Biogeosciences

Discussions

Open Access

Interactive comment on "What is the importance of climate model bias when projecting the impacts of climate change on land surface processes?" *by* M. Liu et al.

M. Liu et al.

mingliang.liu@wsu.edu

Received and published: 14 March 2014

Referee #2 (bgc-10-C7719): In their paper, Liu et al. present a series of climate change impact experiments wherein bias corrected and non-bias corrected regional climate model data are used to drive multiple models as might be used in climate change impact studies. This is a moderately useful study as people often apply bias correction and or downscaling without much thought; however, there are some major areas that could be improved in the current study.

I have two more serious concerns with the present study. First, their statistical tests of significance are probably inappropriate, and second, their interpretation of the results

over HJ Andrews ignores the fact that these results come from a different model.

Reply: We have addressed these two issues in our revision. Please see detailed responses below.

On the statistical tests, the authors apply a simple student's t-test to test the significance of the difference in the change signal between the BC and NBC results. However, by taking the spatial field of differences as a population, their application of this method is treating real spatial variability as random, independent samples. This is not a fair assumption and likely results in a large number of their tests appearing nonsignificant due to the large (and non-random) spatial variability.

Reply: We agree with this point. However, in this study we emphasize the region wide responses of land surface and ecosystems to BC and NBC climate data as well as spatial variations over the domain by analyzing each grid cell. For regional wide estimation, we evaluated the significance of BC effects on total water availability, yield, SWE, BVOC, NO, and DIN export such that taking the PNW as a whole in the t-test is appropriate. For spatial variations, first we quantified, for each grid cell, the projected changes and differences in project changes driven by BC and NBC (Figure 5). Next, we added results from spatial t-test for each grid cell on major studied variables (Figure 4).

A more appropriate test might include a field significance component where each grid cell (or each small/similar region) is tested independently and then the significance of the differences across the domain are evaluated using a Field test or Walker's test (e.g. Folland and Anderson 2002; Livezey and Chen 1983; von Storch, H., 1982; Wilks, 2006). This is important because the authors are claiming that bias correction does not necessarily matter for some of their results, and I would claim that it would matter for more of their results if they used a more appropriate statistical test. I am not sure that the tests I am suggesting here are the perfect ones and would love to see the authors find better tests if they do not like these methods.

Folland, C., and C. Anderson, 2002: Estimating changing ex- tremes using empirical ranking methods. J. Climate, 15, 2954– 2960. Livezey, R. E., and W. Y. Chen, 1983: Statistical field significance and its determination by Monte Carlo techniques. Mon. Wea. Rev., 111, 46–59. von Storch, H., 1982: A remark on Chervin-Schneider's algorithm to test significance of climate experiments with GCM's. J. Atmos. Sci., 39, 187–189. Wilks, D. S., 2006: On "Field Significance" and the False Discovery Rate. J. Appl. Meteor. Climatol., 45, 1181–1189.

Reply: Thanks for the valuable information. As we argued earlier and in our discussion, we emphasize both the regional and local effects of BC. The Walker's test may provide information for global null hypothesis, but it may over-emphasize local effects (i.e. the least p-value in t-test) that can easily reject the global null hypothesis (Wilks, 2006). According to the Walker's test (Wilks, 2006; Livezey and Chen, 1983; von Storch, 1982), our results indicate that BC do have significant differences in responses to climate change over the PNW, i.e. the global null hypothesis is rejected, because in certain locations BC and NBC climate produced significant differences in responses (Figure 4). However, the rejection of global null hypothesis cannot provide further information on which locations are most sensitivities to BC treatment. Therefore, in this revision we reported the percent of grid cell with significant differences (p < 0.05) on some of the variables from both the larger domain (i.e. the PNW) and smaller domain (i.e. the HJ Andrews). These new information have been included in this revision and a new section (i.e. section 2.5) was added to describe the method for quantifying the significance over regional and grid cell bases.

On their interpretation of results in HJ Andrews, they state that the ET Climate Change estimates in HJ Andrews are affected by bias correction, while ET in the domain as a whole is unaffected. However, the larger domain ET is simulated using VIC at a lower spatial resolution, while in the HJ Andrews subdomain they use RHESSys and what appears to be a higher effective resolution in their model domain. These are two very different models with different embedded (or missing) ET, hydrology, and ecological

equations/assumptions. As a result, I don't think they can say anything specific about HJ Andrews from the results they have presented so far. A more important point may be that the effect of BC on climate change impact may be different for different hydrologic models.

Reply: We totally agree to this point that different model may have different modeled results because of their specific parameterization and simplifications on difference processes and spatial scales. We agree that developing a multi-scale and wholly-coupled regional earth system model (e.g. Adam et al., 2014) is critical to identify key processes and feedbacks in climate change impact studies. This study, however, is not intending to demonstrate the differences and uncertainties in BC impacts due to variations in model structure. Rather, we reported the BC impacts from different scales (e.g. regional as for entire PNW and small watershed as in HJ-Andrew) as case scenarios. Even they are not comparable (we didn't do that), as discussed in this paper, these findings from different scales still provided valuable information for evaluating BC impacts.

In either case, it would be appropriate to compare the results from RHESSys to the results from VIC output just over HJ Andrews. If the results are identical between models, then they can make the statement they made initially, that regions such as HJ Andrews are affected differently than the domain as a whole. If the results differ, then they are left with the conclusion that the choice of hydrologic model matters. This will also be affected by the fact that the two models have different methods of calculating driving meteorological variables from the temperature and precipitation data supplied. Ideally they should also run RHESSys with the MTCLIM forcing generated by VIC and vice versa. This may be beyond the scope of the current paper, but a short note to point out the met forcing could be an issue would be useful to include. Similarly, were both of these models locally calibrated? I'm fairly certain that RHESSys has been calibrated for HJ Andrews at some point, and VIC has been calibrated for the CRB, were those calibrated parameters used here? Are their results sensitive to model parameters?

Reply: Thanks to these comments. As to the much coarser resolution of VIC (∼6 km) than RHESSys (30 meter), the direct comparison over this small watershed (∼2 VIC cells) is not appropriate. The reviewer was correct in that has the right point that VIC was calibrated over much larger sub-basins over the Columbia River Basin while RHESSys has been thoroughly calibrated over the HJ-Andrews watershed with observed streamflow. We agree that further comparisons between RHESSys and VIC over large domain are meaningful but is beyond the scope of this paper.

While those strike me as the two most important problems with the current manuscript, there are other major areas that need to be looked at. First, they note that BCSD is used to generate forcing data from e.g. CMIP5 data. However, the dataset they cite is producing monthly data, not daily data, and it is based off of the GCM precipitation data instead of a moderate resolution regional climate model as in their case. They should include some discussion of the fact that they are only applying relatively modest bias correction after a detailed dynamic downscaling, while statistically downscaling GCM output requires much larger bias corrections, and as a result, it is more likely that climate change signals would be affected by the bias correction / statistical downscaling step in that case.

Reply: We agree that there is a big difference between the current RCM study and BCSD applied to monthly GCM data. While results of this study may be generally indicative of the value of bias correction, the differences are sufficiently large that we have removed that comparison.

Along those lines, I was surprised that there was almost no discussion of the WRF modeling that they performed. This is not an inconsequential component and substantially affects the interpretation of the implication of their results for other studies. At a minimum they should report the physics options used to run the WRF model and mention that the WRF domain was (presumably) that of figure 1. In which case, what did they use for SSTs, are these taken from the driving GCM without bias correction? Reply: A more detailed description of the WRF model simulations have been added to

C9248

section 2.3 to directly address these concerns.

The paper as a whole could also use additional discussion and interpretation to make it more useful to a broad audience. For example, their analysis of the relative importance of temperature and precipitation is useful, but more discussion of why they see the results that they see would be useful. For example, they note that precipitation has a greater affect on SWE, but this is probably only true in parts of the domain (this is one of the problems with averaging over a large and heterogeneous domain).

In particular, the lower elevation bound of the seasonal snowpack is probably substantially affected by small changes in temperature because of the temperature threshold at 0C. These sorts of thresholds are critical in understanding how the results from this study may affect other variables or locations. If the system response is highly non-linear or crosses some threshold, than BC will matter, if it is linear than BC may not matter.

Reply: Very good points. We have added more detailed analyses and discussions in the revision. For example in Section 3.2.1.3: This simulation result indicates that under hotter and dryer conditions (BC data), the continuous warming may have fewer relative impacts on snowpack accumulation than under cooler and wetter background conditions (NBC data), particularly over high mountainous areas (Fig. 5-e). The SWE of a cooler baseline (NBC data) has more potential to be affected by additional warming as compared to a warmer baseline (BC data) subject to the same additional warming (Fig. 5-e), especially if the additional warming moves grid cells from below freezing temperatures to above freezing temperatures. Also, a larger expanse of negligible snowpack areas in the BC scenario as compared to the NBC scenario can result in larger areas categorized as having no change in SWE for the BC scenario.

Similarly, additional discussion of the importance of coupled simulations would be useful. The authors premise and primary conclusion is that there are tradeoffs between using BC data offline and NBC data in a coupled framework. However, the authors do

C9249

not seem to have presented any results or substantial discussion showing the utility of a coupled simulations. I would assume the primary utility would be for processes that feedback to the atmosphere and affect the local weather patterns, for example, ET and possibly VOCs that form aerosols (if the atmospheric model actually accounts for aerosols explicitly), as both will affect precipitation, but I do not believe this is ever described in the paper.

Reply: In the revised manuscript, we have added discussions of important feedbacks processes that can only be captured in coupled models.

Finally, the figures as they are presented in the print and discussion versions of this paper are illegible when printed (they are far too small). I suspect this may be due to formatting potentially applied by the publisher after the authors submitted it, regardless it needs to be fixed at some step in the process.

Reply: Figure 5 has been divided into two panels.

Even after fixing the size issue, figure 6 might be better with different colors for some of the lines, the subtle shades of grey are hard to distinguish. There are also many minor grammatical errors in the paper too numerous to document here, hopefully the authors can get a proof-reader to fix these errors. Reply: This figure (i.e. figure 6) has been converted into color diagrams. The revised paper has been carefully edited by native English speakers.

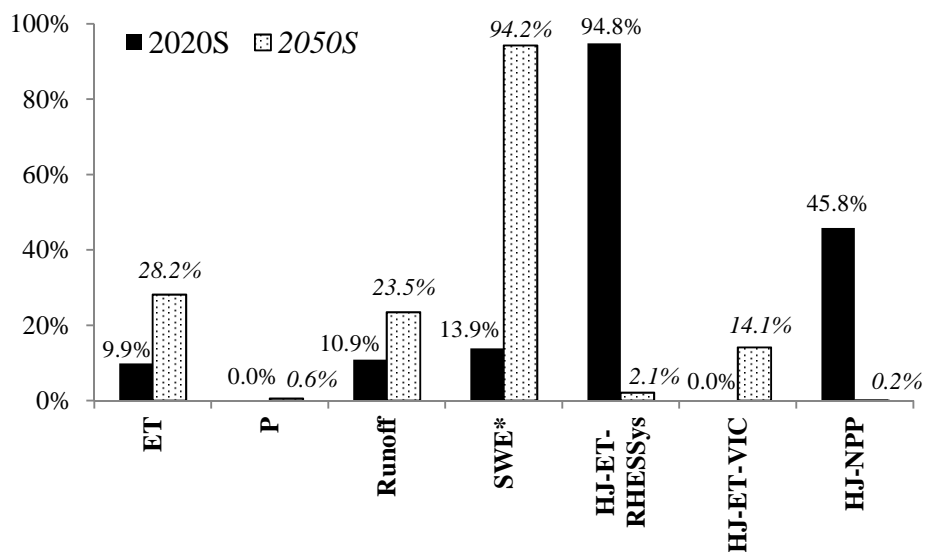Interactive comment on Biogeosciences Discuss., 10, 17145, 2013.

Fig. 1. Figure 4 Percentage of grid cells with significant differences (p < 0.05) between the projected changes using BC and NBC climate data.

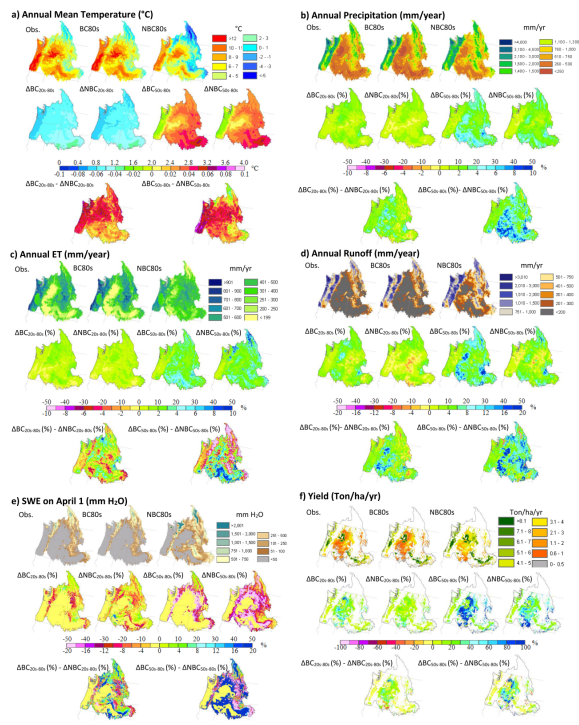a) Annual Mean Temperature (°C)
b) Annual Precipitation (mm/year)
c) Annual ET (mm/year)
d) Annual Runoff (mm/year)
e) SWE on April 1 (mm H₂O)
f) Yield (Ton/ha/yr)

Fig. 2. Figure 5

C9252