First we would like to thank referee John Pitlick for reviewing our manuscript. Original comments by the referee are in normal text and our responses are in bold.

This paper presents a relatively thorough analysis of historical trends in precipitation and runoff in four large river basins in the Upper Midwest, United States. The main question addressed in the paper is whether the observed increases in runoff in this region are due to increases in precipitation, or due to increases in the efficiency of drainage caused by expansion of agricultural drainage networks and tile drainage systems. The instrumental records of streamflow and precipitation in this region are quite good, with continuous records of discharge and precipitation dating back to the early 20th century at hundreds of locations. The data sets that allow for historical reconstruction of land cover are also quite good. A comparative analysis of hydrologic trends in four river basins with different land use/land cover is used to assess the joint roles of precipitation and artificial drainage on increased streamflow.

The results of the analysis suggest that in the three basins that have experienced the greatest change in land cover, increases in annual streamflow are disproportionately higher than changes in annual precipitation; in the fourth basin where changes in land cover have been minimal, there has been no change in discharge over time. Taken together, the results indicate that modifications to land cover and land use in the Upper Midwest have had widespread impacts on the hydrology of the region. This topic has generated much discussion over the last 10 years, and the present paper will certainly help guide future efforts to understand the effects of land cover change.

I have several comments and suggestions for the authors to consider:

1. I suggest adding a brief discussion of the rationale for using PRISM 'data' to evaluate trends in precipitation rather than station data. Presumably, there are dozens of weather stations in each basin, and it wouldn't be too hard to utilize these observations, unless there are significant gaps in the records. Weather station data would also be useful for evaluating changes in temperature, which might have as much influence on ET as changes in land cover. It's not necessary to re-do the analysis, but please consider adding a comment.

Response: This analysis considers monthly and annual precipitation. We rely on PRISM precipitation rasters because they provide continuous spatial and temporal coverage of our study areas. In the Central USA mean absolute errors between PRISM monthly grids and station data are on the order of 3-6 mm/month, while monthly precipitation totals are on average an order of magnitude greater (Daly et al., 2008). Additionally, the ET_a data used in this study were generated by Livneh et al., 2013 (L13) using gridded NCDC COOP station observations of precipitation and daily minimum and maximum temperatures. The precipitation data used by L13 agree well with PRISM, page 11, lines 5-9 (Figure S3). Thus, we feel that the PRISM data are the best available dataset to apply for our study, they are reproducible and defendable, and we do not believe that we could improve on these precipitation estimates using our own algorithm.

2. The 17% adjustment in ET seems unnecessary and it just makes for a more complicated discussion of the results. Livneh (2013) suggested this value based on observations from four Ameriflux sites scattered across the entire US. One of those sites is in the study area. That same site is referenced in Table 2 (Brookings, SD), but here the discrepancy between modeled and observed values is _30%. At the three other sites listed in Table 2, the discrepancy is opposite in sign (+) and perhaps as little as 5%. Estimates of ET are used in this paper primarily

to estimate the storage term, dS/dt, and the same conclusions about storage can be made without a 17% adjustment to a model-derived number.

Response: The referee's point that the same conclusions about the storage term can be made without adjusting the model-derived ET_a is valid. It simply represents our effort to err on the conservative side with the water budget. We briefly discuss our rationale for reducing JJA ET_a by 17% in the methods (page 13, lines 27-31; page 14, lines 1-10) and supplement, and plan to point to the fact that adjusting ET_a does not alter the analysis and main results of our study. The water budget conclusions are based on the relative differences in storage not the magnitude of the storage terms individually (page 28, lines 14-20). The 17% reduction only affects the magnitude of the storage term. In light of the reviewer's comments, we will revise the text related to this point in the results and discussion section.

3. The paper could be shorter. Some of the figures and associated text could be eliminated, e.g. Figure 5, Figure 7, Figure 9. Also, consider whether it is really necessary to consider 7 different flow metrics, e.g. 7-day low flow winter, high flow days versus extreme flow days (what is the difference between high flow days and extreme flow days?).

Response: We have done our best in the revised manuscript to reduce its overall length and improve the flow. We believe that removing the figures suggested would take away some of our main messages, but we will consider removing 1-2 figures as part of the revision. Regarding flow metrics, more complete definitions of the seven different flow metrics can be found in Novotny and Stefan, 2007. The number of high and extreme flow days refers to the number of days in a given year that are one and two standard deviations above the mean of the entire record, respectively. As the tails of distributions may contain useful information and it does not take up much additional text or space to retain all seven of the streamflow metrics, we would like to retain the analysis as is.

4. Minor points:

- p. 6, line 20, numbers don't seem to add up. - p. 8, how have reservoirs in the Chippewa River basin affected peak flows? - p. 13, line 3, remind the reader what LCT is; line 27, perhaps say that the storage term is sensitive to estimates of ET, rather than saying "we don't have high confidence: : :" - p. 15, line 2, eliminate 'grossly' - p. 18, line 8, changes in the IRB do not appear dramatic / systematic to me. - p. 19, line 8, where are the results discussed here, and on the next page, shown? - p. 20, Figure 5 is very difficult to interpret; same with Figure 7 - p. 24, Figure 8: This figure is very useful, could appear earlier - pp. 26-27 and Table 4. First, Table 4 is confusing with the different breakpoints and adjustments in ET. Is it really necessary to show results using the LCT breakpoints, since they are essentially the same as results for the 1974/1975 breakpoint? In addition the LCT breakpoints result in uneven record lengths, and in one case (CRB), the record length is only two years. I suggest using only the 1974/1975 breakpoint since that produces roughly equal intervals of time. Finally, p. 27, reverse the order of discussion of IRB and RRB (or reverse the values in the table). - p. 42, Table 1, change references to Livneh from L14 to L13.

Response: We would like to thank the referee for the suggested comments. -p.6, line 20: Total since 1999 reflects the cumulative total miles of tile installed between 1999-2015. -p.8: Reservoirs on the Chippewa River have reduced peak as well as intermediate and base flows (Barnes, 1997). –p.18, line 8: Although the normalized flow metrics in Figure 4 do not appear to be as systematic or dramatic as the MRB and RRB, five of seven flow metrics have increased significantly (all but peak daily flows). Since dams are most likely to affect peak daily flows, it is not surprising that no significant changes

have occurred for these flows and that overall trends in the IRB are less obvious than in the MRB or RRB. –p. 19, line 8: the results are found in table S1 as noted on page 20, line 15. This information should be stated at the beginning of the section and we have made that change in the revised manuscript. We disagree with the referees comments regarding Figures 5 and 7 and argue that they further demonstrate the similarities and differences between precipitation and streamflow changes in each of the basins. –We have used different breakpoints for comparing records to demonstrate that the results of the statistical tests are robust to breakpoint changes. We will consider removing the LCT results from Table 4 or moving them to the supplement.

References

Barnes, W. J.: Vegetation Dynamics on the Floodplain of the Lower Chippewa River in Wisconsin, J. Torrey Bot. Soc., 124(2), 189–197 [online] Available from: http://www.jstor.org/stable/pdf/2996584.pdf?_=1461109555076 (Accessed 20 April 2016), 1997.

Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G. H., Curtis, J. and Pasteris, P. P.: Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States, Int. J. Climatol., 28(15), 2031–2064, doi:10.1002/joc.1688, 2008.

Livneh, B., Rosenberg, E. A., Lin, C., Nijssen, B., Mishra, V., Andreadis, K. M., Maurer, E. P. and Lettenmaier, D. P.: A Long-Term Hydrologically Based Dataset of Land Surface Fluxes and States for the Conterminous United States: Update and Extensions, J. Clim., 26, 9384–9392, 2013.

Novotny, E. V. and Stefan, H. G.: Stream flow in Minnesota: Indicator of climate change, J. Hydrol., 334(3–4), 319–333, doi:10.1016/j.jhydrol.2006.10.011, 2007.