\xi\left(\frac{z-\mu}{\sigma}\right)^{-\frac{1}{\xi}}\right] \right\} \qquad (1)$$

Formatted: Right

Where z is the streamflow maxima value of interest and θ is the parameter set (μ, σ, ξ) used to specify the distribution such that the center is given by the location (μ), the spread by the scale(σ) and the behavior of the upper tail by the shape (ξ). Based on the shape parameter, the GEV can take one of three forms: Gumbel, or light tailed, when ξ is zero; Fréchet, or heavy tailed, if ξ is positive; and Weibull, or bounded, when ξ is negative. Following the methodology of Towler et al. [2010], GEV parameters (μ, σ, ξ) are fitted using the Maximum Likelihood Estimation (MLE) technique.

Formatted: Font: Italic

In traditional stationary flood frequency analysis, it is assumed that observations are independent and identically distributed (IID), and therefore model parameters (μ, σ, ξ) are derived from the observed flood record and are assumed to remain constant across the period of record and into the future. Here, we introduce non-stationarity into the distribution by allowing location and scale parameters to change with relevant covariates. Such that:

$$\mu(t) = \beta_{0,\mu} + \beta_{1,\mu}x_1 + \cdots + \beta_{n,\mu}x_n \tag{2}$$

$$\sigma(t) = \beta_{0,\sigma} + \beta_{1,\sigma}x_1 + \cdots + \beta_{n,\sigma}x_n \tag{3}$$

Where the β variables represent the coefficients, and the x variables are the covariates. In keeping with previous studies the shape parameter, which is the most difficult to estimate, is assumed constant [e.g. Obeysekara and Salas, 2014; Salas and Obeysekera, 2013; Towler et al., 2010].

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.

To determine the optimal set of covariates for a non-stationary model, additional statistical methods must be employed. The Akaike Information Criterion [AIC; Akaike, 1974], given in Equation 4, weighs the goodness of fit for a model with the level of complexity.

$$AIC = -2(llh) + 2K \tag{4}$$

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 ~~AIC~~ 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. ~~and~~ 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 ~~to ensure~~ statistically significant (~~threshold~~ alpha of 0.05) ~~differences between scores~~improvement in model performance.

Following the methodology described above, GEV distributions are fit to time series of maximum monthly historical (1950-1999) one day simulated stream flows (detailed in Section 2) for the cool season. Although, there are some unregulated historical flow estimates, the available dataset only covers six storms. Therefore, to be consistent we fit our model only to the simulated flows. The dataset includes maximum daily streamflows for each month in the cool season defined by the block of months November through April, as opposed to the more traditional single value per year. This technique was also used by Towler et al. [2010] who noted that expanding the dataset helps avoid the problems associated with using maximum likelihood estimate on small datasets. 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. Floods during the cool season generally last between one and four days. Here we focus on the one day flood peak, as opposed to multi-day flood volumes, because this is a representative metric for flood damage. Additionally, using the one day flood maximum focuses the analysis on flood magnitude rather than duration.

Two covariates were considered, monthly total precipitation (P) and mean temperature (T) averaged over the upstream area for each gauge. As discussed in Section 2, precipitation is a relevant covariate because many of the floods in this season are rain on snow events or extreme rainfall events. Similarly, temperature drives snowmelt and is an important contributor to UTRB flood events (e.g., January 1997 event). Both stationary and non-stationary GEV models were evaluated using the extRemes package [Gilleland and Katz, 2011] in the 'R' statistical computing environment.

3.2 Time varying risk assessment

Traditional flood planning relies on the concept of return periods, which are usually calculated as the inverse of annual exceedance probability for a given flood magnitude, assuming a stationary distribution. For example, the log-Pearson Type III (LP3) distribution described by the Interagency Advisory Committee on Water Data Bulletin 17B [IACWD, 1982]. However, when non-stationary models are used, the distribution parameters, and hence the exceedance probabilities vary with time. Table 1 compares various flood probability calculations between stationary and non-stationary approaches [Salas and

Obeysekera, 2014]. As shown here, when the flood distribution is stationary, the return period for a given flood magnitude is constant and relies only on the exceedance probability (4a). However, if distribution parameters are non-stationary then the return period will vary based on the period of interest (4b). This concept is easily extended to flood risk (here defined as the probability of a flood of a given magnitude occurring, not expected loses). In traditional analyses, the risk of a flood occurring in a given period depends only on the length of the period (5a), while in a non-stationary analysis risk depends on both the length of time considered and the time period itself (5b). This is the concept of design life level proposed by Rootzén and Katz [2013]. Here, we adopt the design life level risk framework given in (5b) and calculate the risk of flood for a range of future periods and design life lengths.

4 Results and Discussion

Results are grouped into three sections. First we present the development of the non-stationary GEV models (4.1). Next the models are verified by comparing simulated results to observations (4.2). Finally we present future projections of flood frequency analysis (4.3).

4.1 Extreme value model development

A suite of models were fit to the logarithms of block (cool season, November-April) maxima flows (simulated by the calibrated VIC model) with different non-stationary parameter combinations. The model structures tested include stationary, non-stationary location, non-stationary scale and non-stationary location and scale. For all model structures model fit was tested using one or both covariates (i.e. precipitation (in) and temperature (F)). Models were also tested using the block maxima flows directly; however, performance was improved considerably with the logarithmic transformation. Validation of the VIC simulated flows as well as the GEV models are presented in the following section.

Table 2 summarizes negative log-likelihood (NLLH) and Akaike Information Criterion (AIC) scores ~~scores~~ for each model configuration. The deviance statistic (D) ~~and the p-values of~~ for pairwise comparisons of ~~AIC~~ NLLH scores and the p-values calculated for each D based on a chi – squared distribution are also provided (Note that the bottom rows provide~~s~~ the number of parameters in each model and the model number that was used for the pairwise comparisons). As shown here the models with non-stationary location and scale relying on both precipitation and temperature as covariates have the best (i.e. lowest) NLLH scores for both stations, and are a statistically significant improvement over the other models listed in Table 2. Figure 2 plots, stationary and non-stationary location and scale models with histograms of observed flow for both gauges. Qualitatively, the stationary model fits well with the center of the distribution but overestimates the tails. The non-stationary models overestimate the median values but are a closer fit to the extreme values.

The coefficients for equations 2 and 3 for the selected models are provided in Table 3. 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). Thus, for every future month there is a separate GEV curve for each of the 234 climate projections.

To address uncertainty, models of the same form (i.e. non-stationary location and scale with precipitation and temperature as covariates) were also fit to the historical simulation period (1950-1999) using downscaled precipitation and temperature from all 234 climate projections. Because each climate projection seeks to reproduce historical behaviour over the historical period, the variability between projections in this time frame is a measure of uncertainty in model coefficients given the same physical system. This differs from the variability between climate projections in future periods (i.e. after 1999) which is a measure of uncertainty in future forcing conditions. Table 3 shows the interquartile range of model coefficients calculated from the 234 historical GCM simulations.

Using these parameters the return period of the design flood at Reno (37,600cfs) was calculated for every set of model parameters using observed historical precipitation and temperature. The observed model estimates a return period of 45 years while the interquartile range (IQR) using the simulated model parameters (i.e., the model parameters estimated from each of the 234 historical GCMs) with observed precipitation and temperature varies from 28 to 247 years. Note that the return period of 45-years estimated from observed meteorology is within the IQR of 28 to 247 years. Although the IQR is large it should be kept in mind that some of the uncertainty in this range is a result of the downscaling methodology. The monthly BCSD algorithm used for downscaling GCM climate only constrains the monthly precipitation and temperature statistics (total precipitation and mean monthly temperature) over the historical 1950-1999 period. Furthermore, uncertainty is introduced when monthly total precipitation and mean temperature are translated to daily values. Thus the estimated

IQR implicitly captures downscaling uncertainties, in addition to explicitly representing parameter uncertainty. The need to consider uncertainties at each and every step of the process starting with, for example, downscaling methods (statistical, dynamical or some combination of statistical and dynamical methods) is a topic of ongoing research.

4.2 Hydrologic and GEV model validation

Because we utilize modeled VIC flows for flood analysis there are two considerations for model validation. First, we compare VIC simulated one day flood events to the observed ~~ungauged~~ unregulated flow estimates (i.e. validating that our calibrated VIC model is accurately simulating flood flows). Second, we compare the GEV modeled floods to the VIC simulated flows and the observed flow estimates (i.e. validating that the GEV models we fit to the simulated data match both the observed flows and the VIC simulated flows).

Although, ~~ungauged~~ unregulated flows are not available for the entire period of record, one-day maximum ~~ungauged~~ unregulated flow estimates are available at Reno for six historical floods [USACE, 2013b]. Figure 3 plots the observed flow (blue triangle) with the one-day VIC flow that was simulated using historical observed forcings from Maurer et al. [2002] (red triangle), and a boxplot of the non-stationary GEV distribution for the same month generated using the same monthly historical precipitation and temperature [i.e. Maurer et al., 2002]. Comparing first the one day maximum VIC simulated flow with the observed flow ~~there is generally good agreement between the two~~the maximum percent difference between the natural logarithm of simulated and observed flows is 12%. ~~, although t~~There does appear~~s~~ to be a slight positive bias in the VIC simulations (i.e. VIC simulated flows are greater than observed flood flows). Still, the simulated flood values (red circles) generally fall within the interquartile range of the GEV distribution, except in the case of the February 2, 1963 flood and the January 2, 1997 flood.

In these instances the VIC simulation matches very closely (percent difference in the natural logarithm of flows are 0.5% and 1.2% respectively) with the observed flow, however, the GEV model underestimates the events. This discrepancy is caused by the flood timing. In both cases the flood occurs at the very beginning of the month. In the GEV framework the precipitation and temperature are used as covariates for the flow of the same month. However, for these storms flooding is linked to precipitation and temperature in the month of flooding and the preceding month. Therefore, the GEV model simulates the flood in the preceding month and/or underestimates the flood magnitude if the precipitation is split between two months. While this is a limitation for matching individual historical events, primarily timing, it is not a major concern in future projections. This is because, for the purposes of risk calculations, it really doesn't matter in which month the GEV model simulates the flood event as long as it realistically captures flood magnitude behavior.

Comparing the GEV model distribution to the other observed floods (blue triangles), the distribution encompasses the observed flood magnitude (within the 5th and 95th percentile) for all except for two of the floods (1955 and 1963). For 1963, the VIC simulated and observed floods are in close agreement (the difference between the natural logarithm of simulated and observed flows is the smallest of any event at 0.5%) and the discrepancy with the GEV model is ~~explained by~~consistent with the flood timing described above. The 1955 flood resulted from 38 cm ~~15 inches~~ of melted snow combined with 33 cm ~~13 inches o~~of rainfall over a three day period [O'Hara et al., 2007]. In the historical forcings used to drive the VIC model December 1955 has 75 cm ~~29.6 inches~~ of precipitation which is the highest December precipitation value in the historical period. In this instance the VIC simulated flow falls within the interquartile range of the GEV model, ~~is in good agreement with the VIC simulated flow~~ but the high monthly precipitation results in an overestimate of the flood magnitude. Again, this is a limitation of using monthly forcings because the total December precipitation is used as a covariate and not a storm specific value though in many cases the storm specific values constitute the bulk of the monthly precipitation totals.

Figure 4 is a time series plot of VIC historical simulated flow along with the median and 5th to 95th percentile flow of the GEV model. As would be expected from the model fit demonstrated in Figures 2 and 3, Figure 4 shows that the VIC simulated flows are generally close to the median GEV modeled flow and nearly always fall within the 5th to 95th percentile range. Although there are differences in the simulation of individual events discussed above, the median simulated flood magnitudes are only greater than the maximum observed flood in two instances.

In general, Figures 3 and 4 show that ~~we~~the VIC simulated flows match closely with the observed floods (based on percent difference in the natural logarithm of flows) and that the interquartile range of the GEV distributions encompass the observed and simulated flows in most instances. ~~note good agreement between simulated and observed flood magnitudes even though~~ Figure 3 does illustrate~~s~~ some of the complications in matching individual events, however based on analysis of the driving forces behind each individual event we are able to document the sources of these discrepancies. Based on this analysis we conclude that the model behaviour is a reasonable match with the natural system.

~~. Figure 4 is a time series plot of VIC historical simulated flow along with the median and 5th to 95th percentile flow of the GEV model. As would be expected from the model fit demonstrated in Figures 2 and 3, Figure 4 shows that the VIC simulated flows are generally close to the median GEV modeled flow and nearly always fall within the 5th to 95th percentile range. Although there are differences in the simulation of individual events discussed above, the median simulated flood magnitudes are only greater than the maximum observed flood in two instances. This demonstrates that the model behavior is a reasonable match to the natural system.~~

4.3 Future flood risk

Future flood risk is calculated using equation (5b) from Table 1. For the first part of this analysis we define 'flood' as one-day flow exceeding 1,065 cms (37,600 cfs). This is the maximum historical ~~ungauged~~ unregulated flow at Reno from the January 2, 1997 event and is considered to be the design flood for flood protection infrastructure design. For each simulation month (1950-2099 November –April) exceedance probabilities are calculated for every climate projection (234 in total) using the selected non-stationary GEV models from Table 3 (fit to the historical observations) and the projected monthly precipitation and temperature. As detailed in the section 3.2, when exceedance probabilities are time dependent, the flood risk (refer to equation 5b, Table 1) is a function of both the length of the design life and the period of operation. Figure 5 plots the risk of flood versus project life for three time periods, 1950 to 1999, 2000 to 2049 and 2050 to 2099. In other words this is the risk of a flood exceeding 1,065 cms ~~37,600 cfs~~ in the next n years if you are standing in 1950, 2000 or 2050. The median and interquartile ranges show the distribution of the 234 climate projections simulated. Here we use the interquartile range, as opposed to the 5th and 95th percentile, to focus on the central tendencies of each time period and not the variability between projections. Note that the ranges presented here express the variability between climate models. Uncertainty of the VIC model is not investigated directly here. For more detailed analysis on uncertainty in VIC simulations the reader is referred to Elsner et al. [2014].

For both Farad and Reno there is a clear positive shift in flood risk between the three time periods. In all cases the median risk for each subsequent time period falls outside the interquartile range of the preceding time period although the prediction spread for Reno is greater than Farad. It is important to note that the flood risk is actually higher at Farad than Reno in both the historical and future periods despite the fact that the observed flow distributions at the two stations are very similar (refer to Figure2). This shift between Farad and Reno is caused by the differences in the shape parameters (refer to Table 3). Farad has a heavier tailed distribution and therefore flood risks are increased. The sensitivity of the model parameters (and the associated flood risk) to small differences in the flow and covariate distributions is further demonstrated by Figure 6.

Figure 6 presents the project life risk from Figure 5 for three project life periods (10, 20 and 30 years). Boxplots show the non-stationary model results for the 234 climate projections with the different time periods compared side by side. Also, the risk calculated using a stationary GEV model and a stationary LP3 model (i.e. the distribution prescribed by Bulletin 17B fit using the L-moments [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 ~~choice~~ approach and non-stationary parameters. For instance, both stationary models are fit to the same historical simulated flows ~~data~~ (one using MLE and the other using L-moments) so differences between the stationary lines reflect the impact of model choice ~~on~~ 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. Finally, variability between the boxplots for a given design period demonstrates the evolution of risk over time (i.e. the impact of climate trends on risk). The latter (i.e. changing risk over time), is the purpose of this analysis, however before assessing trends over time we must first discuss the impact of model choice and parameters on risk estimates.

For both of the stationary methods, the risk increases with project life following equation (5a) from Table 1. The distinction between these lines and the non-stationary approaches is that, with the stationary approach, a single exceedance probability is calculated for the given flood magnitude and this probability is assumed to remain constant throughout the design life. Also, for both stationary approaches the model is fit directly to the historical one day maximum flow distribution and no covariates are required (note that stationary models are not fit to the future time periods because this would require future simulated flows). Comparing between the GEV (blue line) and the LP3 (red line) stationary models there is a 10-20% increase in risk between the two models. This difference is purely a function of model form and highlights the sensitivity of the risk calculations to model choice.

Contrasting the difference between the stationary (blue line) and the non –stationary GEV for the historical time period (grey boxplot) illustrates the effect of adding non-stationary parameters to a given model form. Recall that in both cases the GEV model is fit to the historical simulated flows. However, for the stationary approach, model fitting results in a single set of parameters (location, scale and shape) whereas with the non-stationary approach we derive the shape parameter and a set of coefficients for linear models to determine the location and scale parameters based on precipitation and temperature values. Thus, for the non-stationary approach, different location and scale parameters are calculated for every historical cool season month and GCM model (234).

Overall, there is close agreement between the stationary (S) and average non-stationary (NS) location parameters (6.55 S vs. 6.64 NS at Farad and 6.63 S vs. 6.78 NS at Reno). However, for both gauges the scale parameter is lower with the non-stationary approach (1.30 S vs. 0.94 NS at Farad and 1.28 S vs. 0.96 NS at Reno). At Reno the shape parameter is similar (-0.24 S vs. -0.27 NS), but at Farad the difference is somewhat larger (-0.24 S vs. -0.18 NS). Differences in model parameters are reflected in the distance between the stationary GEV model (blue line) and the median historical non-stationary GEV boxplots (center of the grey boxplots) in Figure 6. For Reno the stationary line is closer to the historical boxplots. However, at Farad, the non-stationary boxplots are consistently higher than the stationary line. The larger differences between the stationary and non-stationary models for Farad result from changes in the shape parameter between the stationary and non-stationary model fits. This change demonstrates the sensitivity of model results to changes in model parameters.

As with Figure 5, Figure 6 shows significant increases in risk moving into the future and subsequently larger differences between the stationary and non-stationary approach. By the second future period the differences between the stationary and non-stationary models can be as much as 50% or more. For both gauges difference in risk between the non-stationary and stationary approaches grows over time, indicating greater potential to underestimate risk looking further into the future if non-stationary parameters are not adopted.

Although the figures are not shown here, results were also grouped by RCPs ~~–~~to analyze connections between greenhouse gas emission rates and changes in flood risk. We observed no clear trend in flood risk based on the different RCPs. This indicates that the variability between GCM model form and initial conditions likely overwhelms the influence of greenhouse gas emissions when comparing between scenarios. In other words, the variability between projections within any RCP scenario is larger than the difference between RCP scenarios.

Given the sensitivity of projected risk to model parameters, an obvious question is whether increases in risk over time are similarly sensitive. For the 1,065 cms ~~37,600 cfs~~ flood plotted in Figure 6, the increased risk with added project life (i.e. 20 years vs. 10 years) is greater with the non-stationary models than the stationary one at both stations. This is intuitive, given the increased flood risk with time demonstrated in Figure 5 for the non-stationary models. Although, Farad has higher risk overall, the relative increase in risk between time periods is similar between the two stations. For example, the median ten year flood risk increases by 21% for Farad comparing between the first (1950-1999) and second (2000-2049) time periods compared to 29% for Reno.

Next, analysis is expanded to a range of flood magnitudes. Figure 7 plots the flood risk over a ten year project life starting in 1950, 2000, and 2050 for flood values ranging from 283 to 1,416 cms (10,000 to 50,000 cfs). As would be expected the ten year flood risk decreases with increasing flood rate. The shapes of the curves are slightly different between Farad and Reno; flood risk decreases more sharply with increased flow at Reno than Farad. Again this behavior is a function of the shape of the distribution. Despite these differences, both gauges display clear shifts between time periods similar to Figure 5. Here again, the median risk for each subsequent period consistently falls outside the interquartile range of the preceding period.

Changes in the median flood risk (i.e. differences between the solid lines on Figure 7) between each future period and the historical period are plotted in Figure 8 for both gauges. As would be expected based on the qualitative differences in Figure 7, the shape of the Farad and Reno difference curves are slightly different. However, the salient point for this analysis is that the increased risk between periods is generally within 10% between the two stations.

Overall the increased risk between the first future period (2000-2050) and the historical period (1950-1999) is between 10 and 20% for flows from 600 to 1,200 cms ~~20,000 to 40,000 cfs~~. Similarly, the increased risk from the historical period to the second future period (2050-2099) is between 30 and 50%. Differences for the highest and lowest flows are difficult to assess because the median is skewed by the upper and lower limits of risk (i.e. 0 and 100%).

5 Summary and Conclusions

The analysis presented is unique in its incorporation of non-stationary GEV analysis using CMIP 5 projections and the design life level risk assessment. We present our findings as a relevant case study and an example application of recent developments in non-stationary flood assessment. Lacking sufficient ~~ungauged~~ unregulated flow data we simulate historical floods using the VIC model. Subsequently we use the simulated floods to fit non-stationary GEV models with downscaled monthly precipitation and temperature as covariates. Although there are some discrepancies between individual simulated and observed floods, we demonstrate that the VIC model adequately captures the range of flood magnitudes. Furthermore, we show that that the GEV modeled historical floods are in good agreement with both the VIC simulated floods and the published flood events [*USACE*, 2013b].

Discrepancies between historical and simulated events often result from the monthly time step used for covariates. This can affect the ability to model floods that are generated by precipitation that occurs in two months. Also, because the climate variables are monthly aggregates, and not event based, large floods can be generated in months with high precipitation even if that precipitation does not occur in one concentrated event. Despite these differences, comparison with historical floods demonstrated that the GEV model does a good job of encompassing historical flood magnitudes, even if some individual historical events are not matched exactly.

Using the derived non-stationary GEV models, we generate flood distributions for 234 CMIP5 climate projections from 1950 to 2099. For the historical one-day design flood magnitude of ~~37,600 cfs~~1,065 cms, results show significant increases in the frequency of high flow events in the future. From a water management standpoint this finding translates directly to increased flood risk. For example, we calculate a 21% (29%) increase risk of a ~~37,600 cfs~~1,065 cms flood over a 10 year design life for Farad (Reno) from the historical time period to the first future period, and similar increases from the first future period to the second.

Increased risk between time periods is also relatively consistent for longer design life periods and similar shifts in flood risk are noted across a range of flood magnitudes. For both stations the increased risk from the historical to the first future period is between 10 and 20% and from the historical to the second future period is between 30 and 50% for floods ranging from 600 to 1,200 cms.~~20,000-40,000 cfs.~~

The significant increases in flood risk through time indicate the importance of non-stationary flood frequency analysis for future infrastructure planning and the potential to underestimate risk when stationarity is assumed. For both stations the difference between the stationary and no-stationary approach increases over time. By the second future period differences in risk calculations between the stationary and non-stationary models can be 50% or larger. This finding is in keeping with a number of recent studies [e.g. *Griffis and Stedinger*, 2007; *Katz et al.*, 2002; *Towler et al.*, 2010] that have highlighted potential applications for non-stationary analysis of flood frequency.

An important consideration for this approach is the sensitivity of results to model parameters. In all cases the flood risk is higher at Farad than Reno due to the heavier tailed distribution that was fit. Estimated model parameters differed by station despite the fact that the flow, precipitation and temperature distributions for both locations are very similar. While these changes effected the overall risk projections the relative increase in risk over time remained consistent between stations. This indicates that the more robust metric from this analysis is the relative increase in flood risk and not the absolute values. This finding is further supported by the fact that absolute flood risk estimates could be impacted by model bias. By focusing on differences in risk we specifically highlight the impact of non-stationarity on risk assessment, as opposed to parameter sensitivity. Similarly, it is important to note that this analysis is based on natural flow estimates and does not include infrastructure development or operation. As such results indicate the potential increase in the underlying natural flood risk and not the potential increase in flood damages.

References

Akaike, H. (1974), New look at statistical-model identification, *IEEE Trans. Autom. Control*, *19*, 716-723.

Allan, R. P. (2011), Climate change: Human influence on rainfall, *Nature*, *470*, 344-345.

Cayan, D. R., K. T. Redmond, and L. G. Riddle (1999), ENSO and Hydrologic Extremes in the Western United States, *Journal of Climate*, *12*(2881-2893).

Cayan, D. R., S. A. Kammerdiener, M. D. Dettinger, J. M. Caprio, and D. H. Peterson (2001), Changes in the Onset of Spring in the Western U.S., *Bulletin of the American Meterorological Society*, *82*(3), 399-415.

Christensen, N. S., and D. P. Lettenmaier (2007), A multimodel ensemble approach to assessment of climate change impacts on the hydrology and water resources of the Colorado River Basin, *Hydrology and Earth System Science*, *11*, 1417-1434.

Christensen, N. S., A. W. Wood, D. P. Lettenmaier, and R. N. Palmer (2004), Effects of climate change on the hydrology and water resources of the Colorado river basin, *Climate Chnage*, *62*(103), 337-363.

Das, T., D. W. Pierce, D. R. Cayan, J. A. Vano, and D. P. Lettenmaier (2011), The importance of warm season warming to western U.S. streamflow changes, *Geophysical Research Letters*, *38*(L23403).

Dettinger, M. D., and D. R. Cayan (1995), Large-scale Atmospheric Forcing of Recent Trends toward Early Snowmelt Runoff in California, *Journal of Climate*, *8*(3), 606-623.

Douglas, E. M., R. M. Vogel, and C. N. Kroll (2000), Trends in floods and low flows in the United States: impact of spatial correlation, *Journal of Hydrology*, *240*(1-2), 90-105.

Easterling, D. R., G. A. Meehl, C. Parmesan, S. A. Changnon, T. R. Karl, and L. O. Mearns (2000), Climate Extremes: Observations, Modeling and Impacts, *Science*, *289*(5487), 2068-2074.

Elsner, M. M., S. G. Gangopadhyay, T. Pruitt, L. D. Brekke, N. Mizukami and M. P. Clark (2014), How does the choice of distributed meterological data affect hydrologic model calibration and streamflow simulations?, *Journal of Hydrometeorology*, 15, 1384-1403.

Franks, S. W. (2002), Identification of a change in climate state using regional flood data, *Hydrol. Earth Syst. Sci.*, *6*, 11-16.

Gangopadhyay, S., T. Pruitt, L. D. Brekke, and D. A. Raff (2011), Hydrologic Projections for the Western United States, *EOS*, *92*(48), 441-452.

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

Gilroy, K. L., and R. H. McCuen (2012), A nonstationary flood frequency analysis method to adjust for future climate change and urbanization, *Journal of Hydrology*, *414-415*, 40-48.

Griffis, V., and J. R. Stedinger (2007), Incorporating climate change and variability into Bulletin 17B LP3 model, paper presented at ASCE World Env. & Water Resour. Congress.

Gutowski, W. J., G. C. Hegerl, G. J. Holland, T. R. Knutson, L. O. Mearns, R. J. Stouffer, P. J. Webster, M. F. Wehner, and F. W. Zwiers (2008), Causes of Observed Changes in Extremes and Projections of Future Changes in Weather and Climate Extremes in a Changing Climate. Regions of Focus: North America, Hawaii, Caribbean, and U.S. Pacific Islands*Rep.*, Washington, DC.

Hirsch, R. M. (2011), A perspective on nonstationarity and water management, *Journal of the American Water Resources Association*, 47(3), 436-446.

Jain, S., and U. Lall (2001), Floods in a changing climate: Does the past represent the future?, *Water Resources Research*, *37*(12), 3193-3205.

Katz, R. W. (2010), Statistics of extremes in climate change, *Climate Change*, *100*, 71-76.

Katz, R. W., M. B. Parlange, and P. Naveau (2002), Statistics of extremes in hydrology, *Advances in Water Resources*, *25*, 1287-1304.

Kunkel, K. E. (2003), North American Trends in Extreme Precipitation, *Natural Hazards*, *29*, 291-305.

Liang, X., E. F. Wood, and D. P. Lettenmaier (1996), Surface soil moisture parameterization of the VIC-2L model: Evaluation and modification, *Global and Planetary Change*, *13*(1-4), 195-206.

Liang, X., D. P. Lettenmaier, E. F. Wood, and S. J. Burges (1994), A simple hydrologically based model of land surface water and energy fluxes for general circulation models, *Journal of Geophysical Research*, *99*(D7), 14415-14428.

Madsen, T., and E. Figdor (2007), When it Rains it Pours - Global Warming and the Rising Frequency of Extreme Precipitaiton in the U.S., edited, Environmental America Research and Policy Center.

Maidment, D. R. (1993), Handbook of Hydrology, McGraw-Hill, ISBN 0070397325.

Mailhot, A. and S. Duchesne (2010), Design criteria of urban drainage infrastructures under climate change, *Journal of Water Resources Planning and Management*, 136(2), 201-208.

Maurer, E. P., L. D. Brekke, T. Pruitt, and P. B. Duffy (2007), Fine-resolution climate projections enhance regional climate change impact studies, *Eos Tran. AGU*, *88*(47), 504.

Maurer, E. P., A. W. Wood, J. C. Adam, D. P. Lettenmaier, and B. Nijssen (2002), A Long-Term Hydrologically-Based Dataset of Land Surface Fluxes and States for the Conterminous United States, *Journal of Climate*, *15*(22), 3237-3252.

Meehl, G. A., et al. (2000), An Introduction to Trends in Extreme Weather and Climate Events: Observations, Socioeconomic Impact, Terrestrial Ecological Impacts, and Model Projections, *Bulletin of the American Meterorological Society*, *81*, 413-416.

Merz, B., S. Vorogushyn, S. Uhlemann, J. Delgado and Y. Hundecha (2012), More efforts and scientific rigour are needed to attribute trends in flood time series, *Hydrol. Earth Syst. Sci. Opinions*, 16, 1379-1387.

Milly, P. C. D., J. Betancourt, M. Falkenmark, R. M. Hirsch, Z. W. Kundzewicz, D. P. Lettenmaier, and R. J. Stouffer (2008), Stationarity Is Dead: Whither Water Management, *Science*, *319*(5863), 573-574.

Min, S.-K., X. Zhang, F. W. Zwiers, and G. C. Hegerl (2011), Human contribution to more-intense precipitation extremes, *Nature 470*, 378-381.

Mote, P.W., A.F. Hamlet, M.P. Clark, And D.P. Lettenmaier (2005), Declining Mountain Snowpack In Western North America, *Bulletin of the American Meterorological Society*, 39-49, doi: 10.1175/BAMS-86-1-39.

Mullet, C. J., P. A. O'Gorman, and L. E. Back (2011), Intensification of precipitation extremes with warming in a cloud resolving model, *Journal of Climate*, *24*(2784-2800).

Nijssen, B., D. P. Lettenmaier, X. Liang, S. W. Wetzel, and E. F. Wood (1997), Streamflow simulation for continental-scale river basins, *Water Resources Research*, *33*(4), 711-724.

Hall, J. et al. (2014), Understanding flood regime change in Europe: a state-of-the-art assessment, *Hydrol. Earth Syst. Sci.*, 18, 2735-2772.

Interagency Advisory Committee on Water Data (IACWD) (1982), Guidelines for determining flood flow frequency: Bulletin 17B of the Hydrology Subcommittee, Office of Water Data Coordination, U.S. Geological Survey, Reston, VA., 183 p.

Obeysekera, J. and J. D. Salas (2014), Quantifying the uncertainty of design floods under nonstationary conditions, *Journal fo Hydrologic Engineering*, 19, 1438-1446.

O'Gorman, P. A., and T. Schneider (2009), The physical basis for increases in precipitaiton extremes in simulations of 21st century climate change, *Proceedings of hte National Academy of Sciences*, *106*, 14773-14777.

O'Hara, B.F., G.E. Barbato, J.W. James, H.A. Angeloff and T. Cylke (2007), 'Weather and climate of the Redno-Carson City- Lake Tahoe Region, Nevada Bureau of Mines and Geology, Special Publication 34.

Pall, P., T. Aina, D. A. Stone, P. A. Stott, T. Nozawa, A. G. J. Hilberts, D. Lohmann, and M. R. Allen (2011), Anthropogenic greenhouse gas contribution to flood risk in England and Wales in autumn 2000, *Nature*, *470*, 382-385.

Payne, J. T., A. W. Wood, A. F. Hamlet, R. N. Palmer, and D. P. Lettenmaier (2004), Mitigating the effects of climate change on the water resources of the Columbia River basin, *Climate Change*, *62*(1-3), 233-256.

Pierce, D. W., T. Das, D. R. Cayan, E. P. Maurer, N. L. Miller, Y. Bao, M. Kanamitsu, K. Yoshimura, M. A. Snyder, L. C. Sloan, G. Franco, M. Tyree (2012), Probabilistic estimates of future changes in California temperature and precipitation using statistical and dynamical downscaling, *Climate Dynamics*, *40*, 839-856.

Raff, D. A., T. Pruitt, and L. D. Brekke (2009), A framework for assessing flood frequency based on climate projection information, *Hydrol. Earth Syst. Sci.*, *13*, 2119-2136.

Ralph, F. M., and M. D. Dettinger (2012), Historical and National Perspectives on Extreme West Coast Precipitation Associated with Atmospheric Rivers during December 2010, *Bulletin of the American Meterorological Society*(June), 783-790.

Reclamation (2010), Truckee River Basin Study, Fact Sheet, available at http://www.usbr.gov/WaterSMART/bsp/docs/fy2010/Truckee_Basin_Factsheet_Final.pdf.

Reclamation (2011), West-wide climate risk assessments: Bias-corrected and spatially downscaled surface water projections, Tech. Memo., 86-68210-2011-01, 138 pp., Tech. Serv. Cent., U.S. Dep. of the Inter., Denver Colo.,March.

Reclamation (2013). Downscaled CMIP3 and CMIP5 Climate Projections: Release of Downscaled CMIP5 Climate Projections, Comparison with Preceding Information, and Summary of User Needs. Downscaled CMIP3 and CMIP5 Climate and Hydrology Projections. U.S. Department of the Interior, Bureau of Reclamation, Technical Service Center, Denver, Colorado, 116 p.

Reclamation (2014), 'Downscaled CMIP3 and CMIP5 Climate and Hydrology Projections: Release of Hydrology Projections, Comparison with preceding Information, and Summary of User Needs', prepared by the U.S. Department of the Interior, Bureau of Reclamation, Technical Services Center, Denver, Colorado. 110 pp.

Regonda, S. K., B. Rajagopalan, M. Clark, and J. Pitlick (2005), Seasonal Cycle Shifts in Hydroclimatology Over the Western U.S., *Journal of Climate*, *18*(2), 372-384.

Rootzén, H., and R. W. Katz (2013), Design Life Level: Quantifying risk in a changing climate, *Water Resources Research*, *49*, 5964-5972.

Salas, J., and J. Obeysekera (2014), Revisiting the Concepts of Return Period and Risk for Nonstationary Hydrologic Extreme Events, *Journal of Hyrol. Eng.*, *19*(3), 554-568.

Slack, J.R., A.M. Lumb, and J.M. Landwehr (1993). 'Hydroclimatic data network (HCDN): A U.S. Geological Survey streamflow data set for the United Sates for the study of climate variation, 1874-1988,' *USGS Water Resour. Invest. Rep.*, 93-4076.

Small, D., S. Islam, and R. M. Vogel (2006), Trends in precipitation and streamflow in the eastern U.S.: Paradox or perception?~~,~~ *Geophysical Research Letters*, *2006*(3).

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

Stokes, J. (2002), Draft Farad Diversion Dam Replacement Project Environmental Impact Report*Rep.*, State Water Resources Control Board, Sacramento, CA.

Sun, Y., S. Solomon, A. Dai, and R. W. Portmann (2007), How Often Will it Rain?, *Journal of Climate*, *20*, 4801-4818.

Taylor, K. E., Stouffer, R. J. & Meehl, G. A. (2012), A Summary of the CMIP5 Experiment Design. *Bull. Am. Meteorol. Soc.* 93, 485-498.

Towler, E., B. Rajagopalan, E. Gilleland, R. S. Summers, D. Yates, and R. W. Katz (2010), Modeling hydrologic and water quality extremes in a changing climate: A statistical approach based on extreme value theory, *Water Resources Research*, *46*(W11504).

USACE (2013a), Final environmental impact statement for the Truckee Meadows Flood Control Project: General Reevaluation Report Volume 1. US Army Corps of Engineers, Sacramento.

USACE (2013b), Truckee Meadows Flood Control Project, Nevada: Draft General Reevaluation Report *Rep.*, US Army Corps of Engineers, Sacramento.

Van Rheenen, N. T., A. W. Wood, R. N. Palmer, and D. P. Lettenmaier (2004), Potential implications of PCM climate change scenarios for Sacramento-San Joaquin River Bain hydrology and water resources, *Climate Change*, *62*(1-3), 257-281.

Villarini, G., F. Serinaldi, J. A. Smith, and W. F. Krajewski (2009), On the stationarity of annual flood peaks in the continental United States during the 20th century, *Water Resources Research*, *45*(8).

Vogel, R. M., C. Yaindl, and M. Walter (2011), Nonstationarity: Flood magnification and recurrence reduction factors in the United States, *JAWRA*, *47*(3), 464-474.

Walter, M., and R. M. Vogel (2010), Increasing trends in peak flows in the northeastern United States and their impacts on design, paper presented at 2nd Joint Federal Interagency Conference, Las Vegas, NV.

Wood, A. W., Leung, L. 5 R., Sridhar, V., and Lettenmaier, D. P. (2004), Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs, *Climatic Change*, 62, 189–216.

Wu, H., R. F. Adler, Y. Tian, G. J. Juffman, H. Li and J. J. Wang (In press), Real-time global flood estimation using satellite-based precipitation and a coupled land surface routing model, *Water Resources Research*.

Table 1: Flood calculations using stationary and non-stationary distributions (adapted from *Salas and Obeysekera* [2014])

Eqn. #	Description	a. Stationary	b. Non Stationary
1	Exceedance probability (Probability of flood occurring in year *x*) [1]	p	p_x
2	Probability of the first flood occurring in year *x* [2]	$f(x) = (1-p)^{x-1}p$	$f(x) = p_x \prod_{t=1}^{x-1} 1 - p_t$
3	Probability of a flood occurring before year *x* [3]	$F(x) = \sum_{i=1}^{x} f(i)$	
		$F(x) = 1 - (1-p)^x$	$F(x) = 1 - \prod_{t=1}^{x}(1-p_t)$
4	Return Period (Expected waiting time between flood occurrences [4,5])	$E(X) = \sum_{x=1}^{\infty} x * P(X = x)$	
		$E(X) = 1/p$	$E(X) = 1 + \sum_{x=1}^{x_{max}} \prod_{t=1}^{x}(1-P_t)$
5	Probability of a flood occurring before the design life *n*	$R = P(X \leq n) = F(n)$	
		$R = 1 - (1-p)^n$	$R = 1 - \prod_{t=1}^{n}(1-p_t)$

[1] Flood is defined as a flow exceeding a predefined threshold

[2] $f(x)$ = Probability density function of X

[3] $F(x)$ = Cumulative distribution function of X

[4] X = Random variable denoting the waiting time for the first flood occurrence

[5] x_{max} = Time when p_x equals 1

Table 2: Negative log likelihood(NLLH) and Akaike information Criterion (~~NLLH~~AIC) scores for each model, as well as the deviance statistics (D) ~~and p-values for~~of pairwise comparisons of different model configurations (P = precipitation only, T= temperature only P&T= both) and the p-values of each D score based on a chi-squared distribution. The number of parameters in each model and the models used for comparison are listed at the bottom of the table. The selected model for each station is shaded in grey.

Station	Metric	Stationary	Non stationary Location			Non stationary Scale			Non stationary Location and Scale		
			P & T	P	T	P & T	P	T	P & T	P	T
		1	2	3	4	5	6	7	8	9	10
~~Farad~~	~~NLLH~~	~~508.9~~	~~422.9~~	~~467.1~~	~~499.7~~	~~487.3~~	~~500.9~~	~~506.5~~	~~416.4~~	~~462.2~~	~~496.9~~
Farad	p-value		< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05
Reno	NLLH	505.4	418.4	462.5	496.0	484.4	497.6	503.1	408.8	457.4	493.2
Reno	p-value		< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05
Model # compared to for pval			1	1	1	1	1	1	2	3	4

Station	Metric	Stationary	Non stationary Location			Non stationary Scale			Non stationary Location and		
			P & T	P	T	P & T	P	T	P & T	P	T
		1	2	3	4	5	6	7	8	9	10
Farad	NLLH	508.9	422.9	467.1	499.7	487.3	500.9	506.5	416.4	462.2	496.9
Farad	AIC	1023.7	855.9	942.3	1007.4	984.6	1009.8	1021.1	846.8	934.4	1003.8
Farad	D		171.8	83.4	18.3	43.1	15.9	4.7	13.0	9.9	5.7
Farad	p-value of D		< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05
Reno	NLLH	505.4	418.4	462.5	496.0	484.4	497.6	503.1	408.8	457.4	493.2
Reno	AIC	1016.8	846.8	932.9	1000.0	978.8	1003.2	1016.1	831.7	924.8	996.5
Reno	D		174.0	85.9	18.8	42.0	15.6	4.7	19.1	10.1	5.5
Reno	p-value of D		< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05
# of model parameters		3	5	4	4	5	4	4	7	5	5
Model # compared to for pval			1	1	1	1	1	1	2	3	4

Formatted: Caption, Keep with next

Formatted: Line spacing: 1.5 lines

Table 3: Summary of derived model covariates for equations 2 and 3 based on historical observations (Historical Observed) and using historical simulated data from the 234 CMIP 5 Projections (Historical Simulated Interquartile Range, IQR).~~.~~

~~-~~	~~Farad~~	~~Reno~~
~~$\beta_{0\mu}$~~	~~2.155~~	~~2.582~~
~~$\beta_{1\mu}$~~	~~0.175~~	~~0.18~~
~~$\beta_{2\mu}$~~	~~0.115~~	~~0.105~~
~~$\beta_{0\sigma}$~~	~~0.211~~	~~0.530~~
~~$\beta_{1\sigma}$~~	~~-0.013~~	~~-0.018~~
~~$\beta_{2\sigma}$~~	~~0.027~~	~~0.017~~
~~Shape (ξ)~~	~~-0.178~~	~~-0.275~~

Formatted Table

	Farad			**Reno**		
-	Historical Observed	Historical Simulated IQR		Historical Observed	Historical Simulated IQR	
$\beta_{0\mu}$	2.155	1.738	4.794	2.582	2.135	4.827
$\beta_{1\mu}$	0.175	0.053	0.148	0.180	0.066	0.152
$\beta_{2\mu}$	0.115	0.046	0.138	0.105	0.046	0.124
$\beta_{0\sigma}$	0.211	0.517	1.673	0.530	0.569	1.748
$\beta_{1\sigma}$	-0.013	-0.020	0.006	-0.018	-0.023	0.008
$\beta_{2\sigma}$	0.027	-0.012	0.022	0.017	-0.015	0.019
Shape (ξ)	-0.178	-0.389	-0.094	-0.275	-0.389	-0.070

Formatted Table

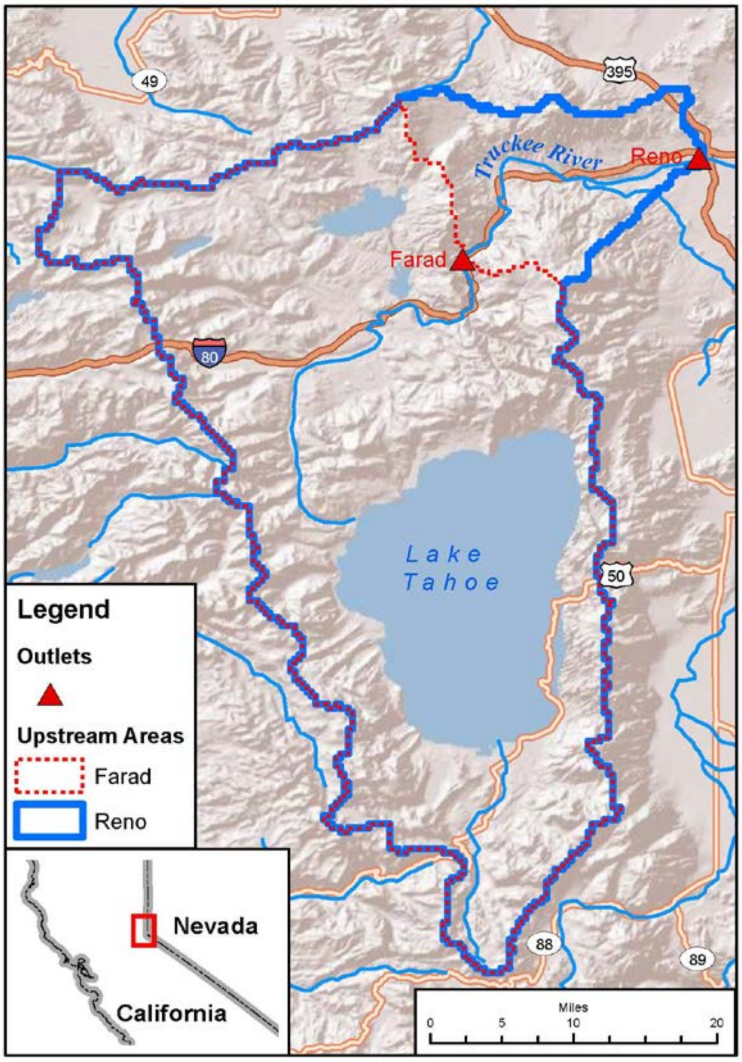
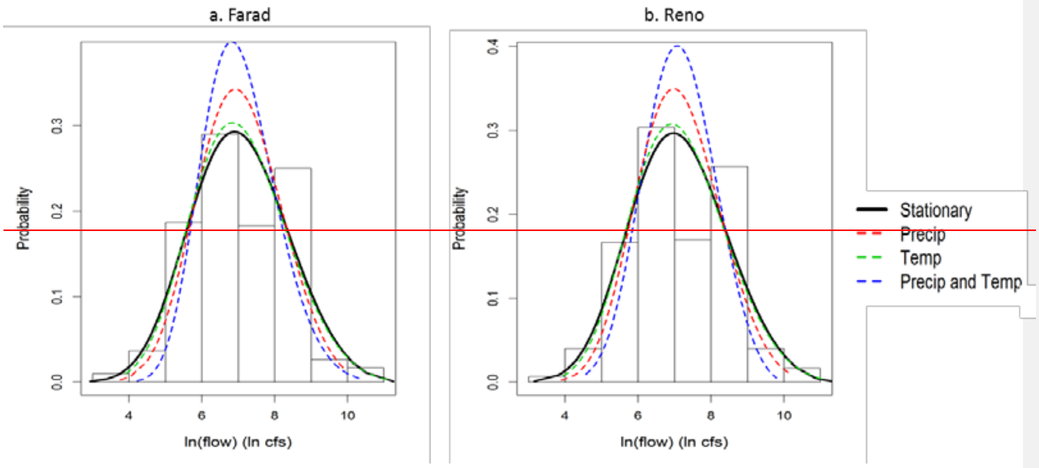


Figure 1: Map of model domain including the Farad and Reno gauges and their drainage areas.



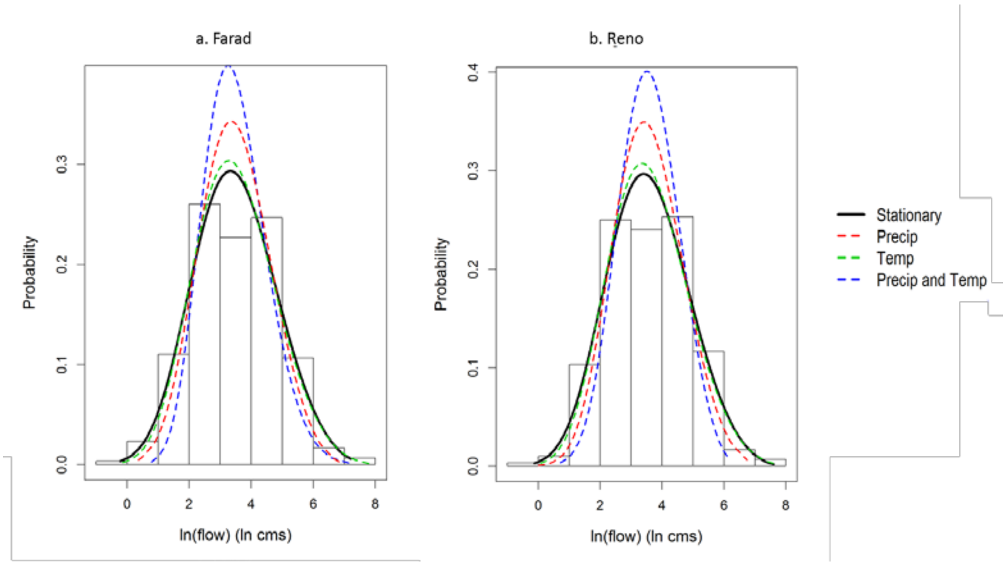


Figure 2: PDFs of fitted stationary (solid black) and non-stationary (dashed) GEV models compared to historical VIC simulated flow histogram.

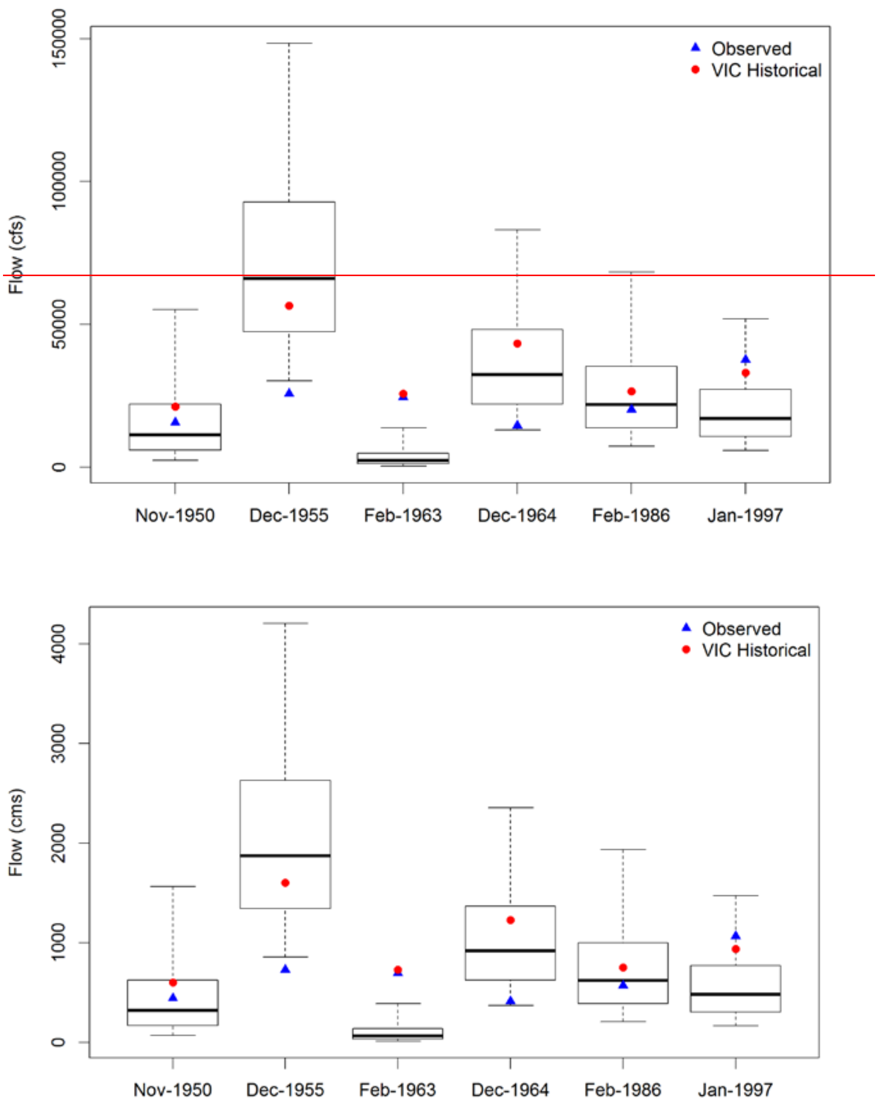


Figure 3: 'Observed' ~~ungauged~~ unregulated flow estimated from gauge records (blue triangle) compared with VIC simulated flow (red circles) and the simulated GEV distribution. Boxes span the 25th to 75th percentile of the GEV distribution for a given month and the whiskers extend to the 5th and 95th percentiles.

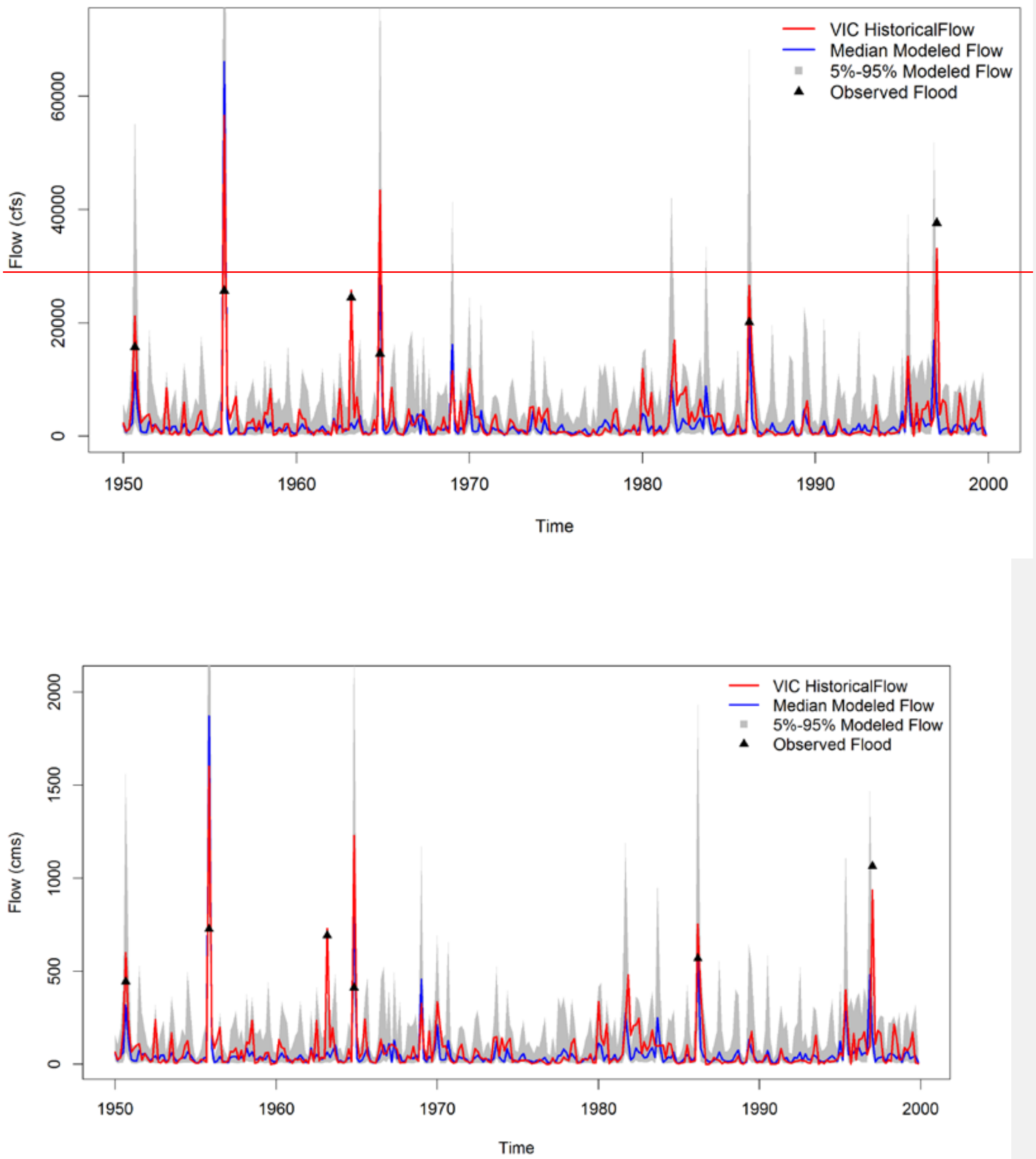


Figure 4: VIC simulated one- day flood maximums for November through April 1950 to 1999 (red lines) compared with the historical GEV distributions (blue line is median and grey shading is the 5th to 95th percentile range) and the six observed flow rates.

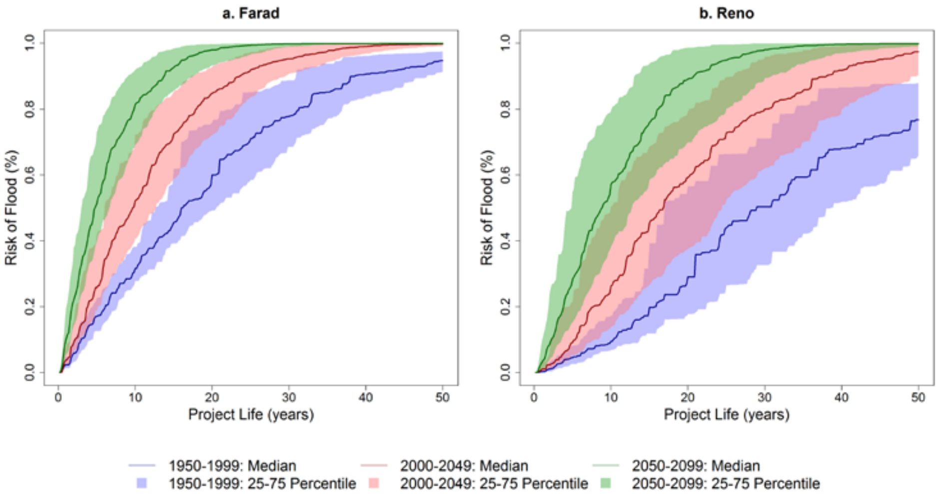


Figure 5: ~~Risk~~ Probability of one day flood exceeding historical maximum of 37,600 cfs (risk) at Farad and Reno. Solid lines represent the median risk of the 234 climate projections and shading covers the interquartile range (i.e. 25^{th} to 75^{th} percentile).

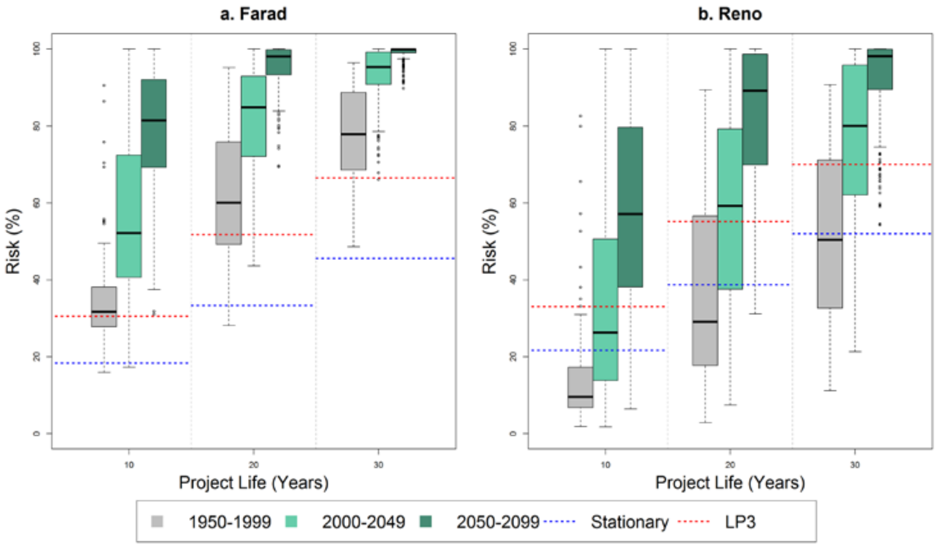
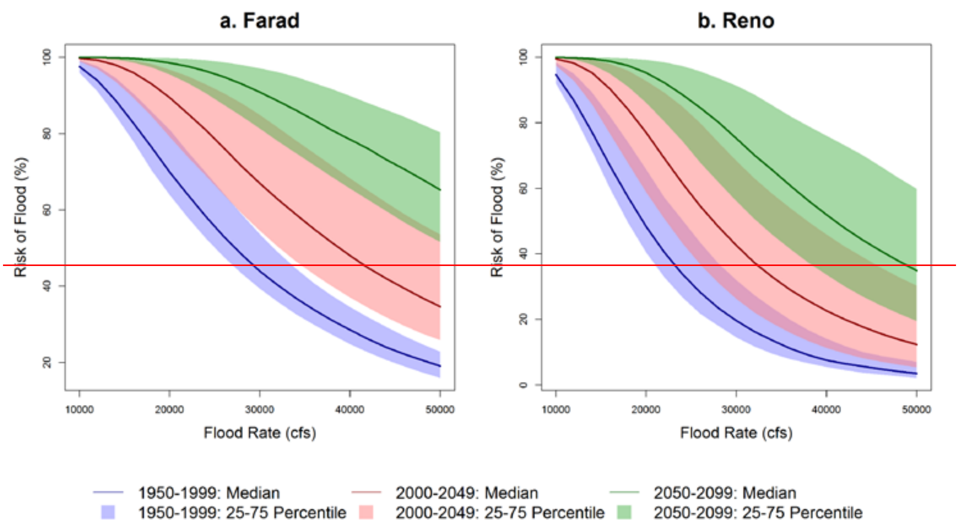


Figure 6: Boxplots of the ~~risk~~ probability of a one-day flood exceeding 37,600 cfs (risk) for three project life lengths (10, 20 and 30 years). Results are grouped by time period (1950-1999, 2000-2049 and 2050-2099). Blue dashed lines show the flood risk calculated from the stationary GEV model fit to the historical data.



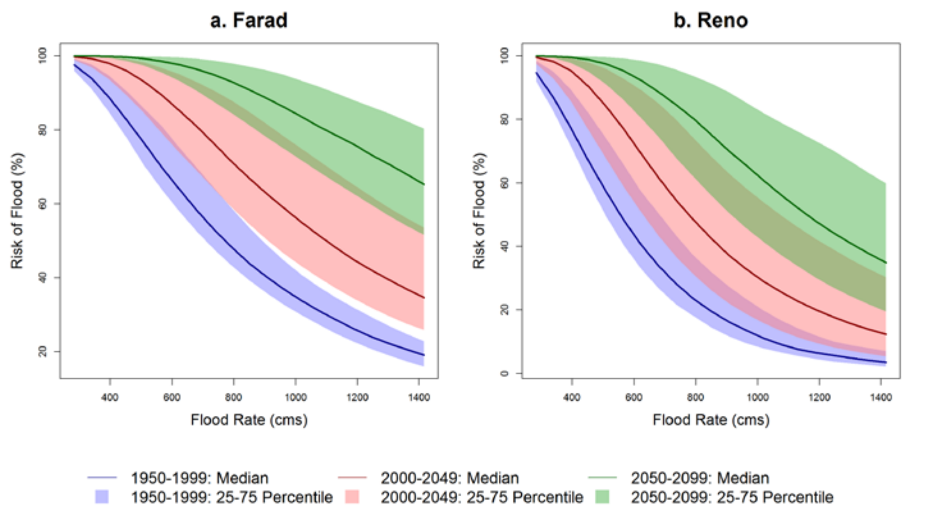
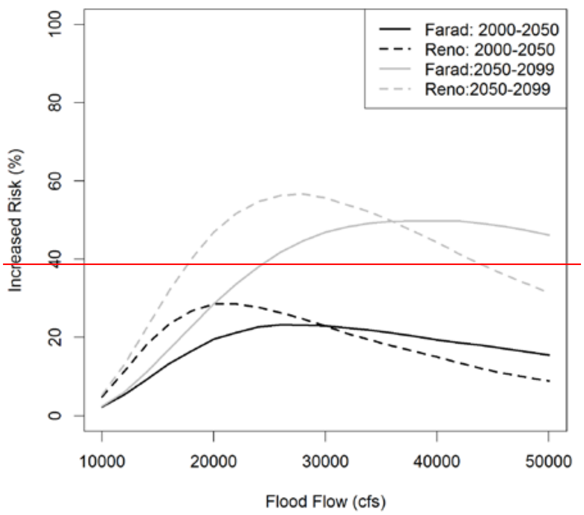


Figure 7: ~~Risk~~ Probability of flood in a ten year project life (risk) vs. median one day flood rate (a) Farad and (b) Reno for three time periods 1950-1999 (blue), 2000-2049 (red) and 2050-2099 (green). Solid lines represent the median of the 234 climate projections and shading covers the interquartile range (i.e. 25^{th} to 75^{th} percentile).



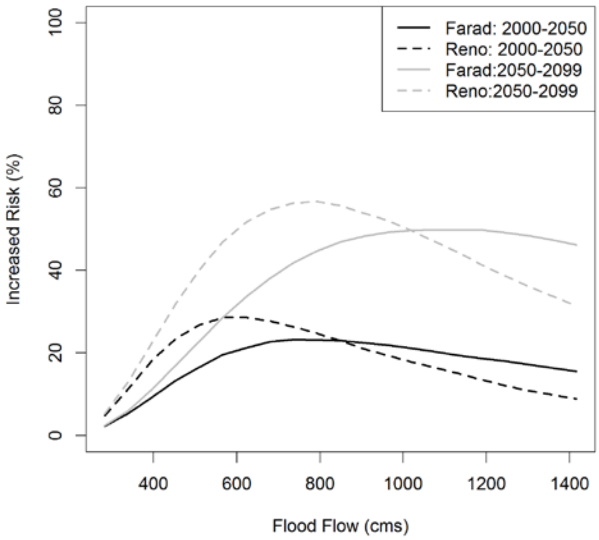


Figure 8: Increased probability of flood occurrence ~~flood risk~~ for a 10 year project life (risk) from the historical period (1950-1999) to each of the two future periods 2000-2050 (black) and 2050-2099 (grey). Farad is plotted with a solid line and Reno is a dashed line.