**Responses to Reviewers**

| ***No*** | ***Comments*** | ***Responses*** |
| --- | --- | --- |
| **Reviewer #1** | | |
| 1 | This manuscript is very well written and organized. It is very easy to be followed. The main focus of the study is on evaluating the performance of CFS for Vietnam. The authors also use a dynamical downscaling approach to improve the forecasts of CFS. However, the whole study lacks the mechanistic analysis and also does not have clear science questions. I have some specific questions: |  |
| 1) | The CFS has a high kill forecast for precipitation for Vietnam over April while it has a low skill in predicting July precipitation. What are the physical reasons behind these different forecasts? |  |
| 2) | Why does the forecast skill increase when increasing the lead times? This can bee seen in several graphs in the manuscript. |  |
| 3) | I think the dynamical downscaling is a very good idea for this study. However, it seems to me that the major improvement with the downscaling is due to the higher model resolution that better describes the topography for the study region. How about the model physics? If the horizontal spatial resolution of the CFS increases to that of RegCM4.2, can the CFS achieve similar results? I think the downscaling is the most interesting part of the study, and the authors should give in-depth analysis to see how and why the regional model produces the better results than the original CFS. |  |
| 2 | In general, I think this is a very interesting study but lacks in-depth physical analysis. I would suggest that the manuscript should be revised significantly before it could be published in this journal. |  |
| **Reviewer #2** | | |
| 1 | General: This manuscript evaluates NCEP products and its downscaling outputs over the Vietnam. They found the CFS\_Rfc can capture the seasonal variability of the Asian monsoon circulation and rainfall distribution. The downscaling is advantageous over the raw CFS in specific climatic sub-regions, particularly in the Northern part in January, and in the Central Highlands of Vietnam in July. Results are useful for seasonal forecast in Vietnam. Manuscript is well written. Suggest it acceptable after the followings are addressed. |  |
| 2 | Over 80% of this manuscript is evaluations of CFS datasets. Only less than 20% is evaluation of downscaling. However, from the introduction, it reads like the purpose of this study is obtain local climate information using downscaling. Then, what is the ultimate purpose of this study, seasonal forecast ability of CFS or downscaling over the Vietnam? If it is the latter, evaluation of the downscaling is short and weak. |  |
| 3 | The CFS reanalysis (CFSR) are used to validate circulations of CFS\_Rfc and CFS\_Ope, in which CFSR is used as initiation. It seems like self-validation. Valid validation should use the third-part dataset. Suggest author use another source reanalysis in validation. For instance, as the latest generation well-known reanalysis, ERA-Interim is found better presents surface air temperature and precipitation characteristics in many places of East Asia (Zeng and Wang 2012; Gao et al. 2014). |  |
| 4 | CFS precipitation is evaluated compared to GPCC and observation. What is the purpose using two reference datasets, to address the uncertainty or something else? |  |
| 5 | CFS precipitation is evaluated compared to station observations. However, Vietnam is narrow and small with so many observation stations. However, CFS has a resolution of 1 degree. It is very likely that many stations locate in one CFS grid cell. Is there any post-processed adjusting done from the CFS grid cell precipitation to comparison with station records? |  |
| 6 | L318-320, "even though there are some moderate to high correlation coefficients (0.6-0.8) in April and January, the correlation coefficients are generally low (~0.2-0.4), or even negative." needs to be clarified. |  |
| **Reviewer #3** | | |
| 1 | This study is an important contribution to the emerging body of literature helping to establish the efficacy of a regional atmospheric modeling paradigm for improved sub-seasonal to seasonal forecasting. Vietnam is an interesting demonstration test case for CFSv2 downscaling, as a country whose climate is strongly governed by convective precipitation during monsoon in southeast Asia and where there may be some climate relationship to large-scale atmosphere-ocean variability that yields potential seasonal predictability. So from all those standpoints, there is good motivation to publish this manuscript.  However, there are necessary improvements that needs to be made in this work to make it suitable for publication, that I note in specific comments below, and these would constitute major revision. |  |
| 2 | I have three overarching comments. First, in a revised version I would like the authors to consider much more what physically would drive climate predictability in this part of the world. As a requirement for revision, I suggest that they need to devote a dedicated section in their manuscript to this topic. This new section would present the relationship of seasonal precipitation anomalies (e.g. in Fig. 4 and similarly structured figures) to possible climate drivers. This could include, for example, ENSO, MJO indices, and Indian-ocean SST dipole indices. Some possible sources for these indices are included as an example from the NOAA ESRL website, but there would be many others:  https://www.esrl.noaa.gov/psd/gcos\_wgsp/Timeseries/DMI/  https://www.esrl.noaa.gov/psd/data/climateindices/list/  The selection of a particular index to correlate with observed precipitation should be motivated by some sort of documented relationship of precipitation in Vietnam to known climate driver(s). The regions where there is relatively higher skill in CFSv2 and CFSv2-RegCM hindcast and forecast simulations can then be directly compared to the areas where large-scale climate variability significantly influences seasonal precipitation. My thought is that higher skill should correspond to geographic areas where at least a significant relationship of precipitation to ENSO exists, if that is already being used in Vietnam in current operational practice. |  |
| 3 | Second, in my opinion what is of greatest interest here, in terms of novelty of work, are the downscaling of the operational CFSv2 data with RegCM4. But we have really only three years of forecast data to consider, and that is not enough to make robust assessments of the climatological performance of a regional climate model seasonal forecast system, if creating one for operational purposes in Vietnam is the ultimate objective here. The dynamical downscaling can also be done with the CFSv2 reforecast data, as well as the operational data, so the only reason not to do it would be for a lack of personnel time and/or computing resources. If that would be the case, in the absence of downscaling more years to create the CFSv2 downscaled climatology, the authors need to focus on the specific climate drivers within the three years they have downscaled, and see if the downscaled model is better realizing any regional climate signals (related to ENSO, Indian ocean variability,  MJO, etc.). But I would suggest creating a longer climatology of downscaled reforecasts, if at all possible. |  |
| 4 | Third, this paper would be substantially stronger if the authors would include a more rigorous quantification of the statistical significance of their results, as this was a repetitive theme in many of the specific comments. It will provide greater focus to results that are important to emphasize versus what are not. |  |
| 5 | Specific comments: |  |
| 1) | Introduction, paragraph beginning line 74: The authors should also mention that dynamical downscaling better addresses the issue of possible non-stationarity and better accounts for the physical processes of precipitation generation at the mesoscale, especially for convection in sub-tropical to tropical environment. |  |
| 2) | Last sentence of paragraph beginning line 79: The authors are not accurate with referring to some of their referenced works here as 'dynamical downscaling.' Dynamical downscaling specifically implies the use of a regional atmospheric model that is forced at its lateral boundaries by a coarser model (typically a global climate model). What seems to be referenced here are studies what actually utilize global modeling systems. In truth, there are actually relatively few examples where regional atmospheric models have been applied to address the question of sub-seasonal to seasonal forecasting. One such community effort the authors should be aware of and cite within this paper is the MRED experiment (De Haan et al., 2015, Theor. Appl. Climatol.; De Sales and Xue, 2013, Clim. Dyn.; Shukla and Lettenmaier, 2013, J. Geophys. Res. Atmos.). The basic conclusion of MRED is that regional climate downscaling adds value in those regions where the driving global climate model has some existing skill, and I think that should define the driving hypothesis for this particular Vietnam study. To their credit, the authors do cite some dynamical downscaling work with WRF for China later, but I'm wondering if their conclusions were similar to MRED. |  |
| 3) | Paragraph beginning line 102: The authors may want to consider that many readers would not have an a priori sense of the climatology of Vietnam. Maybe present some orienting cartoon type figures here that show the driving physical mechanisms of climate variability as a function of season. The reason the rainy season in central Vietnam occurs more in October and November is especially not clear to me, though my suspicion is that this may be the part of Vietnam that is relatively most influenced by tropical cyclones. |  |
| 4) | Paragraph beginning line 120: I would like to see a lot more information on the operational seasonal forecast modeling system for Vietnam. Is there documentation? A website? How does it compare to operational practice in other countries? Basically, I would look for some documented benchmark for skill of the existing statistical forecast, as that seems a necessary perquisite to motivate the present work. Some sort of corresponding supporting figure would be nice, especially for an international audience. The fact that ENSO is the basis for the existing statistical forecast strongly motivates what I said earlier in the summary comments. |  |
| 5) | Paragraph beginning line 130: In the references to Phan et al. papers, please emphasize more what was done with dynamical downscaling with RegCM2, as this prior work would seem to set the stage for the present paper. |  |
| 6) | Paragraph beginning on line 135: The authors should specifically define here what they mean by 'seasonal.' Just saying something like considering one to six month forecasts is probably fine. |  |
| 7) | Paragraph beginning line 146: Maybe instead of listing out the specific products, just give a more general topical summary of the data types. For example, something along the lines: "This study utilizes data from global seasonal forecast models, atmospheric reanalyses, and gridded and gauge precipitation products." Also, within the data description section, AMS publication standards I believe require the actual web links to the data, so those should probably be included in a revision. But the WAF editorial staff can advise more specifically to this point, should the paper be eventually accepted. |  |
| 8) | Paragraph beginning line 151: The authors should specifically mention the CFS reforecast product (it is part of the larger product of NMME reforecasts). |  |
| 9) | Paragraph beginning line 156: It is an important, though perhaps minor, point that the four times daily temporal resolution of CFSv2 reforecast information at standard pressure levels is sufficient to specify the boundary forcing for a RCM. These data have only become freely available in the last couple of years, so it would have been impossible to attempt a CFS reforecast downscaling exercise prior to that. |  |
| 10) | Paragraph beginning line 167: However, per my previous comment, the authors are only using the CFSv2 operational forecast for the actual downscaling exercise to generate three years of output data. So I'm somewhat dubious about making sweeping conclusions on the value added of the downscaling that the authors do towards the end of the paper based on such a limited sample size. Again, it is actually possible to downscale the reforecast product exactly like the operational forecast, so why not do that to generate more years of downscaled forecasts (similar to aforementioned MRED project). If it is impossible to generate more downscaled forecast data, then I suggest look at some of the specific precipitation anomaly patterns in the individual years in relation to atmospheric circulation and large-scale climate forcing. So for example, if sampling a year with a strong ENSO signal, is that signal effectively realized in the seasonal forecast for that year? |  |
| 11) | Paragraph beginning line 181: The authors should explore the use of more spatially resolved, gridded precipitation datasets that incorporate satellite information (from TRMM, GPM systems). A good product developed specifically for data sparse areas is CHIRPS.  http://chg.geog.ucsb.edu/data/chirps/  If the authors can verify that CHIRPS precipitation yields better comparison with Vietnam gauge data, I suggest maybe use that one in addition to or even to replace GPCC. In any case, something that frustrated me in reviewing this paper was the lack of consistency in use of observational products in comparing to the various model data. In some figures, the gauge data are more the emphasis, and in other GPCC. Whatever precipitation data products are used, the authors need to be more diligent in comparing all these data to their model products in a uniform way. |  |
| 12) | Paragraph beginning line 186: On Fig. 1, the authors specify regions R1 to R7 for precipitation data and then use this classification as a basis for organizing all of their subsequent data analyses. The basis for how these regions were objectively determined must be stated. For example, are these regions reflective of dominant annual modes of precipitation (e.g. through an EOF analysis) or are they bounded by distinct physiogeographic features (mountains, rivers, etc.). If this information is summarized in some other documented material, please cite it. |  |
| 13) | Paragraph beginning line 218: In the description of RegCM2 simulations, it is important to state whether spectral nudging was employed or not. Spectral nudging basically employs an interior nudging term preferentially at larger spatial scales and more away from the model surface. It helps the model retain variability in the large scale circulation. Whether it was employed or not may help to explain some of the precipitation patterns in Fig. 13. |  |
| 14) | Paragraph beginning line 254: The authors are proposing an entirely new metric to explain 'added value' in the RCM simulations. But the physical basis for this metric is not very well explained. I do not understand what the meaning of the denominator is. It is then frustrating because most of the results for the regional modeling are framed in terms of the AV metric. For purposes of this paper, I strongly recommend the authors to not use the AV metric. Just stick to standard forecast verification metrics that are well referenced and used in operational practices. Better model performance will be defined quite well enough by the differences in HSS and PSS—and that is more what they should show. |  |
| 15) | Results, Section a(1): The comparison maps of circulation patterns looks reasonable enough, but the authors may want to more quantitatively establish the maps of CFS\_Rfc and not statistically significantly different than CFSR. |  |
| 16) | Paragraph beginning line 292: In Fig. 4, the authors needs to also include results from CFSR. For the 'R' regions, they need to make sure to reference back to Fig. 1 so it is clear to the reader.  Line 294: Eliminate the word 'predicting' here. There is no specific deterministic prediction involved in evaluating the climatological performance of CFS\_Rfc. |  |
| 17) | Paragraph beginning line 302: I do not have an a priori sense of that values for RME and RMAE should be 'small' or 'big' and I am not sure I understand what is meant by these qualitative descriptions. Is there a way the authors could just show the statistically significant differences and emphasize those? Similar to my previous comment, CFSR again needs to be included in the analyses on Figs. 5 and 6, then it can be evaluated whether or not differences are arising because of the CFSv2 modeling system or the deterministic S2S forecast. |  |
| 18) | Paragraph beginning line 316: Are these correlation coefficients statistically significant? Field significant? When are patterns field significant and how does this change with time? Maybe show domain averaged correlation values as a table. |  |
| 20) | Line 329: Never use the word 'significantly' unless there is a quantitative assessment of statistical significance. It is not there. |  |
| 21) | Paragraph beginning Line 334: On Fig. 8 box plots, please explain more what the box and whiskers mean. It looks like it's based on quantiles. Is there a better way to covey the important message that CFS\_Rfc is underestimating variability? Perhaps show this information as a separate figure or table. At the end of this paragraph the authors state '…CFS model which cannot predict the local climate in detail.' This statement needs to be much more clearly explained and related to physical mechanisms of precipitation generation, per my earlier suggestion of modifying the introduction. |  |
| 22) | Paragraph beginning line 348: Can the authors specifically indicate results above the 0.33 threshold and just scale the y-axis above that value? I just drew a horizontal line myself on the figure for clarity. |  |
| 23) | Paragraph beginning line 363: Just a comment, that the results for CFS\_Ope really do not look much different than Fig. 4, except that the scaling has changed. It's basically the same modeling system, whether for reforecasts or operational forecasts, so that is not so surprising. |  |
| 24) | Paragraph beginning line 378: On Fig. 13, should show difference with RCM vs. CFS\_Ope, with some domain averaged statistics. Per one of my overarching comments, there are simply not enough downscaled model years to effectively evaluate climatological performance of a RegCM4-based seasonal forecast modeling system. |  |
| 25) | Paragraph beginning line 388: Again, what does this all mean in terms of statistics and statistical significance? Also, should especially emphasize broadening of IQR with regional model. |  |
| 26) | Paragraph beginning line 395: Authors should consider quantifying some of information in this paragraph in the form of a table. |  |
| 27) | Paragraph beginning line 407: Should have some domain averaged metrics, repeat with GPCC data. Maybe select a few representative stations and look at the precipitation distribution (histograms) to see how shapes of precipitation distributions represented. Then perform goodness of fit tests of model simulated versus observed histograms. |  |
| 28) | The phase 'capturing exclusive local weather features' needs to be better physically explained. |  |
| 29) | On analyses in this paragraph, the authors should reference back to Fig. 6 and show the anomaly correlation. From what I can tell, the regional model appears to add skill where it already exists in the driving CFSv2 global model. Here is where I think the authors should show performance of some of the operational forecast metrics, to know if the forecasts are skillful or not. |  |
| 30) | Paragraph beginning line 432: A more spatially resolved precipitation product like CHIRPS might be very helpful in assessing the regional differences the authors are trying to emphasize with the RegCM4 results. |  |
| 31) | Paragraph beginning line 454: The authors needs to do a better job of referencing back all of their findings to the potential utility for improved seasonal forecasting. What is the threshold for practical usability of this information, in terms of anomaly correlation, skill score, etc? Does downscaling actually help make for more useful seasonal forecasts that could be eventually incorporated into something like the ASEAN Climate outlook? In this regard, it is fine to maybe put some information here about a viable research to operations path forward. |  |