Hydrol. Earth Syst. Sci. Discuss., 12, C2650–C2664, 2015
www.hydrol-earth-syst-sci-discuss.net/12/C2650/2015/

Hydrology and Earth System Sciences Discussions

HESSD

12, C2650–C2664, 2015

Interactive Comment

Interactive comment on "Nonstationarity of low flows and their timing in the eastern United States" *by* S. Sadri et al.

S. Sadri et al.

justin@princeton.edu

Received and published: 20 July 2015

We thank the referee for the very useful comments and suggestions. We have repeated the comments below and our responses are indicated by "Response"

Anonymous Referee #1 This article is well written and provides valuable insight into changes in low flow magnitude and timing in the eastern U.S. It is a more thorough look at historical low flow changes than I have seen previously. The introduction is very well done and referenced. I have two major concerns about the article that I describe below in detail; one concerns the classification of streamflow gauges and one concerns the decomposition algorithm that was used for the analyses. I have also listed minor concerns below.

Classification of flows: I consulted colleagues at USGS and there are no site notes available for low flows. The codes used for this article are presumably from USGS peak streamflow metadata, which are available online. The codes are relevant to the annual instantaneous peak flows and may or may not be applicable to low flows. For example, low flows are likely more sensitive to streamflow regulation than peak flows.

Response: The lack of site notes for low flows is one of the reasons we have developed the proposed statistical methods for identifying stations with low flow time series that are potentially free of regulation or changes in measurement characteristics. We have made this clear at the end of section 2.2 where we state that site notes are not available for low flows, only peak flows, and that low flows may be more sensitive to regulation: "It should be noted that the USGS flags are developed for peak flows and while it is uncertain whether these are directly applicable to low flows, it is likely that low flows are more sensitive to regulation."

Also, some of the codes used by the authors are not meaningful for determining anthropogenic influence. The "base discharge" is a level set to allow 3-4 peaks per year on average to exceed this level. Instantaneous peaks above this level are then recorded.

Response: See response to next comment.

A "change of base discharge" does not indicate any change in actual flows recorded or any anthropogenic change in watersheds above a gauge. The "change of gauge datum" also does not indicate any change in flows or anthropogenic influence. It indicates only that the arbitrary zero gauge height for the rating curve has been changed, normally because of changes in the gauge control point on a river (the riffle or channel section that controls the relation between river height and flow at a gauge). So for example in the abstract, the statement that "about a third of the sites with a decreasing trend were associated with a change of gauge datum" is not a meaningful statement.

Response: We understand that not all codes are related to a change in actual flows. The purpose of the proposed method is to identify useable time series for understanding natural variations in low flows, rather than specifically attributing changes to anthropogenic influences, changes in measurement method or something else. This also responds to other reviewer comments that query the assumption that any detected step changes are attributable to management. Rather, we are looking for useable time series. We have edited the introduction here to be clear on this point (and in response to other referee comments): “The goal of this paper is to examine non-stationarity in low flow generation across the eastern U.S. and attempt to systematically identify time series that are potentially free of the effects of human intervention and examine these in terms of the impact of climate variability and change.” and also in the conclusions (section 5.2) we have made some slight changes in the wording to note the distinction between anthropogenic influences and changes in measurement characteristics, and to note that changes in the gauge datum, whilst not necessarily changing the flows, may have had an effect on the time series homogeneity: “Comparison with USGS site flags indicates that the majority of sites with identified step changes and increasing trends are noted to be regulated in some way, and some are documented as having a change of stream gauge datum or undergone urbanization. For sites with decreasing trends, about one third have USGS flags and these tend to be for a change in gauge datum height; a similar proportion of sites with no trend are also flagged for a change in gauge datum, and so it is unclear whether this type of change has influenced the low flow time series.” For the line in the abstract that refers to the impact of the change of datum we have changed this to note that that these sites are less likely to be flagged, rather than making any statement about attribution. “Sites with decreasing or no trend are less likely to have documented influences on flows or changes in measurement characteristics.”. We have added some discussion to section 2.2 “Streamflow data” on the meaning of “change in base discharge” and “change of gauge datum” and how it relates to the identification of anthropogenic influences: “Based on the USGS site notes (available on the NWIS website), we identified sites that are flagged as: regulated, partially regulated, flow below the rating curve limit, dam failure, affected by urbanization, change of base discharge, and change of gauge datum. It should be

HESSD
12, C2650–C2664, 2015

noted that the USGS flags are developed for instantaneous peak flows and while it is uncertain whether these are directly applicable to low flows, it is likely that low flows are more sensitive to regulation. Some of the flags are unrelated to anthropogenic influences, such as "change of base discharge", which is a level above which peak flows are recorded, or "change of gauge datum", which is the arbitrary zero gauge height for the rating curve. Figure 1d shows the location, flag type, and the number of the sites under each flag. Almost half of the sites have no flag and these are located throughout the domain. The majority of regulated or partially regulated sites are concentrated in the northeast, but this is also where the majority of all sites are located. The sites in the mid-Atlantic states are generally more affected by urbanization or have experienced a change of gauge datum. Overall, 271 sites out of 508 sites are somehow affected in terms of anthropogenic influences or changes in measurement characteristics. In the results section, we show how the results of our statistical methods compare with the USGS site flags."

Statements throughout the article will need to be evaluated and many changed or removed. With my stated concerns with the classification of flows and the decomposition algorithm (see below), it's currently not clear to me whether the classification system provides meaningful insight into the study results.

Response: In addition to the edits discussed in previous comments, we have also edited the statements in Section 3.2 to only report the nature of the flags rather than make statements on association/attribution.

Decomposition algorithm: It's not clear to me why this recursive algorithm is used in the way it is for the Ljung-Box, Mann-Kendall, and Pettitt tests. This algorithm needs to be justified and fully referenced. There is currently one reference to the impact of autocorrelation on the Mann-Kendall test.

Response: The description of the algorithm in the manuscript is not completely consistent with the overall approach and we have updated it to better reflect our intention

HESSD
12, C2650–C2664, 2015

and respond to the reviewer comments. Detailed descriptions and referencing for the tests were removed from an earlier version of the manuscript because of page length concerns and the fact that these tests are standard and prevalent tests in the literature. Nevertheless, we agree that some more details are needed with regard to the use of these tests in the decomposition algorithm and we have updated the description in section 2.3.3. We now account for autocorrelation in the MK and Pettitt tests by pre-whitening the data (previously this was avoided for the MK test by ignoring sites with significant autocorrelation). The first step of the algorithm tests for overall autocorrelation using the Ljung-Box test as before, but we note how this does a good job of identifying potential non-stationarity in terms of abrupt changes. This is because the series of values before the change appear to be autocorrelated relative to the values after the change, and vice-versa. We have confirmed this visually on the time series for all sites. The Ljung-Box test is applied to identify sites with possible step changes before the Pettitt test, because the Pettitt test will identify step changes in time series with gradual trends. Similarly the MK test will identify gradual trends in series with step changes. We then test the autocorrelated sites for step changes using the Pettitt test (pre-whitened). Sites with statistically significant step changes are split into two and the Pettitt test is applied to each sub-series. This continues until no step changes are found or the length of the sub-series is less than 30 years. Series or sub-series with no step change are then tested for trends using the MK test (again with pre-whitened data). The number of categories have been expanded six to distinguish between upward and downward step changes. We then choose a subset of sites with continuous records and no step changes for 1951-2005 to carry out the rest of the analysis in section 4. The description of the methodology has been updated in section 2.3.3: "The three statistical tests (Ljung-Box, Pettitt and Mann-Kendall) were combined into a recursive algorithm to identify non-stationarity in the low flow time series and decompose the series into potentially stationary sub-series. In the first step of the algorithm, a Ljung-Box test with length 20 years was applied to the entire time series of each site, and sites with significant overall autocorrelation (5% significance level) were identified. The Ljung-Box test identifies sites that are non-stationary and is able to identify sites with abrupt changes because the series of values before the change appear to be autocorrelated relative to the values after the change, and vice-versa. This was confirmed by visual inspection of the time series. For the sites with significant overall autocorrelation, we then applied the Pettitt test (5% significance level) to confirm the existence of any step change and identify its timing. The series were pre-whitened to remove lag-1 autocorrelation, following Kumar et al. (2009). It is necessary to identify sites with potential step changes using the Ljung-Box test first because the Pettitt test will identify step changes in time series with gradual trends. Similarly the MK test will identify gradual trends in series with step changes. If a significant change is found by the Pettitt test, the series is split into two parts either side of the step change. Each part is assumed to be a new series at the same location, and if it has a record length of 30 years or more, the decomposition algorithm is applied again. If the length is less than 30 years, the site is removed from further consideration. If a statistically significant step change is not identified, we note that the series is autocorrelated overall. We then applied the Mann-Kendall (MK) test (5% significance level) on the remaining sites to identify statistically significant trends in the data. Again, the series were pre-whitened to remove lag-1 autocorrelation. The sites are assigned categories as follows: 1. Category 1: Non-autocorrelated site with no trend (MK=0); 2. Category 2: Non-autocorrelated site with a statistically significant decreasing trend (MK=-1); 3. Category 3: Non-autocorrelated site with a statistically significant increasing trend (MK=1); 4. Category 4: Autocorrelated site with statistically significant decreasing step change, time series split and the sub-series re-categorized recursively; 5. Category 5: Autocorrelated site with statistically significant increasing step change, time series split and the sub-series re-categorized recursively; 6. Category 6: Autocorrelated site with no step change. "

Concerns/questions that I have with the decomposition algorithm:

(1) Why not correct the Mann-Kendall test for autocorrelation rather than not testing

HESSD

12, C2650–C2664, 2015

sites with autocorrelation. This has been an area of much work. There is the issue of removing trends when removing autocorrelation. Certainly a discussion and justification with references is warranted.

Response: We now pre-whiten the data (flows and timing) before applying the Pettitt and MK tests. This is now mentioned in the methods section (2.5) and follows Kumar et al. (2009). The data are first detrended and then an AR(1) model is fitted to the residuals, which is then removed from the data. The trend is then added back. Comparisons of the test results using the original time series and the pre-whitened time series show a small change in the number of sites with statistically significant values.

(2) Why not test all gauges with the Pettitt test? A significant Mann-Kendall test could actually be due to a step change. This can easily be demonstrated by generating say 30 random normal values about a mean and another 30 about a different mean. There is no trend in each set, but the Mann-Kendall test will indicate a significant trend (if the means are far enough apart) if these two sets are treated as one time series. Why not specify the direction of the step change in the article as was done for Mann-Kendall? This would add important information.

Response: Because the Pettit test will identify a gradual trend as a step change and vice-versa, we use the Ljung-Box test to identify sites with potential step changes, and use the Pettitt test to confirm this and the timing of the change. The description of the method has been updated as noted above. We have updated the step change results to show the direction of the change in the figures.

(3) Why use only test sites with significant autocorrelation for the Pettitt test? In addition to the issue in (2), the Pettitt test is sensitive to autocorrelation just like Mann-Kendall (Serinaldi and Kilsby, 2015, The importance of prewhitening in change point analysis under persistence, Stoch Environ Res Risk Assess; Ferguson and Villarini, 2012, Detecting inhomogeneities in the twentieth century reanalysis over the Central United States, J Geophys Research Atmos).

Response: Again, we now account for autocorrelation by pre-whitening the data. The results do not change very much as shown above.

(4) More information is needed on how the Ljung-Box test was applied. It's not clear to me how many and which lags were tested. Also, does this test address long term persistence, which is an important issue in time series testing (see for example Cohn and Lins, 2005 and several articles by Koutsoyiannis).

Response: The Ljung-Box test is applied over 20 lags as noted in section 2.3.2 (now section 2.5). We do not address long-term persistence with this test. As such it may overestimate the significance of the identified trends if long-term persistence did exist. We note that we are testing for trends in the data, irrespective of whether they are driven by, for example multi-decadal variability, or a structural change due to say climate change. The results therefore are with respect to the given time series and the possible drivers of change within the specific time period. We have added a short comment in the introduction on the potential of long-term persistence as a driver of trends: "We therefore assume that step changes (abrupt and visually obvious) in the time series are indicative of an anthropogenic effect, and that gradual trends reflect a climate effect, which may be due to anthropogenic climate change or long-term persistence (Cohn and Lins, 2005)."

(5) The results in Figure 7 don't follow the stated algorithm. All sites are tested for trend in Panel A, regardless of autocorrelation.

Response: We showed the results with autocorrelation (Figure 7a) to show that the spatial pattern of trends is consistent whether we ignore autocorrelation or take it into account. It may be confusing because this is different from the algorithm, so we have changed this wording in the text and in the caption to make it clear.

Other Comments

(1) "Eastern United States" should be defined when the term is first used in the intro-

HESSD

12, C2650–C2664, 2015

duction.

Response: The region is defined as covering the 20 ecoregions of the eastern US as defined by the USGS (USGS, 2012): “In the eastern United States (defined as the area covering the 20 ecoregions of the eastern US (USGS, 2012)), both direct anthropogenic and climate influences may”

(2) Page 2763, line 18, Low flows in the eastern US are controlled by more than just subsurface flows. Precipitation in low flow seasons is important to low flows in the eastern US because of the regular precipitation during low flow seasons. It’s not clear whether a true base flow from only groundwater input actually happens in the eastern US. If you believe it does, please provide references. Otherwise, please mention the importance of rainfall to maintaining low flows.

Response: This paragraph is a general statement about low flows that describes the overall influence of groundwater on low flow generation. We added a sentence to note that there are other factors, including precipitation: “Low flows are generally controlled by subsurface flows sourced from groundwater that maintain flows during the dry periods of the year, such that low flow volumes are related to the physiological and geological make up of the area. In some regions, where precipitation is significant in the warm season, surface flows also play a role in maintaining low flows.”

(3) Page 2765, line 1, Increased urbanization can also lead to increases in low flows due to water supply and wastewater pipe leakage, direct wastewater discharge, reduced evapotransporation, and importation of water from outside of watersheds that decreases groundwater pumping. See for example Brandes et al. (2005, Base flow trends in urbanizing watersheds of the Delaware River basin, JAWRA).

Response: Agreed. We have added a sentence that notes the potential for increases: “On the other hand, urbanization can lead to increase in low flows because of leakages from water supply and wasterwater pipes, direct wastewater discharge, reduced evapotranspiration, and water imports that can offset groundwater pumping (e.g. Brandes et

HESSD
12, C2650–C2664, 2015

al., 2005).” Brandes, D., Cavallo, G. J. and Nilson, M. L. (2005), Base flow trends in urbanizing watersheds of the delaware river basin. JAWRA Journal of the American Water Resources Association, 41: 1377–1391. doi: 10.1111/j.1752-1688.2005.tb03806.x

(4) Page 2766, line 25, It is misleading to say that none of these studies considered the role of anthropogenic influences. What at least some of the studies did was to remove, to the extent possible, anthropogenic influences by use of minimally altered watersheds. This was done to try to isolate climatic signals rather than direct anthropogenic watershed change signals.

Response: Agreed. Our intention here was to state that these studies did not attempt to identify anthropogenic influences but rather used data from catchments with minimal anthropogenic influences. This is in contrast to our study, which tries to identify anthropogenic influences with the ultimate goal of using the data to identify and understand climatic drivers. We have edited this sentence slightly: “Furthermore, these studies focused on sites that were deemed to have minimal anthropogenic influence, and so did not explore the role of anthropogenic influences on changes in low flows, such as land cover change or water withdrawals (Brown et al., 2013).”

(5) Page 2767, line 1, The extensive efforts of the USGS to identify gauges with minimal human influence should be noted in this paragraph. The old HCDN from 1992 and the new HCDN (Lins, 2012) classify minimally influenced gauges. The latest set is the HCDN-2009 (Lins,2012) which looked at many quantitative factors in watersheds, including urbanization and regulation, and consulted with local USGS offices. It did not target just high flows in it’s classification. I think the current study should at least check their classification scheme vs. the HCDN-2009 to see how well it works for minimally disturbed watersheds. It’s not reasonable to assume that the author’s classification system is better, at least for minimally disturbed watersheds.

Response: We certainly did not assume that our classification system is better, but rather had the goal of developing a methodology that was transferable to situations where there was no classification or minimal information on human influence. We did compare our selected sites with those in the HCDN-2009 database and found that out of the 204 HCDN-2009 sites in our domain, only 64 had continuous daily records for our time period (1951-2005) that was common to all sites. We added a sentence in section 2.2: "Only 64 of these sites are in the HCDN-2009 database and have data for the common time period (1951-2005) that is used for analyzing trends across the domain (see section 4)."

(6) Page 2767, line 14, I think it's a bad assumption that step changes are indicative of an anthropogenic effect. This certainly happens, but step changes can occur for natural reasons, such as a change in a large scale ocean/atmosphere system (PDO, AMO, etc.). A step change in annual minimum flows at many minimally disturbed watersheds in the eastern United States was found by McCabe and Wolock (2002) which may be explained by North Atlantic sea surface temperatures (McCabe and Wolock, 2008).

Response: We have added another sentence to this paragraph to note that step changes can also be because of natural causes, but also note that our assumption is based on identifying abrupt and visually obvious step changes, which are likely to be due to anthropogenic influences. Visual inspection of the time series indicates that there are obvious abrupt shifts in many of the time series that are unlikely to be of natural origin, and these are identified by the combination of the Ljung-Box and Pettitt tests are described above. "Low flow time series (and flows in general) can show two general types of non-stationarity: gradually increasing or decreasing trends, and abrupt changes (Villarini et al., 2009) in the mean and/or variability. As McCabe and Wolock (2002) observe, the distinction between a gradual trend and a step change is important, particularly for climate-change impact studies, since climate change usually manifests as a trend and not a step change. We therefore assume that step changes in the time series are indicative of an anthropogenic effect, and that gradual trends reflect a climate effect. As it is possible that step changes may be driven by natural variability (e.g. McCabe and Wolock, 2008) our assumption is based on identifying abrupt and visually obvious step changes."

(7) Page 2768, line 17, Surely some parts of the Pacific Northwest and Alaska are wetter than the eastern United States.

Response: We have edited this: "The eastern US is one of the wettest parts of the country. . ."

(8) Page 2768, line 21, What lakes would cause lake-effect snow in NY City and Philadelphia if not shielded by the Appalachian Mountains?

Response: We have deleted this sentence.

(9) Page 2769, line 4, The water supply in the Northeast comes as least partly from groundwater from limited aquifers and bedrock. As far as I know, water suppliers have shifted from surface water to groundwater sources in recent years because of increased costs from required filtering of surface water. I think this is because of EPA regulations.

Response: Yes, agreed. This sentence does not discount groundwater sources, but highlights the main source as from surface water. We have edited the sentence slightly to make this clear. "The water supply in the northeast is mainly derived from surface waters, which are heavily regulated to meet the water supply demand of urbanized areas such as New York City, although there has been an increase in groundwater sources in recent years."

(10) Page 2769, line 16, "Several tens of stations" is a bit vague and misleading. Based on the HCDN-2009 map in Lins (2012) it looks like there are about 200 or so stations in your study area, though not all of these will have 50 years of data. Also, in the reference on line 15, this should be HCDN-2009, not HCDN. HCDN was the earlier network from 1992.

Response: The actual number of HCDN-2009 stations is 204, and we have corrected this. "several tens of stations" was incorrectly referring to stations with continuous records for our study period, which is now mentioned later in this section. We have corrected the reference: "Previous studies on low flows (e.g. Kroll et al., 2002, 2004; Douglas et al., 2000) have used the USGS Hydro-Climatic Data Network (HCDN; now updated to HCDN-2009; Lins, 2012),"

(11) Page 2770, line 2, Why is the median record length 18 years longer than the mean record length? This doesn't seem right.

Response: Yes, the median value is a typo. The correct values are: "The mean, median, minimum and maximum record lengths are 74, 72, 50, and 120, respectively."

(12) Page 2771, equations 1a-1c, The variables in the equations should be defined.

Response: We have updated the text to provide definitions: "where μ is the sample mean, sigma is the standard deviation and rho is the correlation, with i representing one realization of a time series"

(13) Page 2772, lines 10 and 19, "no trend" should be "no significant trend" as the lack of a significant trend does not prove that there is no trend. Also, the level of significance for the trend tests should be stated.

Response: We updated the text to reflect these suggested changes: "For the sites with significant overall autocorrelation, we then applied the Pettitt test (5% significance level) to..." "We then applied the Mann–Kendall (MK) test (5% significance level) on the remaining sites to identify statistically significant trends in the data. The series were categorized as having no significant trend, negative trend, or positive trend. "

(14) Page 2772, line 16, I disagree with the assumption that the lack of a significant step change combined with significant autocorrelation is necessarily a reflection of management effects. What about multi-year climatic droughts, for example in New England in the 1960s? What about groundwater storage? What is the basis for this assumption of management effects? What management effects would cause autocorrelation of low flows? Are there many reservoirs in the eastern US with the capability for multi-year

HESSD
12, C2650–C2664, 2015

storage?

Response: We based this assumption on visual examination of the time series of low flows, which showed that sites with significant overall autocorrelation often demonstrated an abrupt and obvious step change that is suggestive of anthropogenic effects, including management. As stated in the updated description of the methodology (see above) this was used to systematically identify abrupt step changes (that were then confirmed using the Pettit test). Nevertheless, we do agree that autocorrelation may also be a reflection of the stated drivers mentioned by the reviewer and have removed this sentence. Sites with overall autocorrelation (but no step change) are now noted as such, but not attributed to management effects.

(15) Page 2776, line 23, I don't see any gray colored sites in Figure 7, are the correct panels in this figure?

Response: We removed this sentence – it refers to an earlier version of the figure.

(16) Section 4.3, Are the tests for changes in low flow timing following the algorithm specified in section 2.3.2? This is confusing, for example in line 10, I thought sites with autocorrelation were removed prior to the MK tests. In line 17, results from the MK and Pettitt tests are compared. Is this for tests at the same sites or different sites? In the algorithm, MK and Pettitt were not done at the same sites.

Response: The analysis of timing changes has been updated to make it consistent with the algorithm that is applied to the flow volumes. Figure 9 now shows the results for the MK trends test, after removing the few sites with step changes and sites with overall autocorrelation. The discussion of the results has been updated.

(17) Page 2778, line 21, There is nothing about average month of low flows in my copies of Figure 9.

Response: This refers to an earlier version of the figure. We updated this section as mentioned in the previous point.

(18) Page 2781, line 10, Abrupt shifts can be climate related and gradual changes can be caused by changes in the management of flows by dams.

Response: Agreed, but our phrasing here is on the conservative side. We also note in the introduction and section 2.2 the possibility of alternative explanations for these different types of changes.

(19) Page 2781, line 24, It is the case that the USGS flags relate to high flows and not low flows.

Response: We have updated the sentence: “For sites with documented regulation or a change in measurement characteristics but no detectable signal, the fact that the USGS flags relate to high flows rather than low flows may help explain this, or that the sites are well managed in terms of low flows.”

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 12, 2761, 2015.