Krogh et al. present an interesting analysis comparing climate change sensitivity impacts on streamflow in the western United States between space for time substitution (which they term STS) and more traditional modeling techniques, where they focus on NoahMP-WRF pseudo-global-warming simulations (termed PGW). They introduce a new metric based on diurnal fluctuations in streamflow that are lag-correlated with solar radiation, and then calculate the day of year when 20% of all days with well-correlated diurnal fluctuations have passed. I like the idea and the premise of the paper, but I feel that major revisions are necessary to disentangle all the possible ways that errors in the analysis could lead to misconceptions in the results. I also feel that the number of acronyms and metrics in the paper (STS, PGW, DOS_20, etc.) make the written text hard to follow, and I strongly recommend that the authors minimize their use of acronyms, perhaps provide a table of acronyms and metrics, and overall work to increase clarity. I have organized my comments into requests for Major and Minor revisions below. The authors are welcome to contact me directly if they have questions: Jessica Lundquist, jdlund@uw.edu

Major:

1) You need a clear analysis of how well your diurnal-cycle-correlation metric works across a range of streams.

1a. line 199-200 “more variable mean annual autocorrelation that ranges between roughly 0.1 and 0.6, with a mean value around 0.4” —need to explain what different mean annual autocorrelations refer to. These numbers are really new to most people. It would be helpful to tie this metric to the examples in Figure 1, as well as a discussion of rain vs. snow — a lot of the “snowmelt days” marked with purple circles in Figure 1 look like rain storms to me. The South Fork of the Tolt mostly gets rain, but also rain on snow. How do diurnal cycles that are identified but aren’t really snow melt impact your results?

1b. As an alternate approach to when snowmelt is significant, you could look at the power spectra of your time series. See Figure 6 in Lundquist and Cayan 2002. The days with a sharp increase in power at the once per day cycle indicate snowmelt, whereas rain
exhibits a much more red spectra. I know that power spectra are commonly used by oceanographers and not hydrologists, so your method is likely easier to understand, but it would be nice to have an independent method to check.

1c. In particular, I recommend clearer discussion about the strengths and weaknesses of this approach. It will miss rain-on-snow (signal dominated by rain), as well as early melt into dry soil (no streamflow response). It may also misclassify rain with a diurnal structure to it as snowmelt. Therefore (and you allude to this multiple times in the manuscript but should make it clearer), the method is best at detecting melt in non-rainy locations with fairly-saturated soils. With that in mind, which of your basins do you trust the signal the most.

1d. Section 3.1 explains how well the DOS_20 is related to simpler magnitude metrics (DOQ_25 and DOQ_50) but doesn’t really justify why the DOS_20 is helpful beyond those metrics — can you better explain what we gain by doing this extra analysis. This section also identifies some rain-dominated rivers wherein these metrics appear less correlated. Is this because the method breaks down? Or can we learn important information from this change in relationship?

2) You need to more explicitly discuss the difference between a stream's climate sensitivity of snowfall changing to rainfall vs. a climate sensitivity of earlier snowmelt.

2a. Many of the earlier papers on streamflow sensitivity to climate change highlighted basins in the transitional rain-snow zone as being most sensitive because snowfall shifts to rainfall. From my own experience, the diurnal cycle in streamflow is particularly hard to detect in these basins because rain-induced runoff is such a larger signal than snow-induced runoff, especially when both happen more or less at the same time. Therefore, I imagine that your snowmelt index uniquely does not work well in these basins (e.g., the Tolt example in your paper, or the NF American River example in Lundquist and Cayan 2002 Fig. 6). I could imagine that for these basins, you could even get DOS_20 moving later in the season with warming if early season events are all rain and only a later, non-rainy period exhibits snowmelt.

2b. I imagine that including rain-on-snow or rain-dominated basins would bias your correlations with humidity because these tend to be more humid basins but also may have spurious results.

2c. I encourage the authors to think about rainfall vs snowfall and snowmelt sensitivities separately and to decide if they want to address both in this paper or only focus on the latter. Then, be very clear about this decision in the paper discussion.

3) You need to more clearly evaluate how well your NoahMP-WRF model set up is simulating streamflow timing in the current climate before examining the results of its climate sensitivity.

3a. It appears that you have a biased simulation of NoahMP-WRF — if the historic runoff date is off by 50 days (see line 260), the model is either simulating too much rain and too little snow or melting snow way too early. It’s hard to draw conclusions on sensitivity when using a biased model. Of course, if the model has less snow than the real world, it will be less sensitive to that snow disappearing. The paper would be much more meaningful if you included some evaluation of your NoahMP-WRF simulations — how do they compare to baseline observations and to other models run over the domain (similar
western US climate-change papers).

3b. Also, if the NoahMP-WF simulations perform better in certain regions (if I’m correct, these were only carefully vetted for Colorado), you may also want to focus your analysis on those regions separately. Do you get closer agreement in areas where the model represents snow processes more accurately? Might a check for space-for-time sensitivity against model sensitivity be a good check for model fidelity?

4. Discussion should be better streamlined and organized. This may be a good place to address major comments 1-3 above.

Minor:

Abstract: 1st sentence, ”may cause“ — I think the literature is pretty conclusive that warming does cause snow to melt earlier. Abstract should define what you mean by the 20th percentile of snowmelt days — this is meaningless to someone only reading the abstract. What do you mean by colder places are more sensitive than warmer places? In what way? Earlier snowmelt? If there’s no snow, of course it wouldn’t be sensitive to that.

Line 120: “DAYMET dataset (daymet.ornl.gov), which in turn is based on ground observations” — it’s interpolated from existing ground observations — worth specifying as sometimes this is far from truth.

lines 202-205 The percent of streamflow volume by a certain date vs temperature has been well established in the early literature (Stewart et al. 2005). Also see Lundquist et al. 2004 for a review of different ways to define the “spring onset” from snow pillows and from a hydrograph: https://doi.org/10.1175/1525-7541(2004)005<0327:SOITSN>2.0.CO;2

line 215: Yes, these sites are low elevation, receiving primarily rain, and I think your methodology is identifying rain events as having a diurnal cycle.

line 259: “greatly underestimated“ — I think you mean than it’s modeled as earlier than observed, right? Underestimated makes me think that the magnitude of the streamflow is too low.