**We would like to thank both reviewers for taking the time to offer many constructive and helpful suggestions that have assisted in improving the quality of the manuscript. We have revised our manuscript based on the comments and corresponding responses, and have justified the reasons where proposed revisions are not made. We believe that the resulting manuscript is consequently much improved over the initial submission. Responses to individual comments can be found below.**

**Reviewer #2:**

This is a nice written manuscript. To understand the changes of extreme precipitation is very important for both the water resources and flood management. This study used a variable-resolution CESM at spatial resolution ~30 km to study the extreme precipitation. Overall, the model performance is convincing.  However, my major concern is lack of discussions on the mechanism modulating the changes in extreme precipitation. I ask a revision between minor and major which can mainly address the following comments:

Major comments:

1. Since this study only used one model and a few members (2 for present and 4 for future scenarios), the central question is how robust the predicted results compared to CMIP5 models or CESM large ensemble runs?

This is a good point. There are quite a few papers focusing on the precipitation changes within CMIP5 models globally. For examples, Kharin et al. (2013) found amplified changes (2-3 times) in extreme precipitation compared to mean precipitation in CMIP5. Specifically, from their study, P20 (globally averaged 20-year return values of annual daily precipitation extremes) increases more than 20 % in the RCP8.5 experiment by the end of the 21-st century. They also stated that the majority of the models simulated values in the 4–10 %/°C range and 1.5–2.5 %/°C range, for P20 and annual mean precipitation respectively in the CMIP5 ensemble. This further supports that extreme precipitation follows changes in temperature more closely to the Clausius-Clapeyron relationship than precipitation mean. They also argued that simulated late 20th-century precipitation extremes are plausible in the extratropics, although uncertainty remains very large in extreme precipitation in the tropics and subtropics.

Sillmann et al. (2013) also found that the contribution of very wet days to the annual total wet-day precipitation has generally increased by the end of 21st century (period 2081-2100), compared to the reference period 1981-2000 based on CMIP3 and CMIP5 output. Generally, in their study, the projected changes for the 2046–2065 period intensify toward the end of the century, with overall model agreement on the increases of total wet-day precipitation, very wet days and the heavy precipitation days index (R10mm). Those findings cover western U.S.

What we found in this manuscript are generally consistent with those aforementioned points focusing on the similar time periods, but with much more regional details. However, the disagreement exists mainly over the California, where CMIP5 shows no significant changes in mean precipitation as what we found in VR-CESMs but significant changes for P20, which are not shown in our results. This is probably due to the large inconsistency in how different models in CMIP5 projecting the changes of ENSO, which is one of the main regulators for the precipitation extremes over California as illustrated in details in the manuscript.

In Pendergrass et al. (2015), it is argued that in contrast to mean precipitation, extreme precipitation depends on the warming magnitude rather than emissions scenario in most CMIP5 models. This provides further support of the reduced uncertainty of our result due to the prescribed SST and sea ice and fixed changes of GHGs, making no need to resort to large ensemble runs. In their study, the recently developed and public-shared CESM large ensemble runs (CESM-LENS) are also involved. It is found that the spread of global-mean precipitation changes within CESM-LENS with only internal variability is much smaller (about one order) than the spread across the CMIP5 multi-model due to the extra structural variability. This conclusion also loosely applies to the extreme precipitation.

Pendergrass et al. (2015) also found that the slowed changing trend of the global-mean precipitation per degree in RCP8.5 than RCP4.5 in the CESM-LENS is consistent with the CMIP5 ensemble mean. As they argued, in CMIP5, the intermodal spread is smaller for mean or extreme precipitation over extratropical land (covering our study area) than for all land, i.e. models agree better on the response of extreme precipitation in the extratropics with relatively well represented precipitation-driven dynamics in GCMs.

In Figure 5 (originally the supplemental Figure 3), the results of precipitation features simulated by CESM at ~1 degree are given. It can be seen that, due to the complex topography over the western U.S., the spatial pattern and magnitude of precipitation are poorly presented and generally underestimated in CESM at coarse resolution. It is hard to gain the confidence that the changes of precipitation can be well captured without the incorporation of the fine-scale dynamical processes. The ability for GCMs to simulate extreme precipitation also strongly depends on the horizontal resolution as discussed (Wehner et al., 2010) with precipitation intensifies at high resolution (Rauscher et al. 2016; O’Brien et al. 2016). In addition, given the CMIP5 and CESM-LENS prediction over the 21st century are coupled ocean-atmospheric simulation, it is even implausible to compare directly to our model results here, which is forced by the bias-corrected SSTs from mean coupled CESM output.

For further reference, the changes of mean precipitation, near-surface temperature and wind pattern between period 2056-2080 and historical 1981-2005 can be found at this [website](http://www.cesm.ucar.edu/experiments/cesm1.1/LE/) provided by NCAR for at least five ensemble runs of CESM-LENS output. Even though the spatial features are not resolved, within our expectation, the overall sign of the changes in CESM-LENS over western U.S. are consistent to what we got.

Accounting for the models’ uncertainty is surely important, however, we also want to motivate the use of a highly performed model to reduce the signal noise. Multiple models mean are better for simulating the climate variability, but the accuracy might be reduced due to the compensation of the results from both good models and fair models. Our aim in this study is to know that how precipitation is supposed to be changed in diverse climate regions in the future, and what are the main mechanisms that drive corresponding changes in a well-performed model with well-represented topography. Even using different models, we suppose the physical relationship and spatial characteristics still hold.

References:

Kharin, V. V., F. Zwiers, X. Zhang, and M. Wehner, 2013: Changes in temperature and precipita- tion extremes in the cmip5 ensemble. Climatic Change, 119 (2), 345–357.

O’Brien, T. A., W. D. Collins, K. Kashinath, O. Ru ̈bel, S. Byna, J. Gu, H. Krishnan, and P. A. Ullrich, 2016: Resolution dependence of precipitation statistical fidelity in hindcast simulations. Journal of Advances in Modeling Earth Systems, 8 (2), 976–990.

Pendergrass, A. G., F. Lehner, B. M. Sanderson, and Y. Xu, 2015: Does extreme precipitation intensity depend on the emissions scenario? Geophysical Research Letters, 42 (20), 8767–8774.

Rauscher, S. A., T. A. OBrien, C. Piani, E. Coppola, F. Giorgi, W. D. Collins, and P. M. Lawston, 2016: A multimodel intercomparison of resolution effects on precipitation: simulations and theory. Climate Dynamics, 47 (7-8), 2205–2218.

Sillmann, J., V. Kharin, F. Zwiers, X. Zhang, and D. Bronaugh, 2013: Climate extremes indices in the CMIP5 multimodel ensemble: Part 2. Future climate projections. Journal of Geophysical Research: Atmospheres, 118 (6), 2473–2493, doi:10.1002/jgrd.50188.

Wehner, M. F., 2013: Very extreme seasonal precipitation in the NARCCAP ensemble: model performance and projections. Climate Dynamics, 40 (1-2), 59–80, doi:10.1007/ s00382-012-1393-1.

2. What is the inter-member spread of the results by comparing the four members in future simulations? Are they showing consistent trends of precipitation changes?

Thanks for pointing out this. The inter-member variability has been illustrated in the way of the differences between each member and the ensemble mean over each time period (see the supplemental figures S1, S2 and S3 for the three time periods’ internal variability, respectively). It can be told that the variability within the members’ yearly-averaged output is quite small comparing to the changing signal among different time periods. As further supported by the relevant statistical tests (here, we used the student t-test), different members of the same ensemble are not significantly different, showing a converge results in our conclusions.

3. Any mechanism or discussion related to the changes of extreme precipitation in addition to the modulation from ENSO?

ENSO is definitely one of the main regulators for the changes of extreme precipitation. In addition, the changes of extreme precipitation over western U.S. in this study are mainly due to the increased large-scale water vapor influx from the eastern Pacific Ocean compound with the orographic forcing effects. In the future, as the climate warms, intensified water vapor influxes are expected to cause larger heavy-rainy events. The modulation of ENSO is directly related to the inter-annual variability of precipitation extremes. More details can be found in Section 5 of the manuscript. (Nothing new is added here.)

Minor comments:

The figure quality can be improved.

i.e., Figure 2: The visibility can be improved if the labels of the longitude (right 3 figures) can be removed.

I am sorry; do you mean the latitude instead?

**Reviewer #3:**

The manuscript describes results from a variable resolution CESM with spectral element dynamic core with an overall resolution of ~110km and a grid refinement of ~25km over the western USA. The buffer zone between the two nests is approximately along the central US and well north and south of the Western USA. This variable resolution model has been shown to perform reasonable well without introducing any significant effects on the global scale circulation by Zarzycki et al., 2015, where they used a refinement over the North Atlantic. However, it was noted that model introduced excess precipitation within the refined domain. This paper discusses a refinement of the grid over the western USA.

The manuscript reads like a technical report submitted to a funding agency rather than a paper meant for a journal. A series of contour plots are presented with results from VR CESM and a number of observational data sets. I am not sure how to read this paper as a Journal of climate publication. It does not describe a detailed evaluation of the model performance, a new parameterization or model sensitivity. It reads like a summary of results from a series of simulations. It would be very helpful if there is additional analysis presented that makes a compelling case for the reader to spend time going through the paper either due to a unique aspect (extremes of precipitation? Or since you are using NARR dataset may be the diurnal variability of precipitation) when compared to a non variable grid CESM at 110 km resolution. Overall, it may be useful to present some comparison for this region between the CESM and vr CESM for precipitation, precipitable water,monsoons, atmospheric rivers or column water vapor over this region for seasonal and for monsoon season. This seems like a great new model, what is unique about it and how well does this model compare for may be specific cases (like monsoons ) when compared to very high resolution non hydrostatic models like WRF or REGCM that also aim to derive this type of information? (Those are very general comments. Paul, do you think we need to make responses for this?)

a) The reference figures were generated assuming equal confidence in the datasets from UW gridded dataset, CPC and NARR. This is most likely incorrect, as NARR has been shown consistently underperforms observations (CPC) over the western part of the country (Bukovsky and Karoly, 2007; Bukovsky, M. S., and D. J. Karoly, 2007: A brief evaluation of precipitation from the North American regional reanalysis. J. Hydrometeor, 8, 837–846.).

Thanks for pointing out this. We acknowledge that, as a reanalysis dataset, NARR is not supposed to be as good as the gridded observations as CPC or UW. We agree that these datasets should not be treated as equal confidence. Our purpose is to combine the gridded observations and reanalysis dataset together to account for the uncertainty in different sources of references. We’d like to see if the climate simulation of VR-CESM could be as good as either observations or reanalysis dataset.

In another paper of ours (Huang et al. 2016), we did examine the differences among various references including Daymet, PRISM, UW, CPC and NARR over California. Overall, UW outperforms CPC, and CPC is better than NARR but not that much.

Reference:

Huang, X., A. M. Rhoades, P. A. Ullrich, and C. M. Zarzycki, 2016: An evaluation of the variable resolution-CESM for modeling California’s climate. Journal of Advances in Modeling Earth Systems, doi:10.1002/2015MS000559.

b) Was the observational data regridded to the model output for each of the datasets used or the model data interpolated onto the observational data grids?

We applied interpolation as the former case to maintain the consistency of the data simulations.

c) The performance of VR CESM has been compared to similar resolution WRF simulations and said to compare favorably with only a modest improvement in performance at higher spatial resolutions. Wang and Kotamarthi (2015; Wang, J and Kotamarthi, V. R.: High resolution dynamically downscaled projections of precipitation in the mid and late 21st century over North America, Earth's Future, 3: 268–288. doi:10.1002/2015EF000304, 2015) performed a comparable set of simulations at 12km resolution with WRF and it shows significant improvement over the host CESM simulations for this region.

Thanks for pointing out this. We need to clarify that in the manuscript, it is stated that VR-CESM demonstrated comparable performance to WRF at 27 km with similar downscaling resolution (about 28km) showing significant improvement contrasted to CESM at ~1° resolution. This is consistent with the findings by Wang and Kotamarthi (2015) stating that WRF simulations at 12 km significant improved the host CESM at near 1° resolution. We also pointed out VR-CESM at higher-resolution (about 14 km) did not appear to substantially improve model accuracy compared to the one at 28 km constrained by the lack of scale-aware model parameterization schemes.

Reference:

Wang, J and Kotamarthi, V. R.: High resolution dynamically downscaled projections of precipitation in the mid and late 21st century over North America, Earth's Future, 3: 268–288. doi:10.1002/2015EF000304, 2015

d) It will be very useful for the reader to have a comparison of the CESM and VR CESM over this region for evaluating the model performance. I am not entirely sure if the overall precipitation in the refined grid domain has increased, decreased or stayed the same when compared to the CESM.

Agree. Actually, this work has already been investigated with the relevant plot included in the supplement (originally Figure S3). For a more clear explanation, we have moved this plot to the main content as Figure 5 with further modifications. As aforementioned in the response to the first question by the other reviewer, the results further support that the ability for GCMs to represent precipitation strongly depends on the horizontal resolution with precipitation intensifies at high resolution.

Overall, precipitation patterns over complex topography are poorly represented in the ~1° dataset without capturing the spatial patterns induced by orographic effects. Over the Cascades and the Sierra Nevada, where majority precipitation locates, total precipitation is grossly underestimated by the coarse resolution data, as compared to gridded observations, so does precipitation extremes. Precipitation has otherwise been smoothed out over the coastal areas and the mountainous regions of the northwest U.S at coarse resolution. It is also found that CESM without nested refined domain tends to underestimate the low-rainy days at the windward side but overestimate the ones over the lee side especially for the Cascade ranges (also refers to Figure 3). This bias is reduced with finer resolution though not fully resolved. This result clearly underscores the benefits of high resolution (particularly the representation of topography) in simulating precipitation features.

(Paul, do you think it is necessary to put the plot in the main text? I am kind of concerned with the repeated display of the Pr features from VR-CESM, UW, and CPC.)

e) The abstract indicates that the manuscript is attempting to evaluating the spatial patterns of precipitation produced by VR CESM with observations. It would have been much more helpful to present spatial correlations between the observational data sets and model results to emphasize this aspect of the work. It is really difficult to evaluate this based on a series of contour plots that are presented as difference over each pixel.

Thanks for the suggestion. The spatial correlations have been added in the text to further support the arguments with values ranging from 0.7 to 0.9.

(I feel a table is not necessary. Paul, what do you think?)

f) SDII is referred to as the spatial pattern of precipitation intensity in the text (line 240) and table 1 lists as simple precipitation intensity index (precipitation amount/R1), where R1 is said to be number of days with more than 1mm of precipitation. These two sound very different, which one is correct?

Thanks for pointing out this. We need to clarify that SDII refers to the simple precipitation intensity index. The corresponding sentence has been rephrased to “… for the simple precipitation intensity index (SDII), its spatial pattern agrees well …” from “ … the spatial pattern of precipitation intensity (SDII) agrees well …”.

g) It really doesn’t serve any purpose to provide figures for each of the R metrics from 1mm to 40mm. It would be best if we have the mean and the extreme (> 95 percentile) and the number of days with precipitation. This will make the figures more accessible to a reader and easier to understand.

Thanks for pointing out this. In order to achieve both meaningful and comprehensive analyses, a variety of relevant indices have been explored at the beginning as mentioned in the Methodology section. Those indices include the ones as defined by ETCCDI and other commonly used metrics covering different percentiles. To our standpoint, interpreting the precipitation events with different levels of strengths is more straightforward and informative to water resources management and climate adaptation. As part of the initial analyses, the percentiles do convey the signal of potentially shifted precipitation distribution, but did not add significant information to the indices we defined in the paper. The changes of the extreme in the scope of percentiles are also loosely imbedded in the frequency distributions of precipitation shown in Figure 10.

We agree that showing the percentile like 95th will assist the readers to get the sense of the extreme changes easier. As the mean and the number of days with precipitation have already included Figure 2 and Figure 7. Here, for a complimentary illustration, the 95th percentile (P95) based on all days over each simulation period is added in Figure 11 next to the precipitation frequency distributions.

Again, the shift of the precipitation to a more extreme condition is pronounced when the warming keep intensifies till the end of the 21st century over the northwest U.S. (P95 increased for about 20-30 %). For the dry area including the southwest and intermountain west, precipitation tends to be more extreme (P95 increased for about 15%) with the increase of the mean precipitation and number of rainy days (see Figure 7) from *hist* to *mid*. However, this trend is suppressed when the warming persists till the *end* over southern California and remaining southwest area where convective precipitation dominates. This is due to the insufficient compensation of air water vapor to the exponentially enlarged saturated vapor pressure. (The percentile P95 did not really add new information to the frequency plot.)

h) The increased number of extreme precipitation events noted in lines 250 262, is it something similar to observed in the CESM that is exaggerated by the VR CESM or a new feature that is present in VR CESM?

Thanks for pointing out this. For the suppressed SDII over the Great Plains during the warm season, it is also present in CESM and not really alleviated in VR-CESM. As for the exaggerated precipitation intensity over the western flank of the Sierra Nevada, this is indeed a new feature in VR-CESM, as CESM cannot even represent the orographic precipitation over the Sierra Nevada spatially.

i) The projected changes for the future timeslices – it would be helpful if you could put this in context of CMIP5 model results (either as table or a figure).

Thanks for the suggestion. The information for this point can be found in the response to the first question of the other reviewer. (Paul, do you think it is still necessary to pull out the dataset and add a table here?)

**We would like to thank both reviewers again for taking the time and effort to provide your thoughtful and thorough commentary. We hope that these revisions have addressed your concerns and believe the quality of the manuscript has been greatly improved with your input.**

**Sincerely,**

**Xingying Huang and Paul Ullrich**