Response to referees' comments

We thank the referee for the very useful comments and suggestions. We have repeated the comments below and our responses are indented. Where we have added/edited text, we have highlighted this in italic font.

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.

> 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.

> 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.

> 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 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.

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.

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 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 auto-correlation 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 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.

> 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.

> 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).

> 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).

> 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.

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 introduction.

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.

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).

> 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 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.

> 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.

> 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).

> 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.

> 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?

> 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.

> 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.

> 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.

> 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.

> 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.

> 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 storage?

> 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?

> 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.

> 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.

> 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.

> 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.

> 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.”

Response to referee comments

We thank the referee for the very useful comments and suggestions. We have repeated the comments below and our responses are indented. Where we have added/edited text, we have highlighted this in italic font.

Anonymous Referee #2

General Comments

The manuscript presents an analysis of low flows in Eastern U.S. that is based streamflow annual minima (with different smoothing windows). In order to identify non-stationarity, the authors propose an algorithm (a cascade of 3 statistical tests) for which gauges are sorted into different classes and, depending on the outcome of the autocorrelation test, series are tested for trends or split into sub-series to be tested for trends. On the basis of this classification trend results are provided with a discussion on the possible causes identified in the study.

The writing is clear and results are generally well described and presented.

While I acknowledge that low flows are analyzed at remarkable spatial and temporal coverage, I have a few concerns with the stationarity analysis methods and assumptions (e.g. presence of autocorrelation invalidates the use of MannKendall but allows the use of Pettitt; change-points are assumed as human induced and are only tested for autocorrelated series). I also find the attribution part weak and think it should be titled differently. These issues are raised in the section below along with edits suggested to the text.

Specific Remarks

point-1 Page 2762 Lines 8-9:

You state : "to systematically distinguish the effects of human intervention from those of climate variability", as if this were the main goal of the paper (is it?), while it seems to me this is just a step. I suggest that framing of the overall scope of the paper should come first, after the initial introduction of Lines 1-5.

> Agreed. The main goal of the paper is to examine non-stationarity in low flow generation, with an initial step of identifying sites with time series that are likely not affected by direct anthropogenic influences. We have moved this part of the sentence to earlier in the abstract and edited it to better reflect the main goal:
>
> "*The goal of this paper is to examine non-stationarity in low flow generation across the eastern U.S. and explore the attribution to anthropogenic influences or climate variability*. We use nonparametric tests to identify abrupt and gradual changes in time series of low flows and their timing for 508 USGS streamflow gauging sites in the eastern US with more than 50 years of daily data,."

point-2 Page 2762 Lines 12-14:

Country wide hydrological databases, for as comprehensive as they can be, may not be so accurate on all gauges metadata and on all flow types (high/low): some gauges may be well documented, some others not so much. Of course this is valid for the streamflow data itself too, but in general, as there is no way to check the notes validity without data scrutiny and the help of the data providers, I would be cautious throughout the text referring to the USGS notes and using on them as supporting evidence. Finally, I find this phrase on the USGS notes not so relevant in the abstract.

"Examination of the USGS notes for each site confirmed that many of the step changes and around half of the sites with an increasing trend were associated with regulation."

> We acknowledge that the notes have uncertainties and may have shortcomings in understanding of changes in low flows, in particular because they are specific to high flows. However, we are interested in whether there is overall consistency between statistically identified changes and documentation of changes that might affect flows and their measurement.
>
> We have changed the parts of the abstract that link the quantified changes with documented changes to be more cautious:
>
> "Examination of the USGS notes for each site *showed that many of the sites with step changes and around half of the sites with an increasing trend have been documented as having some kind of regulation*. *Sites with decreasing or no trend are less likely to have documented influences on flows or changes in measurement characteristics*."

point-3 Page 2764 Lines 17-20:

The difficulty of low flow analysis with the advent of non-stationarity could be introduced and developed earlier in the introduction, particularly for the important consequences on hydrological analysis (i.e., the limits non-stationarity poses to the application of statistical tools).

> Agreed. We have moved this statement to earlier in the introduction (end of second paragraph):
>
> "Regulation generally introduces non-stationarity into low flow time series that impedes the development of regional or at-site frequency analysis models. In most instances, such models show a high standard error between modeled and observed quantiles (Kroll et al., 2004)."

point-3 Page 2766 Lines 28-29:

"we analyze the temporal and spatial distribution [..] to systematically distinguish the effects of human intervention from those of climate variability and change". Is this even possible over such a large area, which has been increasingly impacted by anthropogenic influence over the analyzed decades? Maybe use the verb "attempt to".

> Agreed. We have changed the sentence in response to a previous comment but have highlighted that this is an attempt to identify useable time series that are potentially free of influence.

> "*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."

As in point-1, is this the overall scope of the paper? Let the reader know why this distinction is relevant for this work.

> See response to point-1

point-4 Page 2767 Lines 1-3:

"Often the best way to determine ..". I don't agree (see point-2), I would replace with "A way to determine ..".

> We changed the sentence: "*A way to* determine whether a river has been subject to anthropogenic influences, at least in terms of regulation, is to examine the site notes for the gauging station"

point-5 Page 2767 Lines 6-8:

"we develop an alternative approach ". I find this statement somewhat misleading, it seems to suggests that the routine approach is to rely on site notes, and I don't believe it is the case.

> Agreed. We have removed the word "alternative".

" that assumes that the impact of human activities can be detected in the streamflow data in a systematic way". While I recognize the value of this approach for its ability to process virtually any number of sufficiently long streamflow series systematically, I remain skeptical with its efficacy and universal application. Isn't that a simplification rather than an assumption?

> Agreed. While we suggest that the method presented shows promise, we agree that further work is needed to understand the robustness of the method and its transferability. We have updated this sentence slightly:
>
> "that *makes the simplification* that the impact of human activities can be detected in the streamflow data in a systematic way"
>
> and have edited the sentence at the end of the conclusions that highlights the potential of the method to be applied elsewhere to also note the limitations/work to be done to understand its robustness and broader applicability:
>
> "The methods are readily transferable to other parts of the U.S. and globally, given long enough time series of daily streamflow data, *although further work is required to understand their universal application*."

point-6 Page 2767 Lines 14-15:

"We therefore assume that step changes in the time series are indicative of an anthropogenic effect ". Not necessarily, considering that step changes could result from climate variability (e.g. located at turning points from positive to negative phases of AMO, NAO, etc). For instance, you just mentioned that McCabe and Wolock (2002) reported 1970 as a step change, and that large-scale teleconnections may play an important role in driving changes in low flows (e.g. Giuntoli et al., 2013). You could add Mauget (2003) too [Mauget, S.A., 2003. Multidecadal regime shifts in US streamflow, precipitation, and temperature at the end of the twentieth century. J. Climate 16, 3905–3916.].

> See also our response to referee #1 on this comment.
>
> While we agree that the changes reported in these studies identify changes in different indicators (precipitation, temperature, and streamflow) the time scales over which these happen tend to be of the order of a few years, rather than an abrupt change that occurs from one year to another. In response to other referee comments, we have updated the text in various places to make it clear that the identified shifts are abrupt and visually obvious.
>
> We have added the Mauget reference.

point-7 Page 2768 Line 17:

"the" before "wettest"

> We corrected it.

point-8 Page 2769 Lines 6-7:

Probably worth mentioning that Florida's aquifer may have some inertia on the streamflow regimes and therefore low flows analysis is harder to achieve as a typically slower water response can result in drought events that are not always confined to the same water-year.

> Agreed. This is partially accounted for by calculating the autocorrelation, which will identify this type of behavior. We have added a note on this to this section:
>
> "In contrast, the southeast, including Florida, lies on active aquifers (USGS, 2009), *where low flow anomalies, such as in a drought year, may persist for multiple years*."

point-9 Page 2769 Line 18:

Not sure Fig. 1B is much relevant, and it is so crowded with overlapping dots that it's difficult to distinguish colors. I would just go with the selection of 508 sites used in the analysis (Fig. 1C).

> Agreed. We have deleted Figure 1B and updated the related text.

point-11 Page 2770 Line 6:

Can gauges belong to more than one category in Fig. 1D, so be affected by urbanization and have undergone a change in gauge datum?

> Yes, but only for a few sites. In this case we chose the flag with the highest likelihood of having affected the flows. For example, a "regulated" flag is assumed to have a larger effect than a change in gauge datum. We have noted this in section 2.2:
>
> "*A few sites have more than one type of flag and we show the flag associated with a higher likelihood of the flows being affected (e.g. regulated).*"

point-12 Page 2770 Lines 11-12:

Have you compared your results with the HCDN data set you mentioned above (with the gauges you identified as free from human intervention)?

> See response to referee #1.

point-13 Page 2770 Line 13:

The title of this section introduces 2 sub-sections about statistical methods, but this section is actually about low flow indices alone. I suggest to rename this section "2.3 Low flow indices", and maybe go with what is now 2.3.1 and 2.3.2 as 2.4 and 2.5 respectively.

> Agreed. We have updated the section numbering to 2.3, 2.4, 2.5, and changed the title of section 2.3 as suggested.

point-14 Page 2770 Lines 22:

You state that Q90 is useful for reservoir operations. Considering that indices are extracted yearly, 90 days is a very large smoothing window. Is Q90 really relevant to this study?

> We included Q90 because it is relevant to reservoir operations, and shows some interesting differences from the short-term Qn values. However, as the results for Q90 are similar to Q30 we have decided to remove the results for Q90 and only mention that they are similar. The results for Q1 are also now omitted because they are similar to those for Q7.

Response to referee comments

We thank the referee for the very useful comments and suggestions. We have repeated the comments below and our responses are indented. Where we have added/edited text, we have highlighted this in italic font.

Anonymous Referee #3

This work deals with the analysis of trends and step changes in low flow statistics at stations over the eastern part of the US, and attempts to relate findings to qualitative USGS flags. Although of scientific and operational interest, this study has some weaknesses that prevent its publication in HESS in the present form. My comments had been mainly drawn before the publication of other comments in the online discussion, but I can see that many of my points overlap with previously made ones.

In summary, I would suggest to investigate in much more detail the qualitative flags used and check the meaning of "no flag" for each station, all of this necessarily in close cooperation with USGS database managers. This would contribute to improve the conclusions in terms of relations between statistical findings and human disturbances. My two main comments are detailed below, followed by a list of more specific comments.

General comments

Understanding of the hydrometric database

The manuscript shows many examples of misunderstanding of the database flags, the most noticeable being the "change in gauge datum". This seems to reflect a lack of investigation on the meaning of these flags. More generally, such a study should be done in close cooperation with the database managers and field hydrologists. In that sense, the hard work made to identify reference hydrometric networks should be recognized, and more critically, used.

> We agree that closer cooperation would provide benefits. During the development of the study, we consulted with colleagues who had been using the flag information for studying changes in peak flows. We do acknowledge that the discussion of the flags and their use in interpreting the identified changes in low flows needs to be adjusted to better reflect the meaning of the flags. See responses to referees #1 and #2.

Below are some related comments on specific parts of the manuscript:

1. P2770 L3-6: The big question here is: What is the default in the database? Indeed, what is the meaning of a station with no flag? Is it actually a station with minor anthropogenic influence or change, or may it be a station that has not been documented (yet)? I know that other hydrometric databases include stations that are not flagged (by lack of time for a comprehensive overview) but should be. This is an issue that is not even mentioned in the manuscript, while it may have serious consequences on the interpretation of results.

> Table 1 shows that there are many sites that are identified as having statistically significant step changes (which are potentially due to some form of anthropogenic influence or change in how flows are measured – or possibly a climate regime shift) but do not have a USGS flag. The point of Table 1 is to show that the step changes identified are generally consistent with the USGS flags, but that there are also many changes that are not – there are multiple reasons for this as mentioned in the manuscript and by the referee here, including that the flags have not yet been assigned or that the anthropogenic impacts are small. We discuss the uncertainties with using the database, and the possible errors, in response to other referee comments.

2. P2770 L9-11: "The sites in the mid-Atlantic states are generally more affected by [...] change of gauge datum": This sentence implies that a change of gauge datum can be interpreted as a change in the catchment hydrological behaviour. Well, this is simply a change in the reference level for measuring water levels at the station. Besides, the list of flag you mention does not include dates of changes in the rating curve, which may have consequences in computed streamflow values, mainly for stations with unstable riverbed.

> Agreed. The original wording in the manuscript does imply that this is related to a change in the catchment response, which is incorrect. We have edited this sentence in response to other referees' comments.
>
> The flags in the USGS database do not include information on changes in rating curves and so this is another uncertainty – we have added this to the discussion of the streamflow data in section 2.2:
>
> "*Changes in the rating curve used to estimate streamflow from measured water levels are not recorded in the USGS notes but may be a significant source of variation in low flow values that is not accounted for.*"

3. P2773 L21-26 "this is mostly associated with a change in gauge datum" (and similar quotes): Again this serious issue of interpreting the "change in gauge datum" flag.

P2774 "If a site is flagged and its low flow series has a decreasing trend, this is mostly associated with a change of gauge datum"

"if a site is flagged and its low flow series has an increasing trend, this is mostly related to regulation or a change of gauge datum"

> Again, we have edited these sentences in response to other referees' similar comments.

Relating human disturbance and trends or step changes

There are several assumptions in the interpretation of trends and step changes in terms of potential causes that are clearly debatable and that undermine the overall conclusions. Indeed, gradual changes may for example come from either the climate or gradual changes in water abstractions and water management. A step back should be taken to consider all possible causes (climate, water abstraction, water management) to statistical findings.

Below are some related comments on specific parts of the manuscript:

1. P2767 L14-16 “We therefore assume that step changes in the time series are indicative of an anthropogenic effect, and that gradual trends reflect a climate effect”: This is a very strong assumption, and if climate change may indeed mainly cause gradual changes, this is also the case for different anthropogenic actions on the catchment. Examples of such actions can be found in the manuscript itself, for example P2768 L6-16, where you list a number of land cover / land use changes that gradually change the catchment hydrological behaviour. Similar comments may also be applied to gradual increase in water withdrawals, be they for drinking water following urbanization and population growth or for irrigation.

> See our response to referee #1.

2. P2772 L16-17: “Is a statistically significant step change is not identified, we assume that the autocorrelation is a reflection of management effects”. Well, this is again a very string assumption. Indeed, autocorrelation may come from natural long-term memory from e.g. aquifers.

> See our response to referee #1 on this same point.

3. P2776 L22: “regulation” What do you precisely mean by regulation? Regulation may for example aim at sustaining low flows above a given absolute level (for e.g., environmental flows), and this would have in this case a strong effect on Q1day or Q7days, but a limited effect on more temporally integrated indices like Q90days.

> We assume the referee is referring to P2775 L22. We have removed “regulation” from this sentence to be consistent with the slightly altered description of the assumptions about abrupt step changes. Please see earlier responses.
>
> “We further examined the consistency of the change year among the Q_n series, with the expectation that abrupt changes ~~caused by regulation~~ would be identified for the same year across all or most Q_n time series.”

4. P2777 L4: “rather than a direct anthropogenic impact on the low flows” Again it is not clear what you mean by “direct”. I could understand “indirect” through the consequences of anthropogenic climate change. But “direct” in my opinion applies to all human disturbances on the natural catchment hydrological behaviour, whether on land cover/ land use change, water management change, or combination of both.

> By “direct” we mean that the flows are manipulated directly through management. To be clear we have updated the sentence:
>
> “The attribution of trends at these sites is therefore likely related to climate variability/change and/or land use change, *rather than management of flows*.”

Specific comments

1. P2764 L4-6: I don't understand why the two facts should be conflicting. Please rephrase.

Agreed. We changed it as follows:

"Surface water covers 4.5 % of the eastern US, and the majority of streams have been flagged by the US Geological Survey (USGS) as regulated"

2. P2764 L16: I'm not sure that the reference used here is the most relevant one to support your statements.

The original reference cites numerous examples of the anthropogenic influences on low flows (although mostly with respect to ecological impacts) and we have chosen some examples to better directly support these statements. We have also added some other relevant references:

"Generally dams and reservoirs are considered the largest man-made regulations on streamflow, but other sources include farm ponds, surface water extraction, inter-basin transfers, and wastewater treatment plant discharge (*e.g. Walker and Thoms, 1993; Acreman et al., 2000; Brandes et al., 2005; Thomas, 2006; Deitch et al., 2009; Kustu et al. 2010*)."

Acreman, M. C., B. Adams, and B. Connorton, 2000: Does groundwater abstraction cause degradation of rivers and wetlands? *Water and Environment Journal,* 14, 200–206.

Brandes, D., Cavallo, G.J., Nilson, M.L., 2005: Base flow trends in urbanizing watersheds of the Delaware River basin. *J. American Water Resources Association*, 41 (6), art. no. 04114, pp. 1377-1391. doi: 10.1111/j.1752-1688.2005.tb03806.x

Deitch, M. J., G. M. Kondolf, and A. M. Merenlender, 2009: Hydrologic impacts of small-scale instream diversions for frost and heat protection in the California wine country. *River Research and Applications,* 25, 118-134.

Kustu, M. D., Y. Fan, and A, Robock, 2010: Large-scale water cycle perturbation due to irrigation pumping in the US High Plains: A synthesis of observed streamflow changes. *J. Hydrol*., 390 (3-4), 222-244. doi:10.1016/j.jhydrol.2010.06.045

Thomas, B., 2006. Trends in Streamflow of the San Pedro River, Southeastern Arizona, U.S. Geological Survey Fact Sheet 2006-3004, 4 pp., http://pubs.usgs.gov/fs/2006/3004/, accessed April 2011.

Walker, K. F. and M. C. Thoms, 1993: Environmental effects of flow regulation on the lower river Murray. *Australia. Regul. Rivers: Res. Mgmt*., 8, 103-119. doi: 10.1002/rrr.3450080114

3. P2768 L21: Could you elaborate on the "lake-effect snow"? I'm not sure any reader is familiar with it (I am not).

We deleted this sentence.

4. P2769 L8: "(EPA, 2008)": Could you provide any primary and recent literature on this?

> We have added a reference to Hayhoe et al. (2007), which documents historic and future projected changes for the eastern U.S.

5. P2770 L23: Is it the day with the minimum low flow? Please confirm.

> It is based on the Q7 dates. We updated the text:
>
> "We also calculate the day of the year of low flows and use this to identify the primary (and in some regions the secondary) low flow season, as well as any long-term changes in timing. *The timing results are shown based on Q7 flows*."

6. P2771 L1: I assume you wanted to write "A sequence of realizations of a random variable"

> Yes. We updated the text.

7. P2771 L7-9: Please define "i".

> This denotes one realization of the random variable:
>
> "with *i* representing one realization of a time series."

8. P2771 L14-16: Well, this may be true if you have a long enough series, which is rarely the case in hydroclimatology where the quest for understanding natural variability is still ongoing. Plus, I would strongly suggest using hydrological textbooks or papers rather than finance ones as reference works in order to better capture the specificities of the field.

> We updated this section to better reflect the general statistical and hydrological literature (also including updates in response to other comments):
>
> "A sequence of *realizations of* random variables, Y, is stationary if the distribution of the sequence is independent of the choice of starting point (*Kendall et al., 1983*; Ruppert, 2011). Determining stationarity of a time series is not straightforward (*Lins and Cohen, 2011*) and in practice, it is common to look at restricted measures of stationarity. A time series is defined as weakly stationary if it satisfies three criteria:
>
> [equations here]
>
> *where μ is the sample mean, σ is the standard deviation and ρ is the correlation, with i representing one realization of a time series*. This means that for a weakly stationary variable, the mean and variance do not change with time and the correlation between two values depends only on the lag (the time between values). Visual inspection of the time series and the changes therein can be very helpful in determining stationarity, *in that a change in the underlying process leads to changes in values that are obvious (Lins and Cohen, 2011; Koutsoyiannis, 2011)*"

Kendall, M., A. Stuart, and J. K. Ord, 1983: The Advanced Theory of Statistics, Vol. 3, Design and Analysis, and Time Series, 4th ed., 780 pp., Oxford Univ. Press, New York.

Koutsoyiannis, D., 2011: Hurst-Kolmogorov dynamics and uncertainty. *J. American Water Resources Association*, 47 (3), 481-495

Lins, H. F. and T. A. Cohn, 2011: Stationarity: Wanted Dead or Alive?. *J. American Water Resources Association*, 47: 475-480. doi: 10.1111/j.1752-1688.2011.00542.x

9. P2773 L9-11: "therefore a large number of sites appear stationary": why should there be a causal relationship here? 90 days is only one season and there may be trends/changes occurring on one season only. Please rephrase.

We have changed this sentence slightly:

"As we move from Q_1 to Q_{90}, a larger number of sites appear stationary (category 1) and the number of sites identified using the Pettitt test as having an abrupt shift in the time series (category 4) decreases."

10. Fig. 4: Does it show results from the first step of the algorithm? (I assume it does)

Yes. We have updated the caption and the associated text:

"However, there are also many sites in category 1 (45 %; no trend), 2 (34 %; decreasing trend) and 3 (67 %; increasing trend) that are also flagged (see Fig. 4) "

"Figure 4. Categorization of non-stationarity of sites for Q1 with no USGS flags *from the first step of the decomposition algorithm*."

11. P2776 L1-2: "Q1 may be the most appropriate for identifying a change since they are based on the original time series data": I personally disagree. Indeed, Q1 are more prone to measurement errors at so low water levels than more temporally integrated indices. Q7, or MAM(7) as described by WMO (2008), is much more widely used and in my sense more suitable here.

We have updated the sentence to reflect this:

"Although we have identified the change year for all Q_n, the results for Q_7 may be the most appropriate for identifying a change *since the data are close to the original values, but are less affected by measurement errors than* Q_1 *(WMO, 2008)*."

12. P2776 L23: There is no grey point in Fig. 7.

This sentence referred to an earlier version of the figure. The sentence has been removed.

13. Fig. 7: There is some inconsistency between (b) and (c). Plus, did you apply here some MK test taking account of autocorrelation?

Yes. The figure has been updated. See response to referee #1 about the autocorrelation.

14. P2777 L 11-12: “If the onset time of the low flow season for a site occurs 70 to 100

The referee comment appears to be incomplete so we cannot provide a response.

15. Section 5.1: I would recommend changing the section title, as there is no formal attribution performed here, only observations of qualitative correlation.

We changed the title to: “5.1. Potential Drivers of Trends in Low Flows”

16. Section 4.3: So If I understand well, you remove from the analysis all sites that have two low flow seasons. This means that you are removing all sites that could see a shift in absolute minimum flow from one season to the other, and which are the most interesting ones, from a process point of view, but also from a water management point of view. This would completely change the pattern shown in Fig. 9.

Our original analysis looked at all sites irrespective of whether there was a single low flow season or not, to explore not only whether timings have shifted within a season but also from one season to another, e.g. in the northeast where warming temperatures have altered the freezing regime – something that we agree is interesting. Unfortunately, there was not space to include this full analysis and so we decided to focus on the sites with a single season to simplify the analysis and presentation of results. In any case, the evidence for shifts in timing between seasons was minimal. We have added a short discussion of this at the top of section 4.3:

“*Analysis of changes in timing irrespective of the season (not shown) did not show evidence of shifts in timing from one season to another*.”

17. Fig. 10 (a): What is the “warm season”? Plus, what sites are exactly plotted here? I would assume that only unregulated ones (or at least the ones not flagged as regulated) should be presented here.

This plot showed the results for all sites without step changes. We have updated it to show only sites without step changes and without flags – i.e. those without potential regulation.

Technical corrections

1. Figures: they are all very difficult to read (most notably Fig. 5 and 6, but all others). However, there is redundant information that could be removed to make them bigger: axes across subplots, legends across subplots, etc.

We have removed plots for Q1 and Q90 because they are very similar to the results for Q7 and Q30, respectively, and have updated the text throughout. This has enabled us to expand the size of the panels in Figures 5 and 6. We have also edited the other figures where possible to make them easier to read.

Nonstationarity of low flows and their timing in the eastern United States

S. Sadri, J. Kam, and J. Sheffield

Department of Civil and Environmental Engineering,
Princeton University,
Princeton,
New Jersey,
08544
U.S.A.

Corresponding author: Justin Sheffield, Department of Civil and Environmental Engineering, Princeton University, Princeton, New Jersey, U.S.A, 08544 (justin@princeton.edu)

Abstract

The analysis of the spatial and temporal patterns of low flows as well as their generation mechanisms over large geographic regions can provide valuable insights and understanding for climate change impacts, regional frequency analysis, risk assessment of extreme events, and decision-making regarding allowable withdrawals. The goal of this paper is to examine non-stationarity in low flow generation across the eastern U.S. and explore the attribution to anthropogenic influences or climate. We use nonparametric tests to identify abrupt and gradual changes in time series of low flows and their timing for 508 USGS streamflow gauging sites in the eastern US with more than 50 years of daily data, to systematically distinguish the effects of human intervention from those of climate variability. A time series decomposition algorithm was applied to 1-day, 7-day, 30-day, and 90-day annual low flow time series that combines the Box-Ljung test for detection of autocorrelation, the Pettitt test for abrupt step changes and the Mann-Kendall test for monotonic trends. Examination of the USGS notes for each site showed that many of the sites with step changes and around half of the sites with an increasing trend have been documented as having some kind of regulation. Sites with decreasing or no trend are less likely to have documented influences on flows or changes in measurement characteristics. Overall, a general pattern of increasing low flows in the northeast and decreasing low flows in the southeast is evident over a common time period (1951-2005), even when discarding sites with significant autocorrelation, documented regulation or other human impacts. The north-south pattern of trends is consistent with changes in antecedent precipitation. The main exception is along the mid-Atlantic coastal aquifer system from eastern Virginia northwards, where low flows have decreased despite increasing precipitation, and suggests that declining groundwater levels due to pumping may have contributed to decreased low flows. For most sites, the majority of low

Unknown

Deleted: confirmed that many of the step changes and around half of the sites with an increasing trend were associated with regulation

Justin Sheffield 6/15/15 4:27 PM

Deleted: Around a third of the sites with a decreasing trend were associated with a change of gauge datum.

flows occur in one season in the late summer to autumn, as driven by the lower precipitation and higher evaporative demand in this season, but this is complicated in many regions because of the presence of a secondary low flow season in the winter for sites in the extreme northeast and in the spring for sites in Florida. Trends in low flow timing are generally undetectable, although abrupt step changes appear to be associated with regulation.

Keywords: Eastern U.S.; Low flows; Non-stationarity; Abrupt change; Gradual trends; Autocorrelation; Ljung-box test; Mann-Kendall test; Pettitt test

1. Introduction

Low flows - the minimum flow in a river during the dry periods of the year--- are an important part of the streamflow regime that have direct impacts on water supply, water quality, and ecosystem health (Bradford and Heinonen, 2008). Knowledge of low flow characteristics and generation mechanisms over large geographic regions is important for regional frequency analysis, risk assessment of extreme events, decision-making regarding allowable basin withdrawals and water quality, and understanding climate change impacts (Tallaksen and van Lanen, 2004). For example, in every state of the U.S., estimates of low flow statistics are needed for issuing and/or renewing of National Pollution Discharge Elimination System permits, as required by provisions in the Clean Water Act of 1977 (U.S. Senate, 2002). Furthermore, low flow periods are critical to aquatic habitats due to potentially low dissolved oxygen concentrations and/or high pollutant concentration (U.S. Senate, 2002). However, the study of low flow statistics and patterns have received little attention in comparison to droughts and floods (Kroll et al., 2004). Poff et al. (1997) emphasize the need of paying particular attention to low flows because they present critical stresses and opportunities for a wide array of riverine projects.

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. However, our understanding of these low flow generating mechanisms is limited (Smakhtin, 2001), and is further compounded by the sensitivity of low flows to changes in climate, land use and human impacts on stream flow (Rolls et al., 2012). For example, large-scale teleconnections

Justin Sheffield 6/15/15 2:37 PM
Formatted: Font:Not Italic

may play an important role in driving inter-annual to multi-decadal changes in streamflow (e.g. Mauget, 2003) and low flows (e.g. Giuntoli et al., 2013). Regulation generally introduces non-stationarity into low flow time series that impedes the development of regional or at-site frequency analysis models. In most instances, such models show a high standard error between modeled and observed quantiles (Kroll et al., 2004).

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 have impacted low flows, including land use change impacts via changes in sub-surface flow and groundwater recharge, direct impacts on flows via reservoirs and other streamflow management, and changes in precipitation and evaporation that have altered recharge. In particular:

Justin Sheffield 6/15/15 2:36 PM
Formatted: Font:Not Italic

1. In the U.S., more than 85% of the surface runoff is artificially controlled and nearly 1 million km of rivers are affected by dams (Poff et al., 1997). Surface water covers 4.5% of the eastern U.S., and the majority of streams have been flagged by the U.S. Geological Survey (USGS) as regulated. The USGS estimates that the spatial extent of surface water increased by 1.3% during 1973-2000, with most of this increase in the southern coastal plain and southern Florida coastal plain (USGS, 2012) and associated with reservoir developments required to meet the needs of the expanding population. Figure 1a shows the location of major dams in the eastern U.S. (defined as those 50 feet or more in height, or with a normal storage capacity of 5,000 acre feet (~6,200,000 m^3) or more, or with a maximum storage capacity of 25,000 acre feet (~30,800,000 m^3) or more (USACE, 2012)). Generally dams and reservoirs are considered the largest man-made regulations on streamflow, but other sources include farm ponds, surface water extraction, inter-basin transfers, and wastewater treatment

Justin Sheffield 7/16/15 10:02 PM
Deleted: While s

Justin Sheffield 7/16/15 10:02 PM
Deleted: only

plant discharge (e.g. Walker and Thoms, 1993; Acreman et al., 2000; Brandes et al., 2005; Thomas, 2006; Deitch et al., 2009; Kustu et al. 2010).

Unknown

Deleted: Rolls et al., 2012

Justin Sheffield 6/15/15 4:31 PM

Deleted: Regulation generally introduces non-stationarity into low flow time series that impedes the development of regional or at-site frequency analysis models. In most instances, such models show a high standard error between modeled and observed quantiles (Kroll et al., 2004).

2. The eastern U.S. has gone through significant land use change over the past several decades. For example, between 1973 and 2000, 8.2% of the 23,620,000 km^2 of the northeast ecoregion and 8.9% of the 30,000,000 km^2 of the southeast ecoregion experienced changes associated with active timber harvesting and replanting, which may have impacted low flows and related environmental and ecosystem well-being (USGS, 2012). Furthermore, in the expanding urbanized areas of the region with high levels of impervious ground, infiltration has decreased, which may have led to a decrease groundwater recharge and low flow volumes (USGS, 2013). 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 al., 2005).
3. The region is one of the wettest parts of the U.S. receiving 700-1600 mm of precipitation per year. However, due to population growth and associated increased use of surface and groundwater resources, the future is expected to bring water stress for this area (Averyt et al., 2013). Some of these changes are already being observed. For example, USGS (2013) reports on 3-10 km^3 of depletion of unconsolidated and semi-consolidated sand and gravel aquifers of the east coast between 1900 and 2008. Overuse of surface water in turn does not allow recharge of groundwater leading to groundwater depletion. In parts of the eastern U.S., groundwater resources have become limited and hence municipal and industrial water users are increasingly relying on surface waters (e.g. Daniel and Dahlen, 2002). Changes in both surface water and groundwater use have impacts on low flows.

4. Precipitation has likely changed over the past several decades (Karl and Knight, 1998; Small et al., 2006). Evaporation may have changed due to increasing atmospheric demand from higher temperatures (e.g. Walter et al., 2004), although direct measurements of evaporation are limited in spatial and temporal coverage. Each of these changes may impact on low flows and in some cases may combine to exacerbate or counteract changes in low flows. Warmer temperatures may have also impacted winter-time low flows, via changes in snow (Burakowski et al., 2008) and river ice (Hodgkins et al., 2005).

Past evaluations of changes in low flows over the eastern U.S. have mainly been within studies on the entire U.S. and often with respect to mean and high flows. Douglas et al. (2000) estimated trends in both flood and 7-day low flows for three major geographic regions in the U.S. (East, Midwest, and West) over two time periods: 1959-1988 and 1939-1988, and found evidence of upward trends in low flows across the Midwest, but not in the eastern U.S. Other studies have attempted to explain the general patterns of low flow trends. For example, Small et al. (2006) analyzed trends in annual 7-day low flow, average, and high flows along with seasonal precipitation over individual basins in the U.S. for 1948-1997. The number of sites shown to have statistically significant trends in low flows and fall precipitation in the eastern U.S. was small and restricted to the south of Maine, western Pennsylvania, coastal areas of South Carolina, and western Florida. In the northeast and west of Pennsylvania, precipitation showed an increasing trend during the fall but not during the spring and the increase in fall precipitation appeared to result in an increase in low flows in the northeast areas. The only statistically significant decrease in the low flows was found in the south Atlantic-Gulf region, west of Florida, consistent with the findings from Lins and Slack (1999). However, no specific reason for this decreasing trend was given. McCabe and Wolock (2002) examined historic changes in streamflow, using the annual minimum, median, and maximum daily streamflow at 400 sites across the U.S. during 1941-1999. They found an increase in annual minimum and median daily streamflow around 1970 that primarily occurred in the eastern U.S. as a step change, rather than a gradual trend. Andreadis et al. (2006) used model simulations to examine trends in soil moisture, runoff, and drought characteristics over the U.S. for the period 1915-2003. They found increasing runoff over parts of the northeast, which was most evident during winter months, with decreases in hydrological and agricultural drought, and drying trends in the summer in the southeast, with increases in drought. These changes were attributed to changes in precipitation, and they speculated that increasing drought in the southeast was associated with higher atmospheric demand due to warming. Although these studies are generally consistent for the eastern U.S. they tend to focus on the spatial pattern of trends in 7-day low flows only, and were limited to earlier periods available at the time of the study. Furthermore, these studies focused on sites that were deemed to have minimal anthropogenic influence, and so did not explore the role of anthropogenic influences, such as land cover change or water withdrawals (Brown et al., 2013).

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. A way to determine whether a river has been subject to anthropogenic influences, at least in terms of regulation, is to examine the site notes for the gauging station. However, site notes might not be available, complete, or accurate, and examining the notes for multiple sites can be unwieldy. Furthermore, whether a site is determined to be regulated or not is often based on high flows and not on low flows. Here, we develop an approach that makes the simplification

Justin Sheffield 6/15/15 2:45 PM
Deleted: none of

Justin Sheffield 6/15/15 2:45 PM
Deleted: these studies considered

Justin Sheffield 6/15/15 4:37 PM
Deleted: I

Justin Sheffield 6/25/15 10:29 AM
Formatted: Font:Not Italic

Justin Sheffield 6/15/15 4:37 PM
Deleted: n this study, we analyze the temporal and spatial distribution of annual *n*-day low flows in the eastern U.S. to

Justin Sheffield 6/3/15 2:44 PM
Deleted: distinguish

Justin Sheffield 6/3/15 2:43 PM
Deleted: the

Justin Sheffield 6/3/15 2:46 PM
Deleted: from those of

Justin Sheffield 6/15/15 4:41 PM
Deleted: Often the best

Justin Sheffield 6/15/15 4:41 PM
Deleted: alternative

Justin Sheffield 7/17/15 6:56 PM
Deleted: assumes

that the impact of human activities can be detected in the streamflow data in a systematic way. This is generally more efficient and can complement site notes or compensate for errors in them. 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 (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). 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.

Justin Sheffield 7/16/15 4:34 PM
Deleted: .

Justin Sheffield 7/16/15 4:34 PM
Deleted: .

Justin Sheffield 7/16/15 4:34 PM
Deleted:

Our overall approach is to use nonparametric statistical tests to identify abrupt and gradual changes in the value and timing of *n*-day low flows, and identify stationary segments of the time series. Furthermore we analyze the co-variability of low flows with antecedent precipitation to understand the influence of changes in precipitation and atmospheric demand (as quantified by potential evapotranspiration) on changes in low flows. The paper is organized as follows: Section 2 describes the streamflow data and the methodology, including the use of three straightforward and already-established statistical methods, for identifying non-stationarity in annual low flow time series. The results on the systematic identification and characterization of abrupt changes in low flow volumes and timing are presented in Section 3. The results on the variability and trends in are given in Section 4. Finally, we discuss the results, their attribution and implications, and present conclusions in Section 5.

2. Data and Methods

2.1. Study area

Our study area covers the eastern U.S. from Maine in the northeast to Florida in the southeast and westwards to the Appalachian Mountains and the Mississippi River in the south, and is based on the 20 ecoregions of the eastern U.S. (USGS, 2012). According to the USGS (2012), 52.4% of the eastern ecoregion in 2000 was forest. However, both forests and agriculture have been in decline since 1973 and instead, urbanization has increased and continues to increase. Most land cover change has occurred in the southeast and is associated with forest harvesting, agricultural abandonment, and development (USGS, 2012). Changes in the northeast have been mostly associated with timber harvesting. Changes in the north Central Appalachian region have been more heterogeneous and include examples of non-mechanical transitional change. Unlike the northeastern Coastal Plain, the southern Florida Coastal Plain has not experienced loss of agricultural land, but the largest decrease in surface water and significant loss of wetlands (-2.4%). Changes in surface water in the southern Coastal Plain have primarily been due to urbanization (USGS, 2012).

The eastern U.S. is one of the wettest parts of the country (Small et al., 2006), with average precipitation of about 1100 mm per year, with maxima along the coastal plain and the mountains of the Appalachians. Part of the precipitation in the northeast falls as snow in the wintertime (Hayhoe et al., 2007). The eastern seaboard is susceptible to tropical storms and hurricanes during the Atlantic hurricane season, normally running from June to end of November, which enhance precipitation across southern and eastern parts, and play a role in

Justin Sheffield 6/15/15 2:52 PM
Deleted: The Appalachian Mountains largely shield New York City and Philadelphia from lake-effect snow (Colby, 2008).

alleviating drought (Kam et al., 2013). The El Niño-southern Oscillation (ENSO) alters precipitation patterns across the southeast (Colby, 2008). Coastal extra-tropical cyclones bring the bulk of the wintertime precipitation to that region, forming along the natural temperature gradient of the Gulf stream before moving up the coastline (Gurka et al., 1995). Seasonally, there are slight changes in the precipitation distribution through the year. For example, Burlington, Vermont has a summer maximum and a winter minimum while Portland, Maine has a fall and winter maximum, with a summer minimum in precipitation. 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. In contrast, the southeast, including Florida, lies on active aquifers (USGS, 2009). Projections of future climate indicate an increase in precipitation over the eastern U.S. (Hayhoe et al., 2007; EPA, 2008) with consequences for changes in low flows across the region.

Justin Sheffield 6/15/15 2:54 PM
Deleted: The northeast does not have any principal aquifer and, therefore, t

2.2. Streamflow data

Initially, 4878 sites with daily streamflow records were retrieved from the USGS National Water Information System (NWIS) (USGS, 2014) for the eastern U.S. as defined by Hydrological Unit Codes (HUC) of 01, 02, or 03. 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), in part because anthropogenic influences at these sites are deemed to be negligible, but as such, is limited to 204 sites across the domain. Of the original 4878 sites, 2811 were active in the 2000's or later. Among these, 1092 sites had at least 30 years worth of daily data, 740 sites had 50 years or more, and 324 sites had 75 years or more. We used

Justin Sheffield 7/18/15 10:52 AM
Deleted: only several tens of stations

Justin Sheffield 7/25/15 6:23 PM
Deleted: (Figure 1b)

sites with at least 50 years of data as a balance between having enough of data at each site to identify long-term changes and the need to have many sites to characterize the spatial pattern of changes. We included only sites that did not have any missing years of daily data. This reduced the number of sites to 508 (Figure 1b). 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). The drainage area of the candidate sites ranges from very small (5-100km^2) to large (38,000-67,000km^2), with the majority of areas between 200-500 km^2 and these are spread fairly uniformly across the study area. The majority of the 508 sites are clustered on the eastern flank of the Appalachians and the northeast from eastern Virginia to New Hampshire. There is also a cluster of smaller catchments in central Florida. The mean, median, minimum and maximum record lengths are 74, 72, 50, and 120, respectively.

Justin Sheffield 7/25/15 6:24 PM
Deleted: c

Justin Sheffield 7/18/15 10:26 AM
Deleted: 9

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 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. Changes in the rating curve used to estimate streamflow from measured water levels are not recorded in the USGS notes but may be a significant source of variation in low flow values that is not accounted for. Figure 1c 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. A few sites have more than one type of flag

Justin Sheffield 7/25/15 6:24 PM
Deleted: d

and we show the flag associated with a higher likelihood of the flows being affected (e.g. regulated). 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 flagged as affected in terms of anthropogenic influences or changes in measurement method. In the results section, we show how the results of our statistical methods compare with the USGS site flags.

Justin Sheffield 6/3/15 2:12 PM
Deleted: and

Justin Sheffield 7/18/15 10:38 AM
Deleted: somehow

2.3. Low Flow Indices

Justin Sheffield 7/17/15 2:44 PM
Deleted: Statistical Methods

We analyze four variants of low flows based on different time scales, to understand how non-stationarity is dependent on the time scale as the data become smoother, with implications for the detection of non-stationarity. The 1-day minimum low flow, Q_1, is the annual minimum daily streamflow. The other three variants, Q_7, Q_{30}, Q_{90}, are obtained by applying the same analysis to 7-day, 30-day, and 90-day moving average versions of the time series. Together, we refer to the four low flow variables as the *n*-day minimum flows. Q_7 (dry weather flow) is the most widely used low flow statistic in the U.S. (Kroll et al., 2004; Smakhtin, 2001), but the others are important for different applications, such as Q_1 for ecological assessments and Q_{90} for reservoir operations. We also calculate the day of the year of low flows and use this to identify the primary (and in some regions the secondary) low flow season, as well as any long-term changes in timing. The timing results are shown based on Q7 and Q30 flows.

2.4 Identification of stationary time series

Justin Sheffield 7/17/15 2:44 PM
Formatted: Font:Not Italic

Justin Sheffield 7/17/15 2:44 PM
Deleted: 3.1

A sequence of realizations of random variables, *Y*, is stationary if the distribution of the sequence is independent of the choice of starting point (Kendall et al., 1983; Ruppert, 2011). Determining stationarity of a time series is not straightforward (Lins and Cohen, 2011) and in practice, it is common to look at restricted measures of stationarity. A time series is defined as weakly stationary if it satisfies three criteria:

$$\mathrm{E}(Y_i) = \mu, \quad (\forall i) \tag{1a}$$

$$Var(Y_i) = \sigma^2, \quad (\forall i) \tag{1b}$$

$$Corr(Y_i, Y_j) = \rho(|i - j|), \quad (\forall i, \forall j) \tag{1c}$$

where μ is the sample mean, σ is the standard deviation and ρ is the correlation, with *i* representing one realization of a time series. This means that for a weakly stationary variable, the mean and variance do not change with time and the correlation between two values depends only on the lag (the time between values). Visual inspection of the time series and the changes therein can be very helpful in determining stationarity, in that a change in the underlying process leads to changes in values that are obvious (Lins and Cohen, 2011; Koutsoyiannis, 2011; Serinaldi and Kilsby, 2015).

Justin Sheffield 6/15/15 3:09 PM
Formatted: Font:Italic

Justin Sheffield 6/15/15 3:09 PM
Formatted: Font:Italic

Justin Sheffield 6/15/15 3:09 PM
Formatted: Font:Italic

Justin Sheffield 7/20/15 9:47 AM
Deleted: However, weak stationarity does not apply to the quantiles and higher moments of the distribution (Ruppert, 2011).

Justin Sheffield 7/19/15 9:17 AM
Deleted: A time series plot of a stationary series should show oscillation around some fixed level, a phenomenon called reversion to the mean (Ruppert, 2011).

We apply three tests to identify weak stationarity: (1) the Mann-Kendall test (Mann, 1945; Kendall, 1975), which tests for increasing or decreasing trends; (2) the Pettitt test (Pettitt, 1979), which tests for abrupt changes or change points; and (3) the Ljung-Box test (Ljung and Box, 1978), which tests for autocorrelation. An identified change in the mean by either of the first two tests would rule out stationarity, except in the case of autocorrelated data, for which the Mann-Kendall test will characterize too many sequences of the time series as having a trend (Douglas et al., 2000). Therefore, analysis of autocorrelation is carried out before conducting the

Mann-Kendall test. Even when a site is identified as non-stationary, further analysis is required to understand the overall regime of the data at such a site. For example, the time series may have two separate stationary regimes with one change point in between or an overall trend. We then assume that the change year corresponds to human intervention, which is generally borne out by investigating the site notes.

2.5 Decomposition algorithm

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 20 lags 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

Justin Sheffield 7/17/15 2:44 PM
Formatted: Font:Not Italic

Justin Sheffield 7/17/15 2:44 PM
Deleted: 3.2

Justin Sheffield 7/17/15 2:44 PM
Formatted: Font:Not Italic

Justin Sheffield 7/16/15 4:22 PM
Deleted: length

Justin Sheffield 7/16/15 4:22 PM
Deleted: years

Justin Sheffield 7/16/15 4:07 PM
Deleted: p-value < 0.05

Justin Sheffield 7/16/15 4:07 PM
Formatted: Font:Not Italic

Justin Sheffield 7/16/15 4:07 PM
Formatted: Font:Not Italic

Justin Sheffield 7/16/15 4:07 PM
Deleted: We then applied the Mann-Kendall (MK) test on the remaining sites to identify statistically significant trends in the data. The series were categorized as having no trend, negative trend, or positive trend.

Unknown
Deleted: identify step changes

Justin Sheffield 7/16/15 4:09 PM
Formatted: Font:Not Italic

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 series and sub-series are assigned categories as follows:

Justin Sheffield 7/16/15 4:10 PM
Deleted: assume that the autocorrelation is a reflection of management effects and the site is excluded from further consideration

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 step change, time series split and the sub-series re-categorized recursively;
5. Category 5: Autocorrelated site with no step change.

Justin Sheffield 7/16/15 4:11 PM
Deleted: Site removed from further analysis.

3. Stationarity Results

3.1. Categorization of sites

Figure 2 shows the spatial distribution and the number of sites in each category after the first recursive level of the decomposition algorithm. The results for all *n*-day low flow metrics are presented for the available length of record at each site, which ranges between 1891 and 2011. No site has a record length less than 50 years and no site has any gap in the *n*-day low flow series. As we move from Q_1 to Q_{90}, a larger number of sites appear stationary (category 1) and the number of sites identified using the Pettitt test as having an abrupt shift in the time series

Justin Sheffield 7/19/15 9:14 AM
Deleted: the low flow values are calculated from smoother time series values and therefore

(category 4) decreases. The algorithm re-applies the Pettitt test to category 4 sites to identify useable sub-series. For example, the Q_1 time series of 155 sites are split into two parts, which are subjected to further categorization.

Figure 3 summarizes the time periods that were identified as useable at each step of the recursive algorithm for all sites for Q_1. The light blue lines represent the original record length for each site. The vertical axis shows the site number from 1 to 508 ordered from the lowest to highest latitude. Therefore, site 1 is the most southerly and site 508 is the most northerly. The left panel of Figure 3 shows the record length of sites, which, in the first step of categorization, had no significant autocorrelation. These sites are colored according to their MK trend value: 0 (no significant trend), -1 (significant negative trend), or 1 (significant positive trend). The middle panel again shows the original record length for each site in light blue, but highlights the sites that were identified with an abrupt step change by the Pettitt test and were split into two parts. For each part that exhibits no autocorrelation, the trend values were calculated. The right panel shows the parts of the time series that were recovered in the next step of the decomposition algorithm. As long as the record length is greater than or equal to 30 years the algorithm is applied recursively on the remaining parts of the time series. The number of sites shown in the right panel is small but their data are still useful for subsequent analysis.

3.2. Comparison with USGS flags

Table 1 shows the breakdown of the number of sites in each category and the relation to USGS flags for Q_7 and Q_{30}, and indicates that in every category, anthropogenic influences or measurement changes are documented by the USGS. For Q_7, the majority of sites in categories 4 (64%; step change), and 5 (58%; significant autocorrelation) are flagged by the USGS as

Justin Sheffield 9/12/15 11:45 AM
Deleted: $_1$

Justin Sheffield 9/12/15 11:45 AM
Deleted: $_9$

Justin Sheffield 9/12/15 10:50 AM
Deleted: $_1$

Justin Sheffield 9/12/15 11:45 AM
Deleted: 6

somehow affected. This suggests that the algorithm has some skill in identifying managed or altered flow series. However, there are also many sites in category 1 (50%; no trend), 2 (35%; decreasing trend) and 3 (60%; increasing trend) that are also flagged (see Figure 4) suggesting that anthropogenic impacts or changes in measurement characteristics for these sites are minimal and/or are overwhelmed by any climate or land use induced changes. The fact that the majority of stationary sites (category 1) are not flagged is encouraging. Figure 4 shows all the sites from each of the 5 categories that have no flag for Q_7; 237 out of 508 sites are not flagged but only 119 of these 237 sites show absolute stationarity behavior (category 1) and the rest exhibit some form of non-stationary.

From Table 1 we observe that:

1. If a site is flagged and its low flow series has a decreasing trend, the flags are mostly for a change of gauge datum;
2. If a site is flagged and its low flow series has an increasing trend, the flags are mostly related to regulation or a change of gauge datum;
3. If a site is flagged and it exhibits a step change, the flag is mostly associated with regulation, a change of gauge datum, or possibly urbanization;
4. If a site is in category 5 (not considered further due to significant autocorrelation), it may be flagged as regulated or its gauge datum has changed;
5. If a site shows no trend but is still flagged, the flag relates to regulation or a gauge datum change. This suggests that the impact of the flagged change was either minimal or good management practices have been put in place. The majority of these sites are located in the upper Mid-Atlantic in the states of New York, New Jersey, and Virginia.

Justin Sheffield 9/12/15 11:46 AM
Deleted: 45

Justin Sheffield 9/12/15 11:46 AM
Deleted: 4

Justin Sheffield 9/12/15 11:46 AM
Deleted: 7

Justin Sheffield 9/12/15 10:44 AM
Formatted: Not Highlight

Justin Sheffield 9/12/15 10:43 AM
Deleted: 1d

Justin Sheffield 9/12/15 10:43 AM
Deleted: , for example,

Justin Sheffield 9/12/15 10:44 AM
Deleted: $_1$.

Justin Sheffield 9/12/15 10:44 AM
Deleted: For Q_1,

Justin Sheffield 9/12/15 10:45 AM
Deleted: 21

Justin Sheffield 9/12/15 10:45 AM
Deleted: are

Justin Sheffield 9/12/15 10:45 AM
Deleted: somehow

Justin Sheffield 6/3/15 3:13 PM
Deleted: this

Justin Sheffield 6/3/15 3:14 PM
Deleted: is

Justin Sheffield 6/3/15 3:14 PM
Deleted: associated with

Justin Sheffield 6/25/15 10:34 AM
Deleted: this is mostly

Justin Sheffield 6/25/15 10:35 AM
Deleted: is

Justin Sheffield 6/25/15 10:35 AM
Deleted: be

Justin Sheffield 6/25/15 10:35 AM
Deleted: it may be

Justin Sheffield 6/25/15 10:36 AM
Deleted: ed

Justin Sheffield 6/25/15 10:36 AM
Deleted: its

Justin Sheffield 6/25/15 10:36 AM
Deleted: d, but the statistics do not indicate a trend

Justin Sheffield 6/3/15 3:15 PM
Deleted: impact

Justin Sheffield 6/3/15 3:16 PM
Deleted: of regulation

Dam failure, flow below the rating curve limit, or change of base discharge and even urbanization do not appear to have a significant impact on low flows or their measurement for the sites considered.

3.3. Variability in year of abrupt change

For sites that were identified by the Pettitt test as having an abrupt change, Figure 5a shows the variability of the year of change for Q_n. Most of the changes occurred between 1962 and 1986, and as discussed above, most of these are flagged as having regulation or a change of gauge datum. The spatial distribution of changes indicates that stream regulation began in the northeast before spreading to the southeast. The Pettit test tends to identify significant changes away from the either ends of the time series, and so may not identify changes in the earlier or later part of the record. However, earlier or later step changes are identified in the second recursion of the decomposition algorithm.

We further examined the consistency of the change year among the Q_n series, with the expectation that abrupt changes would be identified for the same year across all or most Q_n time series. Figure 5b shows the spatial distribution and the number of sites with a consistent year of change among the Q_n. Out of 176 sites whose time series were identified as having a step change by the Pettitt test, 82 (almost half) showed the same change year for 3 out of 4 Q_n series. Only 7 sites showed the same change year for all Q_n. Although we have identified the change year for all Q_n, the results for Q_3 may be the most appropriate for identifying a change since the data are close to the original values, but are less affected by measurement errors than Q_1 (WMO, 2008).

4. Variability and Trends in Low Flows and Timing

Justin Sheffield 6/3/15 3:17 PM
Deleted: associated

Justin Sheffield 6/3/15 3:17 PM
Deleted: with

Justin Sheffield 6/3/15 3:17 PM
Deleted: and

Justin Sheffield 7/17/15 2:29 PM
Deleted: caused by regulation

Justin Sheffield 6/9/15 4:51 PM
Deleted: ,

Justin Sheffield 6/9/15 4:53 PM
Deleted: they are

Justin Sheffield 6/9/15 4:52 PM
Deleted: based

Justin Sheffield 6/9/15 4:53 PM
Deleted: on

Justin Sheffield 6/9/15 4:53 PM
Deleted: time series

Justin Sheffield 6/9/15 4:54 PM
Deleted: data

Justin Sheffield 6/9/15 4:52 PM
Formatted: Font:Italic

Justin Sheffield 6/9/15 4:52 PM
Formatted: Font:Italic, Subscript

Justin Sheffield 6/9/15 4:54 PM
Deleted: .

4.1. Trends in low flows

We identified a time period (1951-2005) common to all sites for which they have useable data, and calculated statistics of Q_n, including the trend, and the consistency of trends among Q_n values. The MK trends for Q_n for the sites that were categorized as 1, 2, or 3 by the decomposition algorithm are shown in Figure 6a. The sites with significant trends tend to occur in all Q_n (e.g. the sites in Florida). Sites with lower trend magnitudes tend to become non-significant (MK=0) as we move from Q_1 to Q_{90} (e.g. the two sites in the northeast in Maine). Some sites to the east of the Mississippi River do not have significant trends for Q_1 but show a significant decreasing trend for Q_{90}. Overall, the northeastern sites show increasing trends in low flows and the southeast sites show decreasing trends.

A summary of the consistency of trends across *n*-day low flows is shown in Figure 6b. 208 sites (41% of the sites) have the same trend, such that the Q_n series are all increasing, decreasing, or not changing. 162 sites (32%) agree on the sign of trend for three out of four of the Q_n trends, and 87 sites (17%) agree for 2 out of 4 of the Q_n trends. Overall, the consistency in trends among the Q_n series is generally uniformly distributed across the domain.

Figure 7 (top left) shows the spatial pattern of the MK trend test values for Q_7, for all sites (without testing for step changes or autocorrelation), and when we only consider sites without step changes (top right). In both cases, the pattern of increasing trend in low flows in the northeast and a decreasing trend in the southeast is apparent. However, ignoring the effect of autocorrelation may give rise to misleading results by showing a denser pattern of significant trends. The bottom left panel shows the results removing sites with step changes and pre-whitening the data for the remaining sites. The bottom right panel show the trends when sites

Justin Sheffield 7/18/15 11:06 AM
Deleted: 9

Justin Sheffield 9/12/15 11:53 AM
Formatted: Font:Not Italic

Justin Sheffield 7/20/15 5:32 PM
Deleted: *,*

Justin Sheffield 9/12/15 11:53 AM
Formatted: Not Highlight

Justin Sheffield 9/12/15 11:53 AM
Formatted: Not Highlight

Justin Sheffield 9/12/15 11:50 AM
Deleted: when autocorrelation is ignored

Justin Sheffield 9/12/15 11:51 AM
Deleted: left panel

Justin Sheffield 9/12/15 11:52 AM
Deleted: and if we remove sites with significant autocorrelation (top right panel)

Justin Sheffield 9/12/15 11:54 AM
Deleted: ~~The gray colored sites in the right panel are those with significant autocorrelation.~~

Justin Sheffield 9/12/15 11:53 AM
Formatted: Not Highlight

Justin Sheffield 9/12/15 11:53 AM
Formatted: Strikethrough

Justin Sheffield 9/12/15 11:53 AM
Formatted: Not Highlight

that have USGS flags are also excluded, e.g. for sites without documented anthropogenic impacts. The attribution of trends at these sites is therefore likely related to climate variability/change and/or land use change, rather than management of flows.

Justin Sheffield 6/15/15 2:33 PM
Deleted: i.e

Justin Sheffield 7/16/15 10:06 PM
Deleted: a direct anthropogenic impact on

Justin Sheffield 7/16/15 10:06 PM
Deleted: the low flows

4.2. Variability in low flow timing

Figure 8 summarizes the distribution of the onset of the low flow season for Q_1, where the first season is defined as the 4-month period that contains the majority of low flow occurrences (top panels) and the second season as the 4-month period that contains the majority of the remaining low flows (bottom panels). The left panels show the onset month of the season and the right panels show the probability of the onset season in that month. If the onset time of the low flow season for a site occurs 70% to 100% in a specific month, that site is assumed to have only one low flow season. For Q_1, 353 sites out of 395 (almost 90%) sites fit in this category. For sites with one low flow season, the onset of the season changes from north to south. Most of the sites north of North Carolina have low flow seasons starting in July, which is generally driven by the slight decline in precipitation during the autumn as well as the increased evaporation during the summer (Small et al., 2006). In Florida the season starts in April-May. For coastal sites, the season starts earlier (mostly in June), and for sites in the southwestern part of the domain, the season starts mostly in September-October.

The sites that have low flow events occurring 40-70% of the time in one month and 20-40% of the time in a different month are characterized as having two low flows seasons. These sites are mostly in Florida, and along the coastline of Georgia, South and North Carolina, New York, New Jersey, and Maine and their second season occurs mostly in fall. For New York, New Jersey, and some sites along the west coastline of Florida, the second low flow season mostly starts in November and December. Sites near the Gulf of Mexico and some sites in North Carolina have second low flow seasons starting in April. The second low flow season for the far northeast sites begins in December or January and can be related to freezing conditions that may store water as snow and river ice.

4.3. Changes in low flow timing

To determine whether low flow timing has changed over time, we examined sites with one low flow season as defined as 70% of low flow occurrences in the same season, again for the common time period of 1951-2005. Analysis of changes in timing irrespective of the season (not shown) did not show evidence of shifts in timing from one season to another. For Q_7, for example, 47 sites out of the total 508 were removed because their low flow season occurs less than 70% of the time in one season. Out of the remaining 467 sites, 20 sites showed a decreasing (earlier) trend in timing and were mostly in Pennsylvania and the Carolinas (Figure 9) and 14 showed an increasing (later) trend with most of these in the northeast. The MK test for Q_{30} timings showed mainly decreasing (earlier) trends (26 sites), with most overlap with the Q_7 results in Pennsylvania. These sites have low flow seasons starting in July, and half of them are regulated or partially regulated. Only a few sites were identified by the Pettitt test (5% significance) to have a significant step change in either direction.

The tendency for low flows (Q_7 and Q_{30}) to occur earlier in the season in recent years may be because of a shift of low precipitation from the late to mid summer, but given the small number of sites with significant trends and their low spatial coherence, this is speculative. Although the sites in Pennsylvania did not show a trend in low flow volumes, the overall trend for the northeast is an increasing trend in low flow volumes suggesting that early summer low

Justin Sheffield 6/9/15 4:47 PM
Formatted ... [1]

Justin Sheffield 7/20/15 5:02 PM
Deleted: Both the MK trend and Petti ... [2]

Justin Sheffield 7/20/15 5:01 PM
Deleted: ,…395 sites were initially ... [3]

Justin Sheffield 9/12/15 12:18 PM
Deleted: 3… and an additional 17 site ... [4]

Justin Sheffield 9/12/15 12:41 PM
Formatted ... [5]

Justin Sheffield 9/12/15 12:21 PM
Deleted: 13 …and New York states ... [6]

Justin Sheffield 7/20/15 5:04 PM
Formatted ... [7]

Justin Sheffield 9/12/15 12:21 PM
Deleted: 335

Justin Sheffield 7/20/15 5:05 PM
Formatted ... [8]

Justin Sheffield 9/12/15 12:23 PM
Deleted: 11 sites with a …of which 8 ... [9]

Justin Sheffield 7/20/15 5:05 PM
Formatted ... [10]

Justin Sheffield 7/20/15 5:05 PM
Formatted ... [11]

Justin Sheffield 7/20/15 5:05 PM
Deleted: ,… All 8 sites had ... [12]

Justin Sheffield 7/20/15 5:05 PM
Formatted ... [13]

Justin Sheffield 9/12/15 12:26 PM
Deleted: . Half of the 8 sites

Justin Sheffield 7/20/15 5:05 PM
Formatted ... [14]

Justin Sheffield 9/12/15 12:27 PM
Deleted: For Q_7, 5 of the sites were identifi

Justin Sheffield 7/20/15 5:06 PM
Formatted ... [15]

Justin Sheffield 9/12/15 12:27 PM
Deleted: ed…downward … (Figure ... [16]

Justin Sheffield 9/12/15 12:40 PM
Formatted ... [17]

Justin Sheffield 9/12/15 12:29 PM
Deleted: Overall, the results from th ... [18]

Justin Sheffield 9/12/15 12:36 PM
Deleted: Interestingly, these sites sh ... [19]

Justin Sheffield 9/12/15 12:40 PM
Formatted ... [20]

Justin Sheffield 9/12/15 12:38 PM
Deleted: ese 8

Justin Sheffield 9/12/15 12:40 PM
Formatted ... [21]

precipitation might also be increasing. More investigation is required to confirm whether low precipitation is happening earlier in summer, for example during May and June, and whether the amount is increasing.

5. Discussion and Conclusions

5.1. Potential Drivers of Trends in Low Flows

Justin Sheffield 6/9/15 4:47 PM
Deleted: Attribution

We found spatially coherent patterns of increases in low flows in the northeast and decreases in the southeast, which was robust to the presence of USGS flags and autocorrelation in the time series, despite the smaller number of sites. The pattern of increasing low flows in the northeast is consistent with regional scale studies (e.g. Hodgkins and Dudley, 2011) and are consistent with the increases in 7-day low flows and fall precipitation shown in Small et al. (2006) that focused on a smaller set of sites across the eastern U.S. from the HCDN. Several other studies (e.g. Douglas et al., 2000; McCabe and Wolock, 2002; Hayhoe et al., 2007; Andreadis and Lettenmaier, 2006) have identified an overall increasing trend in precipitation over the past 50 years, and a decreasing pattern in soil moisture drought over the much of the U.S. including the northeast (Andreadis and Lettenmaier, 2006). Therefore, an increase in low flow volumes in the northeast is consistent with the overall shift to wetter conditions. The generally decreasing trends in the southeast are also consistent with the results from Small et al. (2006) and Lins and Slack (1999), which is despite an overall increase in precipitation in the region.

To understand the attribution of these trends more comprehensively, Figure 10 shows the Q_7 trend magnitude and the antecedent precipitation for the previous 180 days. This period was chosen as it provides the highest correlation with low flow volumes (Kam et al, 2015), although the results with 150 and 90 days are similar. The precipitation data are taken from the long-term precipitation dataset of Livneh et al. (2013) and are averaged over the basin corresponding to each site. The similarity between the trends in low flows and antecedent precipitation is striking with a clear increasing trend in the north and decrease in the south, although many of the trends are not statistically significant.

The main disparity is in coastal plains of eastern Virginia, Maryland and northwards to Maine, where Q_7 low flows have decreased but antecedent precipitation is increasing (both often statistically significant). The reason for this is unclear, but groundwater is likely playing a role across the coastal plain aquifer of the mid-Atlantic states and up into New England (Dudley and Hodgkins, 2013) either via changes in recharge or indirectly through anthropogenic impacts. Groundwater pumping has reduced levels in the north Atlantic Coastal Plain aquifer system by tens of meters (e.g. Konikow, 2013, USGS, 2006) and has likely reduced discharge to streams in the northeast (e.g. Pucci and Pope, 1995; Brutsaert, 2010; Barlow and Leake, 2012). Similarly, overuse of groundwater resources in the southeast (Konikow, 2013) may be contributing to decreases in low flows across the region (e.g. Bosch et al., 2003; Opsahl et al., 2007; Brutsaert, 2010).

Increases in evaporation (Walter et al., 2004; Nolan et al., 2007; Huntington and Billmire, 2014) may have also led to declines in groundwater recharge and streamflow (Hodgkins and Dudley, 2011), and potentially cancelled out the overall increases in precipitation across much of the U.S. (Andreadis and Lettenmaier, 2006). Figure 10 also shows an estimate of the trend in late summer/early fall potential evaporation based on the NLDAS2 dataset of Xia et al. (2012). Potential evaporation has increased over the eastern U.S. with statistically significant trends over much of the mid-Atlantic states and the southeast. This suggests that increasing atmospheric demand in the southeast may have exacerbated declines in low flows, and this may have offset increasing precipitation somewhat in the northeast. Changes in land use may also explain trends in both regions, whereby land abandonment in the northeast and forest harvesting and urban development in the southeast may have contributed to the respective trends in each region (Cho et al., 2009; Payne et al., 2005; USGS, 2012), although attribution is difficult.

The analysis of trends in timing of low flows showed one cluster of sites with a trend to earlier timing. These sites are mostly in central and west Pennsylvania, and central southern New York. The reasons for the changes are unclear, but may be related to regulation and possibly a shift in the low precipitation season to earlier in the summer. The timing of low flows in the other parts of the domain has not changed based on a 5% significance level.

5.2. Conclusions

This study has examined the presence of non-stationarity in low flows across the eastern U.S. in terms of volumes and timing. We focused on the full period of available data at each site to identify abrupt shifts that may be associated with management, in particular dam construction, and gradual trends that may be an impact of climate change, land use change or surface/ground water withdrawals. A decomposition algorithm was used to identify useable sub-series of the data that could then be further analyzed for trends. 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, 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

Justin Sheffield 6/3/15 2:51 PM
Deleted: also

Justin Sheffield 6/15/15 1:41 PM
Deleted: with

Justin Sheffield 6/15/15 1:57 PM
Deleted:

Justin Sheffield 6/15/15 1:58 PM
Deleted: associated

Justin Sheffield 6/15/15 1:58 PM
Deleted: with

Justin Sheffield 6/15/15 1:59 PM
Deleted: or no trends

Justin Sheffield 6/15/15 1:59 PM
Deleted: the minority

Justin Sheffield 6/3/15 2:52 PM
Deleted: ve

Justin Sheffield 6/15/15 1:59 PM
Formatted: Font:Not Italic

Justin Sheffield 6/15/15 1:59 PM
Formatted: Font:Not Italic

influenced the low flow time series. This implies that our approach is generally capable of identifying sites with documented regulation or a change in measurement method, but that these changes do not always manifest in a detectable change in the low flow time series. This may be because the documented regulation or other change may not have an impact or that the signal is small compared to the variability in the time series. This is particularly the case for higher low flow metrics such as Q_{90}, for which the regulation is generally less detectable. 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. For example, flows are often artificially elevated above the natural levels of low flow to create ``anti-droughts" to manage the restoration of river systems (Bunn et al., 2006).

Justin Sheffield 6/3/15 2:59 PM
Deleted: it may be that the USGS flags relate more to high flows than low flows

Several outstanding questions remain, most importantly what are the low flow generating mechanisms across the eastern U.S. and what are the drivers of long-term changes in the volumes and timing. Potential mechanisms include, but are not limited to: changes in antecedent precipitation and teleconnections with large-scale climate (e.g. the North Atlantic Oscillation; Kam et al., 2015), land use change, surface and groundwater abstraction, and streamflow regulation. The results of this study suggest that low flow variability in the eastern U.S. is driven by a mixture of climatic and anthropogenic effects, with suggestions that changes in climate have played a role in both the northeast and southeast. However, definitive attribution will require detailed analysis of these competing factors and possibly carefully crafted modeling studies.

The results of this study can help in understanding changes in low flows across the eastern U.S., and the impact of anthropogenic and natural changes. It can therefore provide information for water management, and restoration of stream flows and aquatic habitats. The

Justin Sheffield 7/17/15 6:57 PM
Deleted: help

methods are readily transferable to other parts of the U.S. and globally, given long enough time series of daily streamflow data, although further work is required to understand their universal application.

Author Contribution

S. S. and J. S. conceived the study. S. S. performed the analysis with help from J. K. S. S. prepared the manuscript with contributions from the other authors.

Acknowledgements

This work was supported by the USGS (G11AP20215) and NOAA (NA14OAR4310130 and NA14OAR4310218)

References

Acreman, M. C., B. Adams, and B. Connorton, 2000: Does groundwater abstraction cause degradation of rivers and wetlands? *Water and Environment Journal*, 14, 200–206.

Andreadis, K. M., and D. P. Lettenmaier, 2006: Trends in 20th century drought over the continental united state. *Geophys. Res. Letts.*, 33, 1-4.

Averyt, K., J. Meldrum, P. Caldwell, G. Sun, S. McNulty, A. Huber-Lee, and N. Madden, 2013: Sectoral contributions to surface water stress in the coterminous United States. *Environ. Res. Lett.*, 8, 035046 (9pp). doi:10.1088/1748-9326/8/3/035046

Barlow, P. M., and S. A. Leake, 2012: Streamflow depletion by wells—Understanding and managing the effects of groundwater pumping on streamflow. *U.S. Geological Survey Circular*, 1376, 84p.

Bosch, D. D., Lowrance, R. R., Sheridan, J. M. and Williams, R. G., 2003: Ground water storage effect on streamflow for a Southeastern Coastal Plain watershed. *Ground Water*, 41, 903–912. doi: 10.1111/j.1745-6584.2003.tb02433.x

Bradford, M., and J. Heinonen, 2008: Low flows, instream flow needs and fish ecology in small streams. *Canadian Water Resources Journal*, 33, 165–180.

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

Brown, T. C., R. Foti, and J. A. Ramirez, 2013: Projected freshwater withdrawals in the United States under a changing climate. *Water Resour. Res.*, 49, 1259–1276, doi:10.1002/wrcr.20076.

Justin Sheffield 7/17/15 2:04 PM
Formatted: Font:Italic

Unknown
Deleted:

Justin Sheffield 6/15/15 2:41 PM
Formatted: Font:Italic

Brutsaert, W., 2010: Annual drought flow and groundwater storage trends in the eastern half of the United States during the past two-third century. *Theoretical and Applied Climatology,* 100 (1-2), 93-103.

Bunn, S. E., M. C. Thoms, S. K. Hamilton, and S. J. Capon, 2006: Flow variability in dryland rivers: boom, bust and the bits in between. *River Research and Applications*, 22 (2), 179186.

Cho, J., V. A. Barone, and S. Mostaghimi, 2009: Simulation of land use impacts on groundwater levels and streamflow in a Virginia watershed. *Agricultural Water Management*, 96 (1), 1-11

Cohn, T. A., and H. F. Lins, 2005: Nature's style: Naturally trendy. *Geophys. Res. Lett.*, 32, L23402, doi:10.1029/2005GL024476

Colby, F. P., 2008: Mesoscale snow bands in an ocean-effect snowstorm. *Tech. rep., University of Massachusetts Lowell*, Lowell, MA 01854.

Cox, D. R., and A. Stuart, 1955: Some quick sign tests for trend in location and dispersion. *Biometrika*, 42(1/2), 80–95.

Daniel, C., and P. Dahlen, 2002: Preliminary hydrogeologic assessment and study plan for a regional ground-water resource investigation of the Blue Ridge and Piedmont Provinces of North Carolina, Investigations Report 02-4105, USGS.

Deitch, M. J., G. M. Kondolf, and A. M. Merenlender, 2009: Hydrologic impacts of small-scale instream diversions for frost and heat protection in the California wine country. *River Research and Applications*, 25, 118-134.

Douglas, E. M., R. M. Vogel, and C. N. Kroll, 2000: Trends in floods and low flows in the United States: impact of spatial correlation. *J. Hydrology*, 240, 90–105.

Justin Sheffield 7/20/15 7:32 PM
Formatted: Font:Italic

Unknown
Deleted:

Justin Sheffield 7/17/15 2:05 PM
Formatted: Font:Times New Roman, 12 pt

Justin Sheffield 7/17/15 2:05 PM
Formatted: Font:Times New Roman, 12 pt

Justin Sheffield 7/17/15 2:05 PM
Formatted: Font:Times New Roman, 12 pt

Justin Sheffield 7/17/15 2:05 PM
Formatted ... [22]

Justin Sheffield 7/17/15 2:05 PM
Formatted ... [23]

Justin Sheffield 7/17/15 2:05 PM
Formatted ... [24]

Justin Sheffield 7/17/15 2:05 PM
Formatted ... [25]

Justin Sheffield 7/17/15 2:05 PM
Formatted ... [26]

Justin Sheffield 7/17/15 2:05 PM
Formatted ... [27]

Justin Sheffield 7/17/15 2:05 PM
Formatted ... [28]

Justin Sheffield 7/17/15 2:05 PM
Formatted ... [29]

Justin Sheffield 7/17/15 2:05 PM
Formatted ... [30]

Justin Sheffield 7/17/15 2:05 PM
Formatted ... [31]

Justin Sheffield 7/17/15 2:05 PM
Formatted ... [32]

Justin Sheffield 7/17/15 2:05 PM
Formatted ... [33]

Justin Sheffield 7/17/15 2:05 PM
Formatted ... [34]

Justin Sheffield 7/17/15 2:05 PM
Formatted ... [35]

Justin Sheffield 7/17/15 2:05 PM
Formatted ... [36]

Justin Sheffield 7/17/15 2:05 PM
Formatted ... [37]

Justin Sheffield 7/17/15 2:05 PM
Formatted ... [38]

Justin Sheffield 7/17/15 2:05 PM
Formatted: Font:Times New Roman

Dudley, R.W., and G. A. Hodgkins, 2013: Historical groundwater trends in northern New England and relations with streamflow and climatic variables. *J. American Water Resources Association*, 49 (5), 1198-1212.

EPA Climate Change Division, 2008: Precipitation and storm changes, url: www.epa.gov/climatechange/science/indicators/weather-climate/index.html.

Giuntoli, I., B. Renard, J.-P. Vidal, and A. Bard, 2013: Low flows in France and their relationship to large-scale climate indices. *J. Hydrology*, 482, 105–118.

Gurka, J. J., E. P. Auciello, A. F. Gigi, J. S. Waldstreicher, K. K. Keeter, S. Businger, and L. G. Lee, 1995: Winter weather forecasting throughout the eastern United States. Part II: An operational perspective of cyclogenesis. *American Meteo.*, 10, 21–41.

Hayhoe, K., C. P. Wake, T. G. Huntington, L. Luo, M. D. Schwartz, J. Sheffield, E. Wood, B. Anderson, J. Bradbury, A. Degaetano, T. J. Troy, and D. Wolfe, 2007: Past and future changes in climate and hydrological indicators in the U.S. Northeast. *Climate Dynamics*, 28, 381–407.

Hodgkins, G. A., R. W. Dudley, and T. G. Huntington, 2005: Changes in the number and timing of days of ice-affected flow on northern New England rivers. *Climatic Change*, 71 (3), 319-340. doi: 10.1007/s10584-005-5926-z

Hodgkins, G. A., and R. W. Dudley, 2011: Historical summer base flow and stormflow trends for New England rivers. *Water Resources Research*, 47 (7), W07528

Huntington, T. G., and M. Billmire, 2014: Trends in precipitation, runoff, and evapotranspiration for rivers draining to the Gulf of Maine in the United States. *J. Hydrometeor.,* 15(2), 726-743.

Kam, J., J. Sheffield, X. Yuan, and E. F. Wood, 2013: The influence of Atlantic tropical cyclones on drought over the Eastern U.S. (1980-2007). *J. Climate*, 26 (10), 3067-3086. http://dx.doi.org/10.1175/JCLI-D-12-00244.1.

Kam, J., S. Sadri, and J. Sheffield, 2014: Changes in the low flow regime over the Eastern United States (1962-2011): Variability, trends, and attributions, submitted to *Climatic Change*.

Karl, T. R., and R. W. Knight, 1998: Secular trends of precipitation amount, frequency, and intensity in the United States. *Bull. Am. Meteorol. Soc.*, 79, 231–241

Kendall, M. G., 1975: Rank Correlation Methods. Charles Griffin, 202 pp.

Kendall, M., A. Stuart, and J. K. Ord, 1983: The Advanced Theory of Statistics, Vol. 3, Design and Analysis, and Time Series, 4th ed., 780 pp., Oxford Univ. Press, New York.

Konikow, L. F., 2013: Groundwater depletion in the United States (1900–2008): U.S. Geological Survey Scientific Investigations Report 2013–5079, 63 p., http://pubs.usgs.gov/sir/2013/5079

Koutsoyiannis, D., 2011: Hurst-Kolmogorov dynamics and uncertainty. *J. American Water Resources Association*, 47 (3), 481-495

Kroll, C., J. Luz, B. Allen, and R. M. Vogel, 2004: Developing a watershed characterisitics database to improve low streamflow prediction. *J. Hydrologic Engineering*, 9 (2),116–125.

Kroll, C. N., and R. M. Vogel, 2002: Probability distribution of low streamflow series in the United States. *J. Hydrologic Engineering*, 7 (2), 137–146.

Kustu, M. D., Y. Fan, and A, Robock, 2010: Large-scale water cycle perturbation due to irrigation pumping in the US High Plains: A synthesis of observed streamflow changes. *J. Hydrol.*, 390 (3-4), 222-244. doi:10.1016/j.jhydrol.2010.06.045

Unknown
Deleted:

Justin Sheffield 7/20/15 9:49 AM
Formatted: Font:Times New Roman, 12 pt

Justin Sheffield 7/20/15 9:49 AM
Formatted: Normal (Web), Indent: Hanging: 0.25", Space Before: 0.1 pt, After: 0.1 pt, Line spacing: double

Justin Sheffield 7/20/15 9:49 AM
Formatted: Font:Times New Roman

Unknown
Deleted:

Justin Sheffield 7/20/15 9:49 AM
Formatted: Font:Times New Roman, 12 pt

Justin Sheffield 7/20/15 9:49 AM
Formatted: Normal (Web), Indent: Hanging: 0.25", Space Before: 0.1 pt, After: 0.1 pt, Line spacing: double

Justin Sheffield 7/20/15 9:49 AM
Formatted: Font:Times New Roman

Unknown
Deleted:

Justin Sheffield 7/17/15 2:05 PM
Formatted: Font:(Default) Times New Roman

Lins, H. F. and T. A. Cohn, 2011: Stationarity: Wanted Dead or Alive?. *J. American Water Resources Association*, 47: 475-480. doi: 10.1111/j.1752-1688.2011.00542.x

Lins, H., and J. R. Slack, 1999: Streamflow trends in the United States. *Geophys. Res. Letts.*, 26, 227–230.

Lins, H. F., 2012: USGS Hydro-Climatic Data Network 2009 (HCDN–2009): U.S. Geological Survey Fact Sheet 2012–3047, 4 p., available at http://pubs.usgs.gov/fs/2012/3047/.

Livneh, B., E. A. Rosenberg, C. Lin, B. Nijssen, V. Mishra, K. M. Andreadis, E. P. Maurer, and D. P. Lettenmaier, 2013: A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States: Update and extensions. *J. Climate*, 26, 9384–9392.

Ljung, G. and G. Box, 1978: On a Measure of Lack of Fit in Time Series Models. *Biometrika*, 65, 297-303.

Mann, H. B., 1945: Nonparametric tests against trend. *Econometrica*, 13, 245–259.

Mauget, S. A., 2003: Multidecadal regime shifts in U.S. streamflow, precipitation, and temperature at the end of the twentieth century. *J. Climate*, 16, 3905–3916.

McCabe, G. J., and D. M. Wolock, 2002: A step increase in streamflow in the conterminous United States. *Geophys. Res. Letts.*, 29 (24), 1–4.

Milly, P. C. D., J. Betancourt, M. Falkenmart, R. M. Hirsch, Z. W. Kundzewicz, D. P. Lettenmaier, and R. J. Stouffer, 2008: Stationarity is dead: Whither water management? *Science*, 319 (5863), 573–574.

Justin Sheffield 7/20/15 9:50 AM
Formatted: Font:12 pt

Justin Sheffield 7/20/15 9:50 AM
Formatted: Normal (Web), Indent: Hanging: 0.25", Line spacing: double

Justin Sheffield 6/15/15 2:25 PM
Deleted:

Justin Sheffield 6/15/15 2:25 PM
Formatted: Font:Italic

Justin Sheffield 7/17/15 2:52 PM
Formatted: Font:12 pt

Justin Sheffield 7/17/15 2:52 PM
Formatted: Normal (Web), Indent: Hanging: 0.25", Line spacing: double

Justin Sheffield 7/17/15 2:52 PM
Formatted: Font:12 pt, Italic

Justin Sheffield 7/17/15 2:52 PM
Formatted: Font:12 pt

Opsahl, S. P., S. E. Chapal, D. W. Hicks, and C. K. Wheeler, 2007: Evaluation of ground-water and surface-water exchanges using streamflow difference analyses. *J. American Water Resources Association*, 43, 1132–1141. doi: 10.1111/j.1752-1688.2007.00093.x

Payne, D. F., M. A. Rumman, and J. S. Clarke, 2005: Simulation of ground-water flow in coastal Georgia and adjacent parts of South Carolina and Florida—Predevelopment, 1980, and 2000: U.S. Geological Survey, Scientific Investigations Report 2005-5089, 91p.

Pettitt, A. N., 1979: A non-parametric approach to the change-point problem. *Applied Statistics*, 28, 126–135.

Poff, N. L., J. D. Allan, M. B. Bain, J. R. Karr, K. L. Prestegaard, B. D. Richter, R. E. Sparks, and J. C. Stromberg, 1997: The natural flow regime: a paradigm for river conservation and restoration. *Bioscience*, 47, 769–784.

Pucci, A. A., and D. A. Pope, 1995: Simulated effects of development on regional ground-water/surface-water interactions in the northern Coastal Plain of New Jersey. *J. Hydrology*, 167 (1–4), 241–262.

Rolls, R. J., C. Leigh, and F. Sheldon, 2012: Mechanistic effects of low-flow hydrology on riverine ecosystems: ecological principles and consequences of alteration. *Freshwater Science*, 31(4), 1163–1186.

Ruppert, D., 2011: Statistics and data analysis for financial engineering, Springer Texts in Statistics, Springer.

Sadri, S., and D. H. Burn, 2011: A Fuzzy C-Means approach for regionalization using a bivariate homogeneity and discordancy approach. *J. Hydrology*, 401, 231–239.

Serinaldi, F., and C. G. Kilsby, 2015: Stationarity is undead: Uncertainty dominates the distribution of extremes. *Advances in Water Resources*, 77, 17-36

Smakhtin, V. U., 2001: Low flow hydrology: a review. *J. Hydrology*, 240, 147–186.

Small, D., S. Islam, and R. M. Vogel, 2006: Trends in precipitation and streamflow in the eastern U.S.: Paradox or preception?. *Geophys. Res. Letts.*, 33, 1–4.

Tallaksen, T., and H. A. Van Lanen, 2004: Hydrological drought, processes and estimation methods for streamflow and groundwater. Vol. 48, 580 pp., Elsevier Science.

Thomas, B., 2006. Trends in Streamflow of the San Pedro River, Southeastern Arizona, U.S. Geological Survey Fact Sheet 2006-3004, 4 pp., http://pubs.usgs.gov/fs/2006/3004/.

U.S. Army Corps of Engineers, 2012: url: http://geo.usace.army.mil/pgis/f?p=397:12

U.S. Senate, 2002: Federal Water Pollution Control Act, url: www.epw.senate.gov/water.pdf.

USDA Forest Service, 2010: Climate Change and Water Issues for the Eastern US, Tech. rep., USDA.

USGS: Sustainability of the Ground-Water Resources in the Atlantic Coastal Plain of Maryland, USGS Fact Sheet, 2006-3009, 2 p., 2006.

USGS, 2009: Groundwater atlas of the United States, url: www.pubs.usgs.gov/ha/ha730/.

USGS, 2012: Land cover change in the eastern United States, url: www.landcovertrends.usgs.gov/east/regionalSummary.html.

Justin Sheffield 9/12/15 10:36 AM
Formatted: Font:Italic

Unknown
Deleted:

Justin Sheffield 7/17/15 2:10 PM
Formatted: Font:Times New Roman, 12 pt

Justin Sheffield 7/17/15 2:10 PM
Formatted: Normal (Web), Indent: Hanging: 0.25", Line spacing: double

Justin Sheffield 7/17/15 2:10 PM
Formatted: Font:Times New Roman, 12 pt

Justin Sheffield 7/17/15 2:10 PM
Formatted: Font:Times New Roman

USGS, 2013: Groundwater depletion in the United States (1900-2008). Scientific Investigation Report 2013-5079, U.S. Department of the Interior, U.S. Geological Survey, Reston, Virginia.

USGS, 2014: National Water Information System data available on the World Wide Web (Water Data for the Nation), url: http://waterdata.usgs.gov/nwis/

Villarini, G., J. A. Smith, F. Serinaldi, J. Bales, P. D. Bates, and W. F. Krajewski, 2009: Flood frequency analysis for nonstationary annual peak records in an urban drainage basin. *Advances in Water Resources*, 32, 1255–1266.

Walker, K. F. and M. C. Thoms, 1993: Environmental effects of flow regulation on the lower river Murray. *Australia. Regul. Rivers: Res. Mgmt.*, 8, 103-119. doi: 10.1002/rrr.3450080114

Walter, M. T., D. S. Wilks, J.-Y. Parlange, and R. L. Schneider, 2004: Increasing evapotranspiration from the conterminous United States. *J. Hydrometeor*., 5, 405–408. doi: http://dx.doi.org/10.1175/1525-7541(2004)005<0405:IEFTCU>2.0.CO;2

WMO, 2008: Manual on low-flow estimation and prediction. Operational Hydrology Report 30, WMO-N 1029. Geneva, Switzerland

Unknown
Deleted:

Justin Sheffield 7/17/15 2:06 PM
Formatted: Font:Times New Roman, 12 pt

Justin Sheffield 7/17/15 2:06 PM
Formatted: Normal (Web), Indent: Hanging: 0.25", Line spacing: double

Justin Sheffield 7/17/15 2:06 PM
Formatted: Font:Times New Roman

Unknown
Deleted:

Justin Sheffield 6/9/15 4:55 PM
Formatted: Indent: Left: 0", Hanging: 0.25", Space Before: 0.01 line, After: 0.01 line

Table 1. Comparison of the number of streamflow gauging sites in each category of the decomposition algorithm and their USGS flags for Q_7. DamFail: dam failure; RegPar: partially regulated; Reg: regulated; Below: flow below rating curve limit; Urban: affected by urbanization; ChangeDis: Change base discharge; ChangeDatH: change gauge datum.

Category	Q_7	Q_{30}	Flag	Q_7	Q_{30}	Flag type	Q_7	Q_{30}
No Trend	240	260	Flagged	121	126	DamFail	1	2
						RegPar	33	37
						Reg	51	48
						Below	0	0
						Urban	2	4
						ChangeDis	2	3
						ChangeDatH	32	32
			Not flagged	119	134			
Decreasing Trend	62	61	Flagged	22	19	DamFail	0	0
						RegPar	3	1
						Reg	5	5
						Below	0	1
						Urban	0	0
						ChangeDis	1	0
						ChangeDatH	11	12
			Not Flagged	40	42			
Increasing Trend	55	70	Flagged	33	48	DamFail	8	0
						RegPar	8	13
						Reg	15	24
						Below	0	0
						Urban	0	0
						ChangeDat	0	0
						ChangeDatH	10	11
			Not Flagged	22	22			
Step Change	111	89	Flagged	72	60	DamFail	1	0
						RegPar	21	16
						Reg	38	32
						Below	1	0
						Urban	4	5
						ChangeDis	1	1
						ChangeDatH	6	6
			Not Flagged	40	29			
Autocorrelated	38	27	Flagged	22	17	DamFail	1	1
						RegPar	4	2
						Reg	7	7
						Below	0	0
						Urban	1	0

Justin Sheffield 9/12/15 12:41 PM
Formatted ... [39]

Justin Sheffield 9/12/15 10:48 AM
Deleted: $_1$ and Q_{90}

Justin Sheffield 9/12/15 11:19 AM
Deleted: L

Justin Sheffield 9/12/15 12:41 PM
Formatted ... [40]

Justin Sheffield 9/12/15 11:20 AM
Deleted: at

Justin Sheffield 9/12/15 11:43 AM
Formatted Table ... [41]

Justin Sheffield 9/12/15 10:48 AM
Deleted: $_1$ ... [42]

Justin Sheffield 9/12/15 10:51 AM
Deleted: 221… ... [43]

Justin Sheffield 9/12/15 11:27 AM
Deleted: 2 ... [44]

Justin Sheffield 9/12/15 11:28 AM
Deleted: 32 ... [45]

Justin Sheffield 9/12/15 11:37 AM
Deleted: 5

Justin Sheffield 9/12/15 11:20 AM
Deleted: at

Justin Sheffield 9/12/15 11:28 AM
Deleted: 0 ... [46]

Justin Sheffield 9/12/15 11:42 AM
Deleted: 121… ... [47]

Justin Sheffield 9/12/15 10:52 AM
Deleted: 36… ... [48]

Justin Sheffield 9/12/15 11:19 AM
Deleted: Low ... [49]

Justin Sheffield 9/12/15 11:35 AM
Deleted: Reg ... [50]

Justin Sheffield 9/12/15 11:20 AM
Deleted: at

Justin Sheffield 9/12/15 11:29 AM
Deleted: 9 ... [51]

Justin Sheffield 9/12/15 11:42 AM
Deleted: 24… ... [52]

Justin Sheffield 9/12/15 10:52 AM
Deleted: 57… ... [53]

Justin Sheffield 9/12/15 11:30 AM
Deleted: 7 ... [54]

Justin Sheffield 9/12/15 11:30 AM
Deleted: 1 ... [55]

Justin Sheffield 9/12/15 11:30 AM
Deleted: 1

Justin Sheffield 9/12/15 11:30 AM
Deleted: 8 ... [56]

Justin Sheffield 9/12/15 11:43 AM
Deleted: 19… ... [57]

Justin Sheffield 9/12/15 10:53 AM
Deleted: 155… ... [58]

Justin Sheffield 9/12/15 11:31 AM
Deleted: 23… ... [59]

Justin Sheffield 9/12/15 11:32 AM
Deleted: 61… ... [60]

Justin Sheffield 9/12/15 11:33 AM
... [61]

Justin Sheffield 9/12/15 11:20 AM
... [62]

Justin Sheffield 9/12/15 11:33 AM
... [63]

Justin Sheffield 9/12/15 11:43 AM
... [64]

Justin Sheffield 9/12/15 11:33 AM

			Not Flagged	16	10	ChangeDis ChangeDatH	2 7	2 5

Justin Sheffield 9/12/15 11:20 AM
Deleted: at

Justin Sheffield 9/12/15 11:34 AM
Deleted: 1

Justin Sheffield 9/12/15 11:40 AM
Deleted: 1

Justin Sheffield 9/12/15 11:34 AM
Deleted: 11

Justin Sheffield 9/12/15 11:40 AM
Deleted: 2

Justin Sheffield 9/12/15 11:43 AM
Deleted: 7

Justin Sheffield 9/12/15 11:42 AM
Deleted: 6

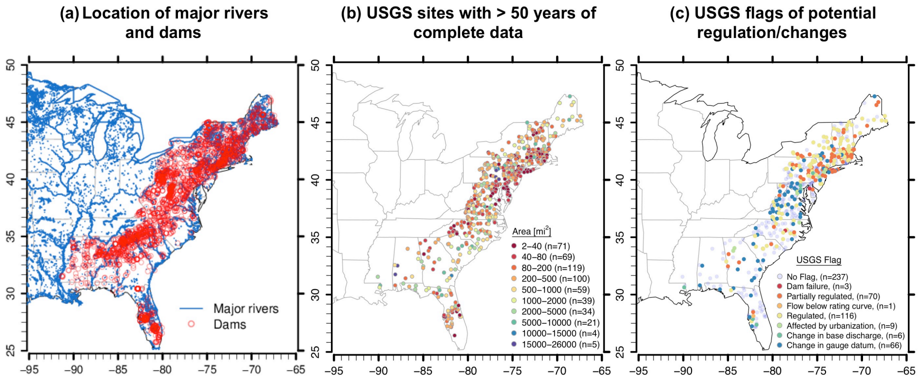

Figure 1. (a) Location of 2,352 major dams in the eastern U.S. (b) Location of the 508 streamflow sites with 50 years or more of complete daily data. (c) Flagged sites according to the USGS.

Justin Sheffield 7/25/15 6:19 PM

Deleted: (b) Location of 2811 USGS streamflow gauging sites that are currently active, or became inactive after the year 2000.

Justin Sheffield 7/25/15 6:19 PM

Deleted: c

Justin Sheffield 7/25/15 6:19 PM

Deleted: d

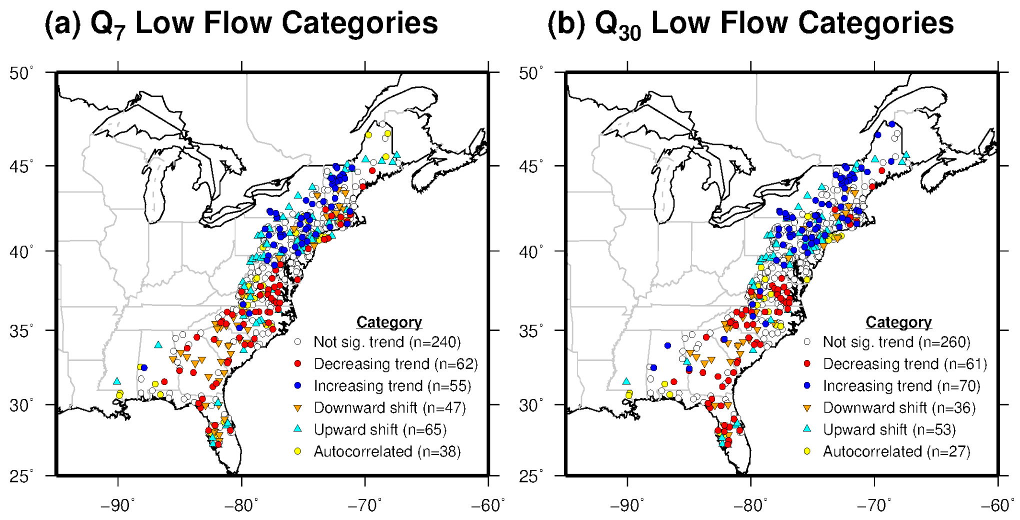

Figure 2. Categorization of non-stationarity of sites for Q_7 and Q_{30}.

Justin Sheffield 7/25/15 6:46 PM

Formatted: Font:12 pt, Not Italic, Not Superscript/ Subscript

Justin Sheffield 7/25/15 6:46 PM

Formatted: Font:12 pt, Not Superscript/ Subscript

Justin Sheffield 7/25/15 6:46 PM

Deleted: „

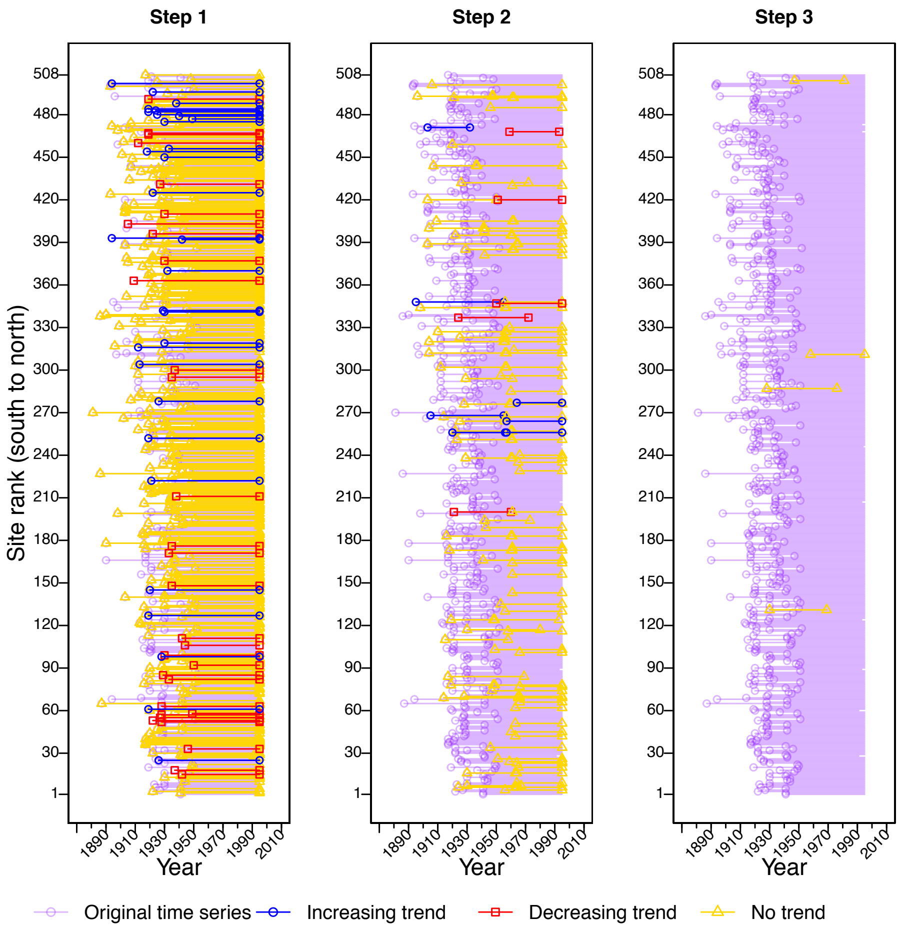

Figure 3. Range of years for each site that are stationary or show a trend, for each step of the decomposition algorithm.

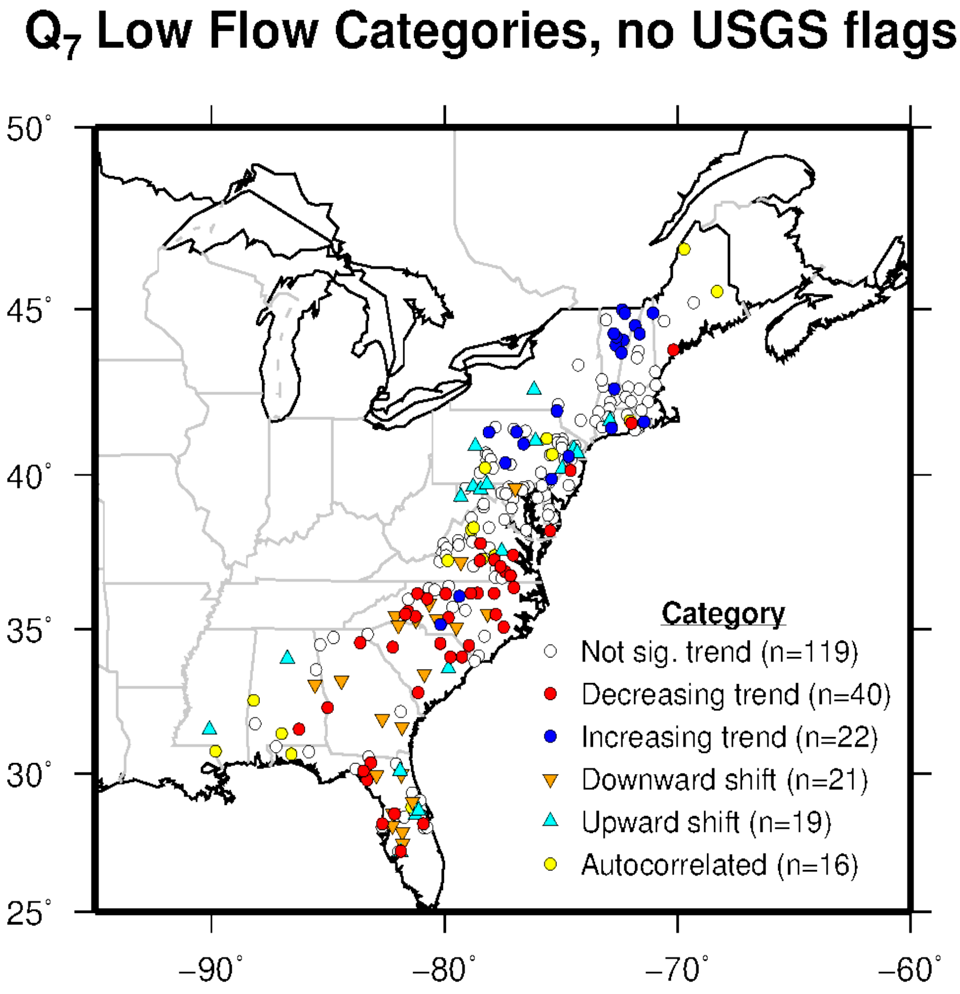

Figure 4. Categorization of non-stationarity of sites for Q_7 with no USGS flags from the first step of the decomposition algorithm.

Justin Sheffield 7/26/15 9:32 AM
Deleted: ,

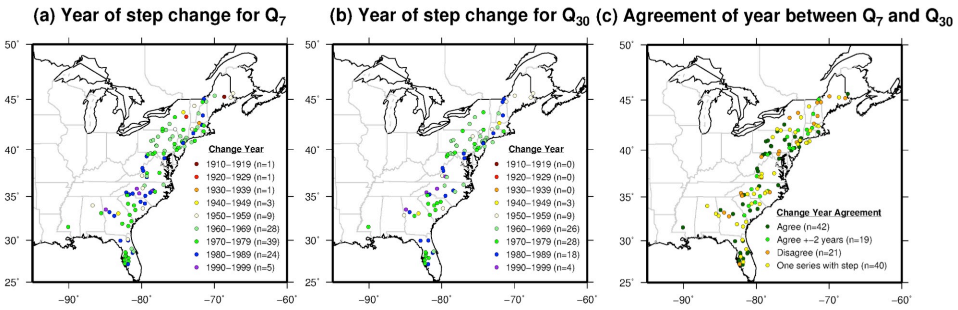
(a) Year of step change for Q_7 (b) Year of step change for Q_{30} (c) Agreement of year between Q_7 and Q_{30}

Figure 5. Year of step change for (a) Q_7 and (b) Q_{30}. (c) Agreement in year of step change between Q_7 and Q_{30} time series.

Justin Sheffield 7/25/15 6:36 PM
Deleted: (a)

Justin Sheffield 7/25/15 6:36 PM
Deleted: a

Justin Sheffield 7/25/15 6:36 PM
Formatted: Font:12 pt, Italic

Justin Sheffield 7/25/15 6:36 PM
Formatted: Font:12 pt, Italic, Subscript

Justin Sheffield 7/25/15 6:36 PM
Formatted: Font:12 pt, Italic

Justin Sheffield 7/25/15 6:36 PM
Formatted: Font:12 pt, Italic, Subscript

Justin Sheffield 7/25/15 6:37 PM
Deleted: across

Justin Sheffield 7/25/15 6:37 PM
Formatted: Font:12 pt, Not Superscript/ Subscript

Justin Sheffield 7/25/15 6:37 PM
Deleted: *n*

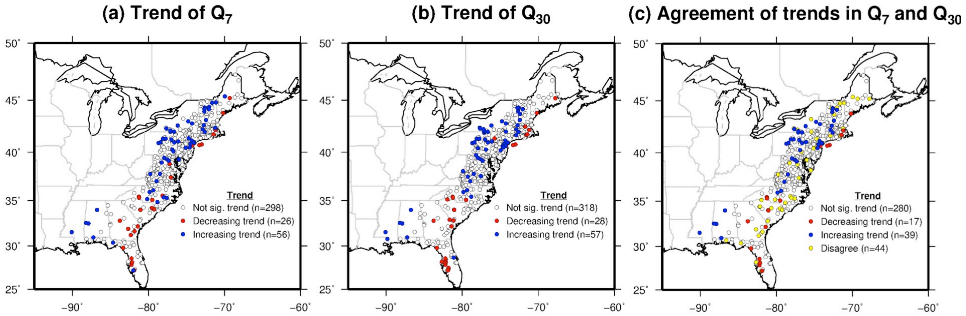

Figure 6. Trends in (a) Q_7 and (b) Q_{30} for 1951-2005 and (c) their agreement.

Justin Sheffield 7/26/15 2:03 PM
Deleted: (a)

Justin Sheffield 7/26/15 9:45 AM
Deleted: $_n$

Justin Sheffield 7/26/15 2:04 PM
Deleted: b

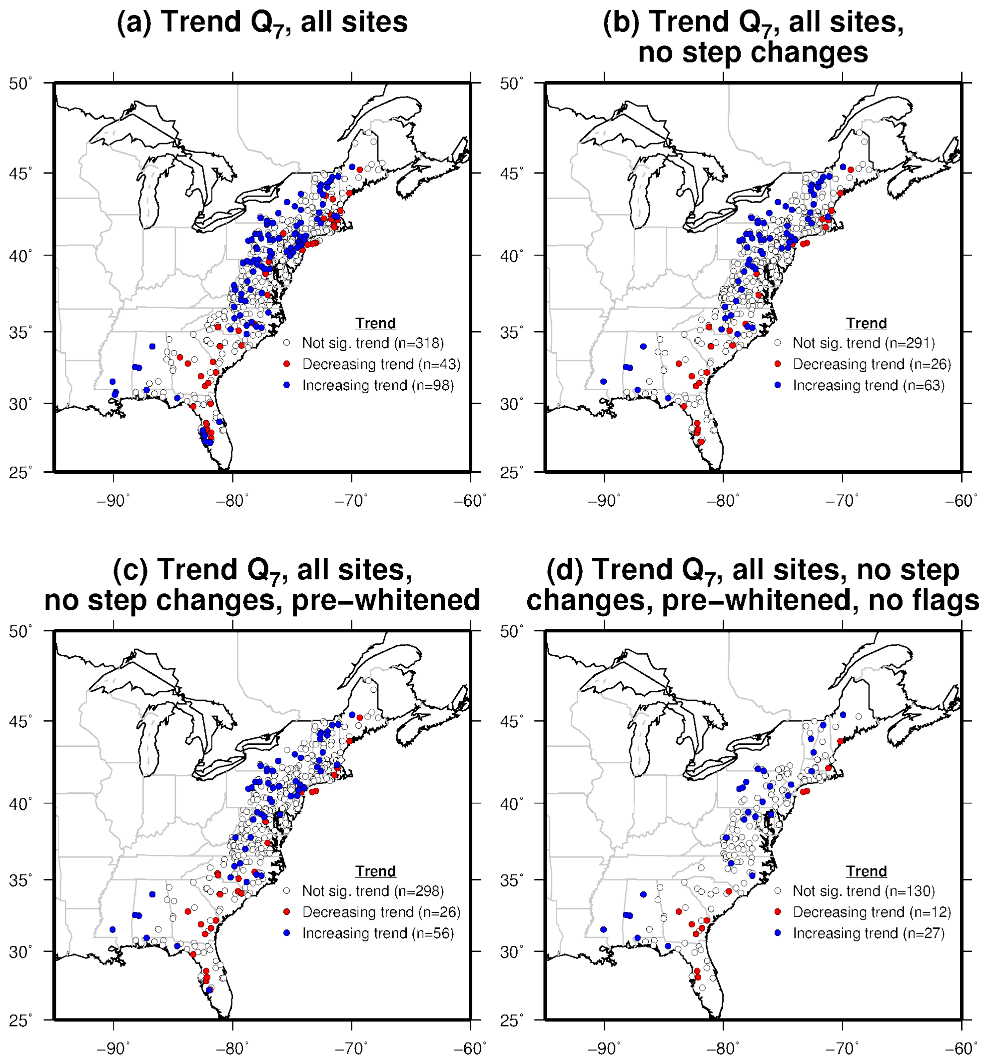

Figure 7. Trends in Q_1 for 1951-2005 for (a) all sites, (b) excluding sites with step changes or overall autocorrelation, (c) as (b) but with pre-whitened data, and (d) as (b) but without USGS flags.

Justin Sheffield 7/26/15 4:35 PM

Deleted: with and without consideration of autocorrelation and USGS flags

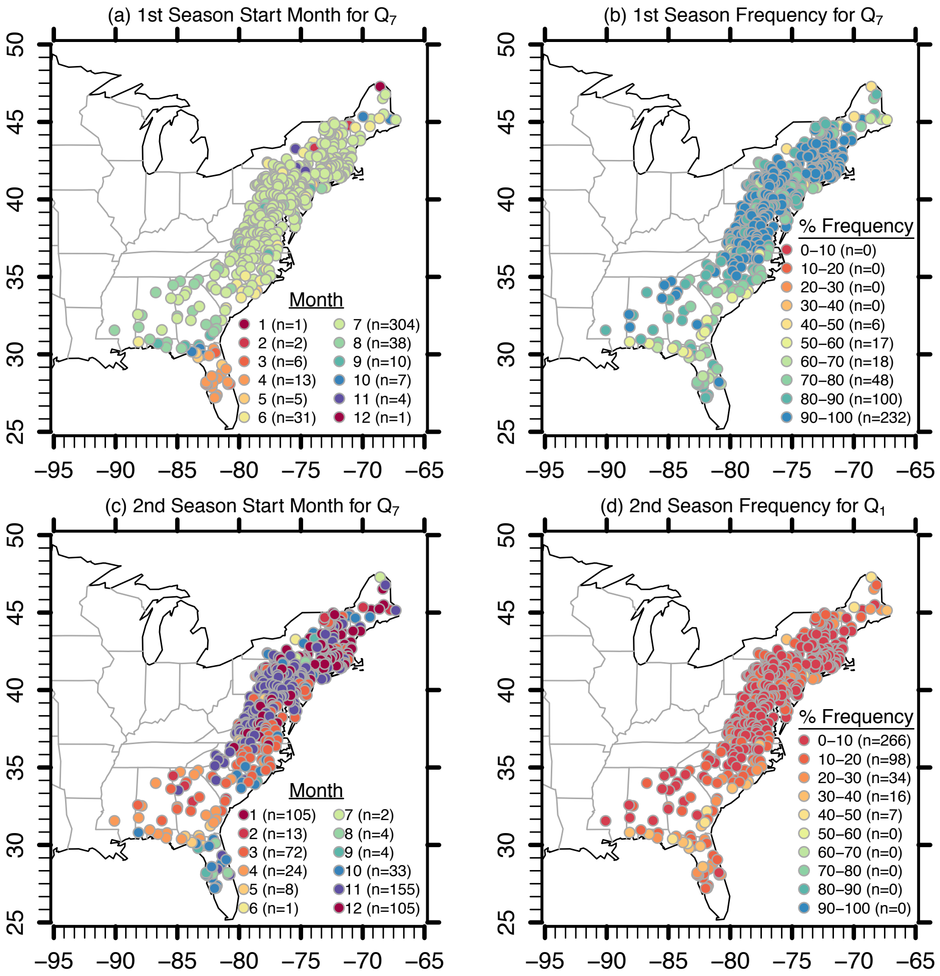

Figure 8. Primary and secondary seasons of occurrence of Q_7 low flows and their frequencies.

Justin Sheffield 9/12/15 12:17 PM
Deleted: $_1$

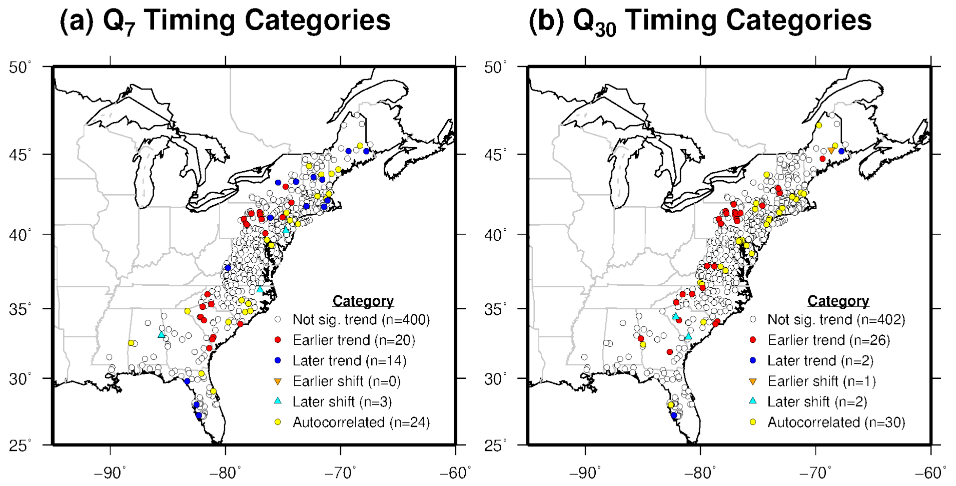

Figure 9. Categorization of non-stationarity of sites for timing of (a) Q_7 and (b) Q_{30}

Justin Sheffield 7/20/15 4:59 PM

Deleted: a

Justin Sheffield 7/20/15 4:59 PM

Deleted: Trend in the onset time of the low flow season for Q_1 and Q_7. Sites with statistically significant autocorrelation are not shown. (b) The year of step change in timing for Q_1 and Q_7.

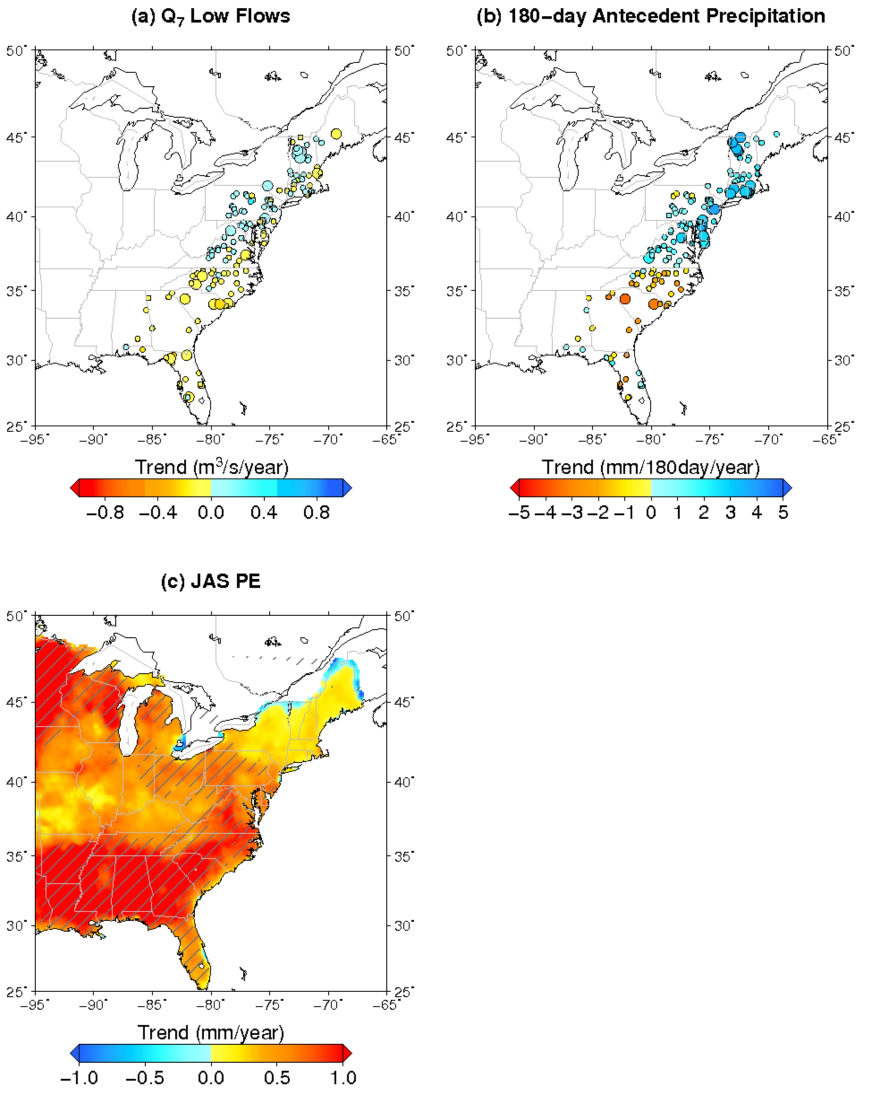

Figure 10. (a) Trend in Q_7 low flows for 1951-2005 for the warm season. (b) Corresponding trend in 180-day antecedent precipitation. For (a) and (b), trends that are statistically significant at the 0.05 level are shown in large symbols. (c) Trend in July-August-September (JAS) potential evaporation for 1979-2012. Statistically significant trends are shown by hatching.

Justin Sheffield 7/18/15 11:05 AM
Deleted: 9

Justin Sheffield 9/12/15 12:42 PM
Deleted: