INTERNATIONAL JOURNAL OF CLIMATOLOGY

RMetS Royal Meteorological Society

On the impact of gaps on trend detection in extreme streamflow time series

Journal:	*International Journal of Climatology*
Manuscript ID	JOC-16-0338.R1
Wiley - Manuscript type:	Short Communication
Date Submitted by the Author:	n/a
Complete List of Authors:	Slater, Louise; University of Iowa, IIHR - Hydroscience and Engineering Villarini, Gabriele; University of Iowa, Civil & Environmental Engineering
Keywords:	Gaps, Trend detection, Annual maximum, Peak-over-threshold, Times series analysis, Streamflow, Hydrology

Submitted to *International Journal of Climatology*

SHORT COMMUNICATION

On the impact of gaps on trend detection in extreme streamflow time series

LOUISE SLATER[1], GABRIELE VILLARINI[1]

[1] IIHR-Hydroscience & Engineering, The University of Iowa, Iowa City, Iowa, USA

Corresponding author: louise-slater@uiowa.edu, +1 319 383 5932

Summary

Streamflow time series often contain gaps of varying length and location. However, the influence of these gaps on trend detection is poorly understood and cannot be estimated a priori in trend-detection studies. We simulated the effects of varying gap size (1, 2, 5, and 10 years) and location (one quarter, one third, and half of the way) on the detection rate of significant monotonic trends in annual maxima and peaks-over-threshold, based on the most commonly-used trend tests in time series of varying length (from 15 to 150 years) and trend magnitude (β_1). Results show that, in comparison with the complete time series, the loss in trend detection rate tends to grow with (i) increasing gap size, (ii) increasing gap distance from the middle of the time series, (iii) decreasing β_1 slope, and (iv) decreasing time series length. Based on these findings, we provide objective recommendations and cautionary remarks for maximal gap allowance in trend detection in extreme streamflow time series.

Key Words

Gaps; Trend detection; Annual maximum; Peak-over-threshold; Times series analysis

1. Introduction

Data gaps are pervasive in hydrological networks globally. In the continental USA, records of mean daily and peak annual streamflow are remarkably long (those exceeding ten complete years are 45 and 49 years long on average, respectively) but data gaps are frequent and sometimes lengthy (36%/47% of these records have at least one gap, lasting 10.6/10.0 years on average, respectively; **Figure 1**). In the UK, similarly, 22% of flow records were less than 95% complete in 2008 (Marsh and Hannaford 2008). Missing values are a common feature of river flow archives (e.g., Stahl et al. 2010, Hannah et al 2011, Whitfield et al. 2012) and are of notable concern, as they may hinder the calculation of summary statistics (Hannaford 2004) and affect statistical trend detection (Helsel and Hirsch 1992).

Gaps in streamflow records arise from a variety of causes. Short-term gaps (single to multiple measurements) may occur as a result of ice effects, backwater, equipment damage during large flood events (Kundzewicz and Robson 2004), or malfunctioning of the recording system (e.g., Carter and Davidian 1968, Rantz 1982). Long-term discontinuities in data collection, in contrast, tend to arise from changes in gauging site establishment (e.g., Juracek and Fitzpatrick 2009), under-funding (e.g., Lins et al. 2010), or station destruction (resulting from major floods or vandalism). The location of the resulting gaps may vary within the streamflow distribution, often clustering at the extremes (i.e. minima and maxima) of the flow range (e.g., Hannaford 2004, Marsh 2002, Gustard and Demuth 2009) when flows are too low or too high to be accurately recorded with the existing equipment.

Various methods have been suggested for dealing with missing streamflow data, depending on the size of the gap. At the sub-annual scale, gaps are typically infilled when it is preferable to add synthetic data than leave a gap in the record (e.g., Lamb et al. 2003). The choice of the most appropriate infilling method depends on the type of site, streamflow variability, gap size, record length, conditions when the gap occurred (rising, falling, or peaking flow), metadata, available tools, and knowledge of the person correcting the data (Gustard and Demuth 2009). For single-value gaps, the local average is preferable to the sample mean (Pappas et al. 2014). For gaps of less than one day, interpolation is preferred (Archer 2007), and for gaps of less than one month, records are often compared with those of neighbor 'donor' gauging stations with similar discharge (Hannaford and Buys 2012), although differences in mean and variance may occur unless they are corrected for (e.g. Grygier et al. 1989). The equi-percentile method is also used for gaps exceeding seven days (Hannaford 2004, Lavers et al. 2010), and has been shown to perform well in comparison with 14 other different infilling methods (Harvey et al. 2012). Sometimes mixed methods are used, e.g., combining linear regression with a streamflow model (Sanderson et al. 2012), or more complex statistical approaches in the case of very large gaps (Gyau-Baokye and Schultz 1994). Overall, it is recognized that these infilling approaches are more reliable for short than for long gaps (e.g., Hirsch and Fischer 2014).

At the annual to decadal scale there are fewer recommendations in the existing literature. One technique is to divide the time series into three sections of equal length, and discard any section with less than 20% of the total coverage (Helsel and Hirsch 1992); however this method is infrequently used (e.g., Asarian and Walker 2016) and there is not a clear rationale for choosing those thresholds over others. Other approaches range from lax to strict. Some ignore the presence of gaps (e.g., Archer 2007, Tomkins 2014) when data availability is limited and a stringent filtering would do more harm than good (Slater 2016). Typically, it is considered that a one- or two-year gap in the middle of a flow record should not disqualify a station from an analysis (Helsel and Hirsch 1992). Others select a maximal gap threshold below which gaps are deemed acceptable, e.g., *x* consecutive days (Zaidman et al 2002, Hannaford and Buys 2012), or a minimal coverage, e.g., 330 days (~90% completeness) per year (Mallakpour and Villarini 2015, Slater et al. 2015). It may also make sense to remove any sites where gaps exceed one (Petrow and Merwade 2009) to five years (Guo et al. 2014), or to require a percentage of completeness over a period of several decades (Adam and Lettenmaier 2008). Last, the most strict approach consists in prohibiting gaps entirely (e.g., Baker et al. 2004, Pinter et al. 2008, Villarini et al. 2009a), but this approach typically eliminates many sites, especially in regions such as Africa, Asia, and South America (Kundzewicz et al. 2005) where there are data continuity issues arising from hydrological reasons, funding, and institutional capacity.

Little is known about the effects of annual gaps on trend detection, and so the choice of maximal gap allowance in trend analyses is seldom fully justified (e.g., Mallakpour and Villarini 2015, Slater et al. 2015). The aim of this work, therefore, is to provide a preliminary framework to better understand the effects of gaps on the detection of monotonic trends in time series of streamflow maxima. We investigate the influence of gap location on trend detection, for gaps of varying length and location in the record, varying trend magnitude, and record length. We do not investigate the influence of autocorrelation and/or short and long term persistence, although these would be valuable questions for further analyses. Based on our results, we provide objective recommendations for maximal gap allowance in trend detection, with a number of caveats.

2. Methods

2.1. Data

For all stream gauges in the continental USA, we downloaded peak annual streamflow data from the U.S. Geological Survey (USGS) National Water Information System (NWIS) at http://nwis.waterdata.usgs.gov/nwis/peak, and mean daily streamflow data using the package ‘dataRetrieval’ (Hirsch and De Cicco 2015) in the open-source software *R*. We computed the number and average length of gaps (i.e., years with less than 365 daily values, or no annual peak value) at each site that had at least ten complete years of data (**Figure 1**).

Across the continental USA, we find that no regions are entirely gap-free, and that 13% (daily) and 17% (annual peak) of sites have at least one gap that equals or exceeds ten years. For the simulation we keep only the historical flow records with at least 30 complete years of daily data (4,525 sites with daily streamflow, and 7,575 sites with annual peak streamflow) to obtain a range of plausible trend magnitudes.

2.2. Simulation Setup

For both the peak-over threshold (POT) and Generalized Extreme Value (GEV) simulations described below we follow a similar procedure, generating synthetic streamflow time series, and conducting sensitivity analyses with these data (see **Figure S1** for a flow chart).

POT

At the 4,525 retained sites with daily data we use a POT approach to compute the number of separate flood events in each year above a threshold. We set this threshold as the streamflow that is exceeded twice per year on average over the entire record. Separate events are selected using an inhibition window of 5 days + log(A) (e.g., Lang et al. 1999), where A is the contributing drainage area in logarithmically-transformed mi^2. Poisson regression is used (because the data are discrete and bounded at zero, see e.g., Dobson 2008) to estimate the temporal trends in annual counts, as:

$$\lambda = \exp(\beta_0 + \beta_1 \times year) \quad (1)$$

The β_0 and β_1 parameters are retained for all 4,525 sites. For the simulation, we choose $\beta_{1,sim}$ parameters based on the observed distribution of β_1 (i.e., measured from the 4,525 sites; 0 and ± 0.001, 0.005, 0.01, 0.02, 0.03, 0.04, 0.05, 0.1) to encompass both small and large observed slopes. The idea is to investigate how the accuracy of trend detection would vary in the case of small versus large underlying trends (unknown β_1 slopes) in the data. The $\beta_{0,sim}$ parameter is set (fixed) as the median of the observed distribution at the 4,525 sites for consistency throughout the simulation ($\beta_{0,sim}$=0.687). To avoid any dependence of the parameters on a specific time-period, we set a common start year equal to 1. Gaps are located one quarter, one third, and half of the way into the time series.

To generate a synthetic Poisson distribution for a sequence of 150 continuous years, we compute the rate parameter λ for every year (λ_i) ranging from 1 to 150, following equation 1. We generate 50,000 streamflow values for every year, using λ_i inside the function rpois(n=50000, λ_i) in *R* (R Core Team, 2015). From the 150-year time series, we subset streamflow time series of varying length, ranging from 15 to 150 years, with a time step of 5 years. In other words, for every time series of length *n*, we subsample *n* streamflow values (one from the 50,000 for every year).

We use Poisson regression to estimate $\hat{\beta}_0$ and $\hat{\beta}_1$ in the simulated series. If $\hat{\beta}_1$ has the same sign as $\beta_{1,sim}$ and is significant at the 5% level, then we consider the detection as a "hit." If $\hat{\beta}_1$ does not have the same sign as $\beta_{1,sim}$ or is not significant, then it is a "miss." We then remove *x* (e.g., 5) years of data from the same time series, at location *m* (e.g., one-third of the way in), and re-compute $\hat{\beta}_1$ and its statistical significance. As before, the trend detection can be either a *hit* or a *miss*. This procedure is iterated 50,000 times, storing the result (*hit* or a *miss*) for both the complete time series and the 'gapped' time series. We then compute the detection rate for both the complete and 'gapped' time series as the fraction of *hits* out of 50,000 iterations (**Figure S2** indicates the trend detection rate for the complete POT time series). The effect of the gap on trend detection is then measured as the percentage difference between detection rates for the incomplete time series with respect to the complete series.

GEV

To examine the impact of gaps on trend detection in annual maxima, we use the GEV distribution (e.g., Coles et al. 2001). The GEV distribution is dependent upon the location parameter μ (which governs the magnitude), the scale parameter σ (the variability), and the shape parameter ξ (the heaviness of the tail). To obtain realistic parameters for the simulation, we fit a GEV distribution with a linear trend in the location parameter μ (with parameters β_0, β_1) and constant σ and ξ using the R package 'ismev' (Heffernan and Stephenson 2014) for every site with at least 30 complete years of peak data. We estimate and retain the β_0, β_1, σ, and ξ parameters for all 7,575 of those sites.

For the simulation, our approach is the same as with the daily data, but adapted to the GEV distribution. The $\beta_{0,sim}$ and $\sigma_{,sim}$ parameters for the simulation are kept constant, using the median of the observed β_0 (equal to 2527.35) and σ (equal to 1467.61) distributions at the 7,575 sites. The $\beta_{1,sim}$ parameters are based on the observed distribution of β_1s (0 and ± 1, 2.5, 5, 10, 25, 50, 100, 150, 200) to encompass both small and large observed slopes. We also explore the sensitivity of the results to different values of ξ_{sim} falling in the domain of attraction of the Weibull, Gumbel or Fréchet distribution (ξ_{sim} equal to -1, 0, or +1).

To produce a synthetic GEV streamflow distribution for a sequence of 150 continuous years, we begin by computing a μ rate parameter for every year (μ_i, from 1 to 150), as:

$$\mu = \beta_0 + \beta_1 \times year \qquad (2)$$

We then generate 50,000 synthetic streamflow values for every year using the GEV distribution in the software *R*, as rgev(n=50000, ξ_{sim}, μ_i, σ_{sim}), with the fExtremes package (Wuertz et al. 2013). From these computed annual streamflow values, we subset streamflow time series of varying length, ranging from 15 years to 150 years, with a time step of 5 years, as before. Any negative values are replaced by zero, as strictly positive streamflow values are desired for the simulation.

For every time series, we compute the value of the Mann Kendall tau and associated p-value (Mann 1945, Kendall 1975), using the *R* package Kendall (McLeod 2011). If the sign of the tau is the same that of $\beta_{1,sim}$ and is significant at the 5% level, then we consider the detection as a "hit." If the tau does not have the same sign as the $\beta_{1,sim}$ or is not significant, then it is a "miss." As before, we then remove *x* (e.g., 5) years of data from the time series, at location *m* (e.g. one-third of the way in), and re-compute tau and the associated p-value. The trend detection can be either a *hit* or a *miss*. This procedure is iterated 50,000 times, storing the result (*hit* or a *miss*) for the complete and the truncated time series. At the end of the iterations, we compute the detection rate for both the complete and the truncated time series as the fraction of *hits* out of 50,000 iterations (**Figures S3-S5** indicate the trend detection rate for the complete GEV time series, with varying ξ_{sim}). The effect of the gap on trend detection is then measured as the percentage difference between detection rates for the incomplete time series with respect to the complete series.

In performing these simulations, we do not consider the effects of serial correlation (e.g., Yue et al. 2003, Pappas et al. 2014) or long-term persistence (e.g., Cohn and Lins 2005), and we assume that the nature of the trends is monotonic (as is assumed in all of the studies that use the Mann-Kendall test). However, we acknowledge that streamflow records may exhibit more complex patterns (e.g., Hall and Tajvidi 2000; Ramesh and Davison 2002, Villarini et al. 2009b) and that there is ongoing discussion about the limitations of a nonstationary description of hydrological processes (Montanari and Koutsoyiannis 2014; Koutsoyiannis and Montanari 2015; Read and Vogel 2015; Serinaldi and Kilsby 2015; Serinaldi 2015; Dimitriadis et al. 2016). One last issue that should also be taken into account is the period of record. In locations where streamflow time series are strongly influenced by interannual and interdecadal climatic variability, the absolute length of the time series or of the gaps may be less important than the start/end dates of the records (and by extension of any gaps) in relation to the behavior of oceanic and atmospheric drivers.

3. Results

The effect of gaps on trend detection rate varies considerably depending on the size of the gap (one, two, five or ten years), the location of the gap within the time series (one quarter, one third, or half of the way in), the length of the time series (from 15 to 150 years), and the magnitude of the β_1 slope (which differs between daily POT time series and annual GEV time series), for both the daily POT and the annual peak GEV time series (**Figures 2 and 3**). For both types of time series, we find that the larger the gap, the lower the detection rate of significant (i.e., non-zero) β_1 slopes in comparison with the detection rate for the complete time series. For gaps of just one or two years, the percentage difference in detection rate between incomplete and complete time series is generally less than 10%. However, a ten-year gap, in comparison with a one-year gap, will decrease the detection rate by roughly one order of magnitude or more (**Figures 2 and 3**).

Gaps that are located centrally within a time series (half of the way in) have a smaller impact on trend detection than gaps that are located towards the extremes (e.g., one quarter of the way in). The color maps suggest that gaps located at the beginning/end of a time series would have an even greater effect on trend detection (**Figures 2 and 3**). Similarly, the further a change in mean or variance is located from the middle of a streamflow time series, the lower the power of statistical tests in detecting these changes (Mallakpour and Villarini 2016, Nayak and Villarini 2016).

Equally important is the length of the time series: long time series are less affected by gaps than short time series, i.e., gap size only matters proportionally to the length of the entire time series. Last, the magnitude of the β_1 slope is important, so gaps have a lesser influence on time series characterized by strong trends (i.e., trends are better detected for larger β_1 values).

The daily POT data exhibit an asymmetric pattern with lower trend detection rates for positive values of β_1 (**Figure 2**) due to the nature of the Poisson regression model, which is bounded at zero to prevent negative values. The asymmetry suggests that Poisson regression is less sensitive to gaps in the presence of negative (vs. positive) trends in the extreme streamflow time series, and thus more successful in detecting decreases than increases in flood frequency.

The annual peak GEV time series exhibit much greater variability than POT (as is to be expected due to greater variability in the maxima), and no asymmetry between positive and negative values of β_1 (**Figure 3**). Additionally, because of the variability in annual maxima, a stronger β_1 slope is needed for significant trends to become detectable in comparison with the daily POT, and there is a much larger band of 'indeterminate' trends (mixed positive and negative values in **Figure 3**) for low values of β_1 as one approaches zero.

The ξ parameter, which determines the heaviness of the tail of the GEV distribution, also affects the rate of trend detection, with lower detection rates for higher values of ξ. Results indicate that for ξ_{sim}=1 (i.e., a streamflow record following a Fréchet distribution) a slightly longer time series (e.g., by 10 years) is required to obtain a similar trend detection rate to ξ_{sim}=–1 (Weibull distribution), regardless of gap location or length (**Figure 3**). Thus, to avoid biases relating to the shape of the GEV distribution, longer time series may be preferable in analyzing time series of peak streamflow data, in contrast with the POTs.

The results of the simulation tests provide us with objective guidelines on the maximal gap length that can be selected to obtain a given level of confidence (80% or 90%) in trend detection rates (**Figure 4**). For example, to obtain an 80%-accurate detection rate in a 30-year daily POT time series at least 80% of the time, gaps must not exceed ten years if they are located one third or half of the way into the time series (**Figure 4**, top left panel), but five years if they are located one quarter of the way. If a 90%-accurate detection rate is preferred for the same 30-year time series, a two to ten year gap would be deemed acceptable depending on gap location (**Figure 4**, bottom left panel). The guidelines for the GEV time series are slightly more conservative, particularly for higher values of ξ, because of the greater variability in annual maxima in comparison with daily POT. For a 90% accurate detection rate in the same 30-year time series and ξ_{sim}=1, a two to five year gap only would be deemed acceptable.

We provide these results with two notable caveats. (1) Here we cannot provide detection rates greater than 90% because the variability in the simulation leaves some gaps in the data (e.g., see GEV ξ_{sim}=0, 90% detection, **Figure 4**). Generally, if a trend test is applied to a time series with substantial gaps, the likelihood of obtaining an accurate trend test result is significantly reduced and performing the trend test may not be recommended. (2) Our results provide guidelines for characterizing trends over broad spatial regions where only limited data are available. At any individual site, however, gaps may have a major impact, particularly in regions with strong interdecadal climatic variability, or when gaps are located at the beginning/end of a time series. It is thus generally advisable to use complete time series if one wishes to fully understand streamflow variability and potential non-stationarity. Therefore, we recommend great caution in using these guidelines.

4. Conclusions

Here, we address the lack of rigorous procedures for estimating the influence of annual to decadal gaps on trend detection by providing objective estimates of the maximum gap size that should be allowed in time series of varying length (15-150 years) to obtain an 80% or 90% accurate detection rate of significant trends. However, this work remains a preliminary investigation, and any further simulations using the same framework could take into account the effects of non-linearities and autocorrelation in the data (because short- or long-term persistence in the record can affect trend detection due to the clustering of extremes) or the effects of non-independent residuals and/or changing variance (i.e. where the residuals are non-identically distributed around the mean). Additionally, the effects of gaps on trend detection could be investigated in rivers of varying catchment size and flow variability, or in the lower tail of the streamflow distribution, where gaps also tend to cluster. Here, because we focus on the upper and most variable part of the flow distribution, our recommendations may serve as a conservative guideline for trend detection in most types of streamflow time series.

5. Acknowledgments

This study was supported by the Broad Agency Announcement (BAA) Program and the Engineer Research and Development Center (ERDC)–Cold Regions Research and Engineering Laboratory (CRREL) under Contract No. W913E5-16-C-0002. Gabriele Villarini also acknowledges financial support from the USACE Institute for Water Resources.

6. Bibliography

Adam, J.C., Lettenmaier, D.P. 2008. Application of New Precipitation and Reconstructed Streamflow Products to Streamflow Trend Attribution in Northern Eurasia. *Journal of Climate*, *21*(8), 1807–1828. http://doi.org/10.1175/2007JCLI1535.1

Archer, D.R. 2007. The use of flow variability analysis to assess the impact of land use change on the paired Plynlimon catchments, mid-Wales. *Journal of Hydrology*, *347*(3-4), 487–496. http://doi.org/10.1016/j.jhydrol.2007.09.036

Asarian, J. E., Walker, J.D. 2016. Long-term trends in streamflow and precipitation in Northwest California and Southwest Oregon, 1953-2012 1. *Journal of the American Water Resources Association*, *04011*. http://doi.org/10.1111/1752-1688.12381

Baker, D., Richards, R., Loftus, T., Kramer, J. 2004. A new flashiness index: characteristics and applications to midwestern rivers and streams. *Journal of the American Water Resources Association*, 503–522. http://doi.org/10.1111/j.1752-1688.2004.tb01046.x

Carter, R. W., and Davidian, J. 1968. Chapter A6 - General procedure for gaging streams, Book 3 - Applications of hydraulics. In *Techniques of Water Resources Investigations of the USGS* (pp. 1–13).

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

Coles, S., Bawa, J., Trenner, L., and Dorazio, P. 2001. *An Introduction to Statistical Modelling of Extreme Values*. Vol. 208. London. Springer.

Dimitriadis P, Koutsoyiannis D, Tzouka K. 2016. Predictability in dice motion: how does it differ from hydrometeorological processes? *Hydrological Sciences Journal*. 61 (9), 1611-1622. http://doi.org/10.1080/02626667.2015.1034128

Dobson, A.J., and Barnett, A. 2008. *An introduction to generalized linear models*. CRC press.

Grygier, J.C., Stedinger, J.R. and Yin, H-B. 1989. A generalized maintenance of variance extension procedure for extending correlated series. *Water Resources Research* 25.3: 345-349.

Guo, J., Li, H., Ruby Leung, L., Guo, S., Liu, P., Sivapalan, M. 2014. Links between flood frequency and annual water balance behaviors. *Water Resources Research*, 937–953. http://doi.org/10.1002/2013WR014374

Gustard, A., Demuth, S. 2009. *Manual on low-flow estimation and prediction*. Operational Hydrology Report No. 50, World Meteorological Organization (WMO)-No. 1029

Gyau-Baokye, P. and Schultz, G.A. 1994: Filling gaps in runoff time series in West Africa. *Hydrological Sciences Journal*, 39(6): 621–636.

Hall, P., Tajvidi, N., 2000. Nonparametric analysis of temporal trend when fitting parametric models to extreme-value data. *Statistical Science* 15 (2), 153–167.

Hannaford, J., 2004. Development of a strategic data management system for a national hydrological database, the UK national river flow archive. In: Liong, S.Y., Phoon, K., Babovic, V. (Eds.), *Proceedings of 6th International Conference on Hydroinformatics. World Scientific Publishing*, pp. 637–644.

Hannaford, J., Buys, G. 2012. Trends in seasonal river flow regimes in the UK. *Journal of Hydrology*, *475*, 158–174. http://doi.org/10.1016/j.jhydrol.2012.09.044

Hannah, D.M., Demuth, S., van Lanen, H. Looser, U., Prudhomme, C., Rees, G., Stahl K, Tallaksen, L.M. 2011. Large-scale river flow archives: importance, current status and future needs. *Hydrological Processes*, *25*(7), 1191–1200. http://doi.org/10.1002/hyp.7794

Harvey, C., Dixon, H., Hannaford, J. 2012. An appraisal of the performance of data infilling methods for application to daily mean river flow records in the UK. *Hydrol. Res.* 43, 618–636 http://doi.org/10.2166/nh.2012.110

Heffernan, J.E. Stephenson, A.G. 2014. ismev: An Introduction to Statistical Modeling of Extreme Values. R package version 1.40. http://CRAN.R-project.org/package=ismev

Helsel, D.R. Hirsch, R.M. 1992. *Statistical Methods in Water Resources*. Studies in Environmental Science 49. Elsevier, Amsterdam, The Netherlands.

Hirsch, R.M., Fisher, G.T. 2014. Past, Present, and Future of Water Data Delivery from the U.S. Geological Survey. *Journal of Contemporary Water Research & Education*, *153*, 4–15. http://doi.org/10.1111/j.1936-704X.2014.03175.x

Hirsch, R.M., De Cicco, L.A. 2015. User guide to Exploration and Graphics for RivEr Trends (EGRET) and dataRetrieval: R packages for hydrologic data (version 2.0, February 2015): U.S. Geological Survey Techniques and Methods book 4, chap. A10, 93 p., http://dx.doi.org/10.3133/tm4A10

Juracek, K. E., Fitzpatrick, F.A. 2009. Geomorphic applications of stream-gage information. *River Research and Applications*, *25*(3), 329–347. http://doi.org/10.1002/rra

Kendall, M. G. 1975. Rank Correlation Methods, Charles Griffin, London

Koutsoyiannis D, Montanari A. 2015. Negligent killing of scientific concepts: the stationarity case. *Hydrological Sciences Journal*. **60**(7-8): 1174–1183. http://doi.org/10.1080/02626667.2014.959959

Kundzewicz, Z.W., Robson, A.J. 2004. Change detection in hydrological records—a review of the methodology / Revue méthodologique de la détection de changements dans les chroniques hydrologiques. *Hydrological Sciences Journal*, *49*(1), 7–19. http://doi.org/10.1623/hysj.49.1.7.53993

Kundzewicz, Z.W., Graczyk, D., Maurer, T., Pińskwar, I., Svensson, C., Szwed, M. 2005. Trend detection in river flow series: 1. Annual maximum flow. *Hydrological Sciences Journal*, *50*(5).

Lamb, R., Zaidman, M.D., Archer, D.R., Marsh, T.J., Lees, M.L. 2003. River Gauging Station Data Quality Classification (GSDQ). *R&D Technical Report W6-058/TR, for the Environment Agency*.

Lang, M., Ouarda T.B.M.J., and Bobée, B. 1999. Towards operational guidelines for over-threshold modeling, *Journal of Hydrology*, 225, 103-117. http://doi.org/10.1016/S0022-1694(99)00167-5

Lavers, D., Prudhomme, C., Hannah, D. M. 2010. Large-scale climate, precipitation and British river flows: Identifying hydroclimatological connections and dynamics. *Journal of Hydrology*, *395*(3-4), 242–255. http://doi.org/10.1016/j.jhydrol.2010.10.036

Lins, H.F., Hirsch, R.M., and Kiang, J. 2010. Water—the Nation's Fundamental Climate Issue: A White Paper on the U.S. Geological Survey Role and Capabilities: U.S. Geological Survey Circular 1347, 9 p., http://pubs.usgs.gov/circ/1347.

Mallakpour, I., and Villarini, G. 2015. The changing nature of flooding across the central United States. *Nature Climate Change*, (February), 1–5. http://doi.org/10.1038/NCLIMATE2516

Mallakpour, I., and Villarini, G., 2016. A simulation study to examine the sensitivity of the Pettitt test to detect abrupt changes in mean, *Hydrological Sciences Journal*, 61(2), 245-254, http://doi.org/10.1080/02626667.2015.1008482

Mann, H. B. 1945. Non-parametric tests against trend, Econometrica, 13, 245–259.

Marsh, T.J. 2002. Capitalising on river flow data to meet changing national needs: a UK perspective. Flow Measurement and Instrumentation, 13: 291–298.

Marsh, T.J. and Hannaford, J. (Eds). 2008 UK Hydrometric Register. Hydrological data UK 689 series. Centre for Ecology & Hydrology.

McLeod, A.I. 2011. Kendall: Kendall rank correlation and Mann-Kendall trend test. R package version 2.2. http://CRAN.R-project.org/package=Kendall

Montanari, A., and Koutsoyiannis D. 2014. Modeling and mitigating natural hazards: Stationarity is immortal!, *Water Resour. Res.*, 50(12), 9748–9756, http://doi.org/10.1002/2014WR016092

Nayak, M., and Villarini, G. 2016. Evaluation of the capability of the Lombard test in detecting abrupt changes in variance, *Journal of Hydrology*, 534, 451-465. http://doi.org/10.1016/j.jhydrol.2016.01.016

Pappas C, Papalexiou SM, Koutsoyiannis D. 2014. A quick gap filling of missing hydrometeorological data. *Journal of Geophysical Research: Atmospheres* **119**(15): 9290–9300. http://doi.org/10.1002/2014JD021633

Pinter, N., Jemberie, A., Remo, J.W.F., Heine, R, & Ickes, B.S. 2008. Flood trends and river engineering on the Mississippi River system. *Geophysical Research Letters*, *35*(23), 1–5. http://doi.org/10.1029/2008GL035987

Petrow, T., & Merwade, V. 2009. Trends in flood magnitude, frequency and seasonality in Germany in the period 1951–2002. *Journal of Hydrology*, *371*(1-4), 129–141. http://doi.org/10.1016/j.jhydrol.2009.03.024

R Development Core Team. 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, http://www.R-project.org

Ramesh, N.I., Davison, A.C., 2002. Local models for exploratory analysis of hydrological extremes. *Journal of Hydrology* 256, 106–119. http://dx.doi.org/10.1016/S0022-1694(01)00522-4

Rantz, S. E. 1982. Measurement and Computation of Streamflow. Volume 1. Measurement of Stage and Discharge. In *U.S. Geological Survey, Water-Supply Paper 2175* (pp. 1–284). Reston, Virginia.

Read LK, Vogel RM. 2015. Reliability, return periods, and risk under nonstationarity. *Water Resources Research* **51**(8): 6381–6398. http://doi.org/10.1002/2015WR017089

Sanderson, M.G., Wiltshire, A.J., & Betts, R. 2012. Projected changes in water availability in the United Kingdom. *Water Resources Research*, *48*(8), 1–11. http://doi.org/10.1029/2012WR011881

Serinaldi F. 2015. Dismissing return periods! *Stochastic Environmental Research and Risk Assessment*. Springer Berlin Heidelberg **29**(4): 1179–1189. http://doi.org/10.1007/s00477-014-0916-1

Serinaldi F, Kilsby CG. 2015. Stationarity is undead: Uncertainty dominates the distribution of extremes. *Advances in Water Resources*. Elsevier Ltd **77**: 17–36. http://doi.org/10.1016/j.advwatres.2014.12.013

Stahl K, Hisdal H, Hannaford J, Tallaksen LM, van Lanen HAJ, Sauquet E, Demuth S, Fendekova M, Jódar J. 2010. Streamflow trends in Europe: evidence from a dataset of near-natural catchments. *Hydrology and Earth System Sciences* **14**(12): 2367–2382. http://doi.org/10.5194/hess-14-2367-2010

Slater, L.J., Singer, M.B., and Kirchner, J.W. 2015. Hydrologic versus geomorphic drivers of trends in flood hazard. *Geophysical Research Letters*, *42*, 1–7. http://doi.org/10.1002/2014GL062482

Slater, L.J. 2016. To what extent have changes in channel capacity contributed to flood hazard trends in England and Wales? *Earth Surface Processes and Landforms*, http://doi.org/10.1002/esp.3927

Tomkins, K.M. 2014. Uncertainty in streamflow rating curves: methods, controls and consequences. *Hydrological Processes*, *28*(3), 464–481. http://doi.org/10.1002/hyp.9567

Villarini, G., Serinaldi, F., Smith, J. a., and Krajewski, W. F. (2009a). On the stationarity of annual flood peaks in the continental United States during the 20th century. *Water Resources Research*, 45(8), 1–17. http://doi.org/10.1029/2008WR007645

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

Whitfield PH, Burn DH, Hannaford J, Higgins H, Hodgkins GA, Marsh T, Looser U. 2012. Reference hydrologic networks I. The status and potential future directions of national reference hydrologic networks for detecting trends. *Hydrological Sciences Journal* **57**(8): 1562–1579. http://doi.org/10.1080/02626667.2012.728706

Wuertz, D. and many others. 2013. fExtremes: Rmetrics - Extreme Financial Market Data. R package version 3010.81. http://CRAN.R-project.org/package=fExtremes

Yue, S., P. Pilon and B. Phinney 2003. Canadian streamflow trend detection: impacts of serial and cross-correlation, *Hydrological Sciences Journal*, 48:1, 51-63

Zaidman, M.D., Keller, V., Young, A.R. 2002. Probability distribution for “x-day” daily mean flow events (from gauged records); Methodology for use with Long Records & Prototype Methodology for use with Short Records R&D Technical Report W6-064/TR2. Centre for Ecology and Hydrology, Wallingford, UK. 247pp.

Figure legends

Figure 1. Gap diagnostic. Number of gaps (top row) and average duration of gaps (bottom row) in U.S. Geological Survey complete (i.e., 365 values per year for the daily streamflow, or one value per year for the peak data) continental streamflow time series with a minimum record of at least ten years.

Figure 2. Difference in trend detection rate for mean daily streamflow POT trends between incomplete and complete time series, of varying length (x-axis), $\beta_{1,sim}$ magnitude (y-axis), gap size (rows), and gap location (columns). Colors indicate the percentage difference in trend detection between the incomplete and the complete time series (after versus before gap removal), ranging from dark blue (difference greater than -10%) to red (difference greater than +1%), through zero (no difference).

Figure 3. Difference in trend detection rate for annual peak GEV trends between incomplete and complete time series, of varying length (x-axis), $\beta_{1,sim}$ magnitude (y-axis), gap size (rows), gap location (columns), and ξ_{sim} (panels). Colors indicate the percentage difference in trend detection between the incomplete and the complete time series (after versus before gap removal), ranging from dark blue (difference greater than -10%) to red (difference greater than +1%), through zero (no difference).

Figure 4. Recommendations for acceptable gap length in mean daily streamflow POT and annual peak GEV time series depending on time series length, gap location, and for different values of ξ_{sim}**.** Color indicates the location of the gap in the time series, ranging from middle (dark blue) to quarter of the way in (light blue). The represented gap location shows the most restrictive requirement; gaps located in the middle of the time series are the least influential for trend detection.

Supporting Information contains Figures S1-S5

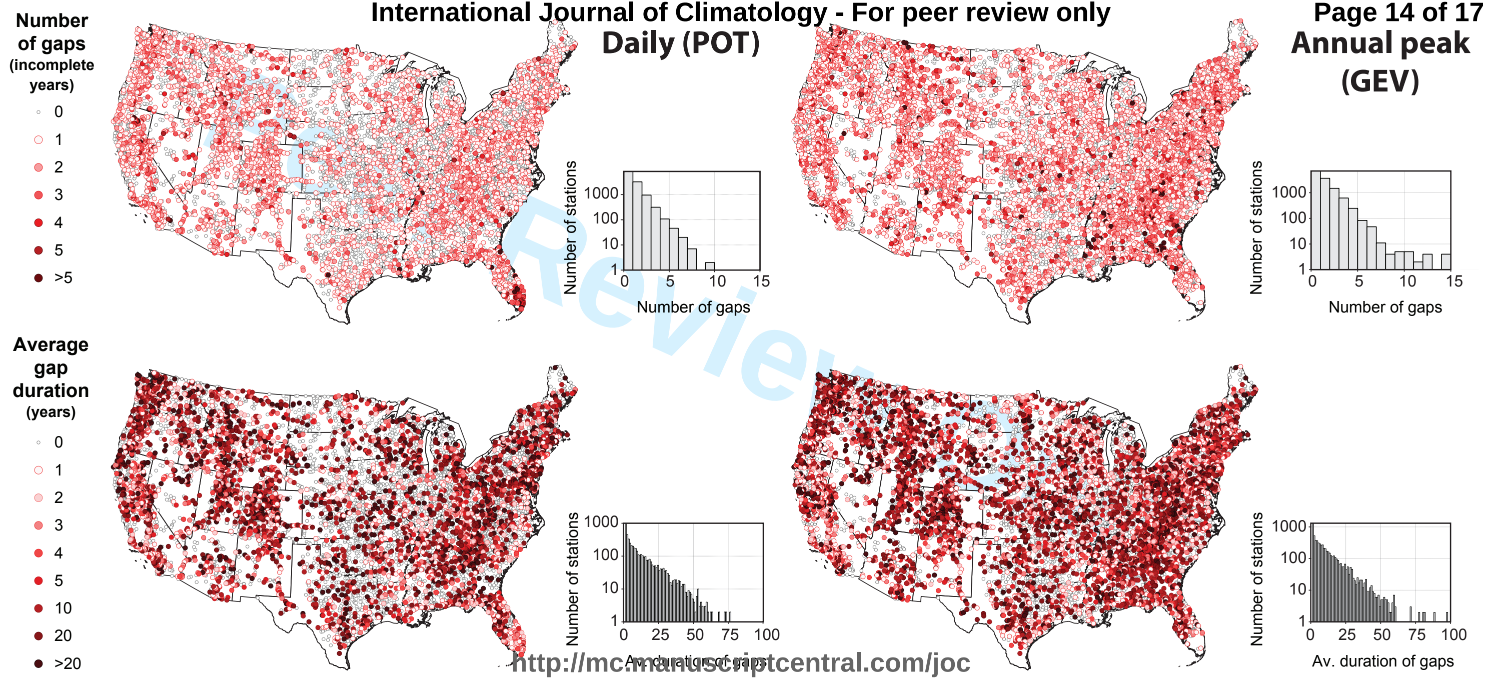
Daily (POT)
Annual peak (GEV)
Number of gaps (incomplete years)
0
1
2
3
4
5
>5
Number of stations
1000
100
10
1
0
5
10
15
Number of gaps
Average gap duration (years)
0
1
2
3
4
5
10
20
>20
Number of stations
1000
100
10
1
0
25
50
75
100
Av. duration of gaps

Figure 1

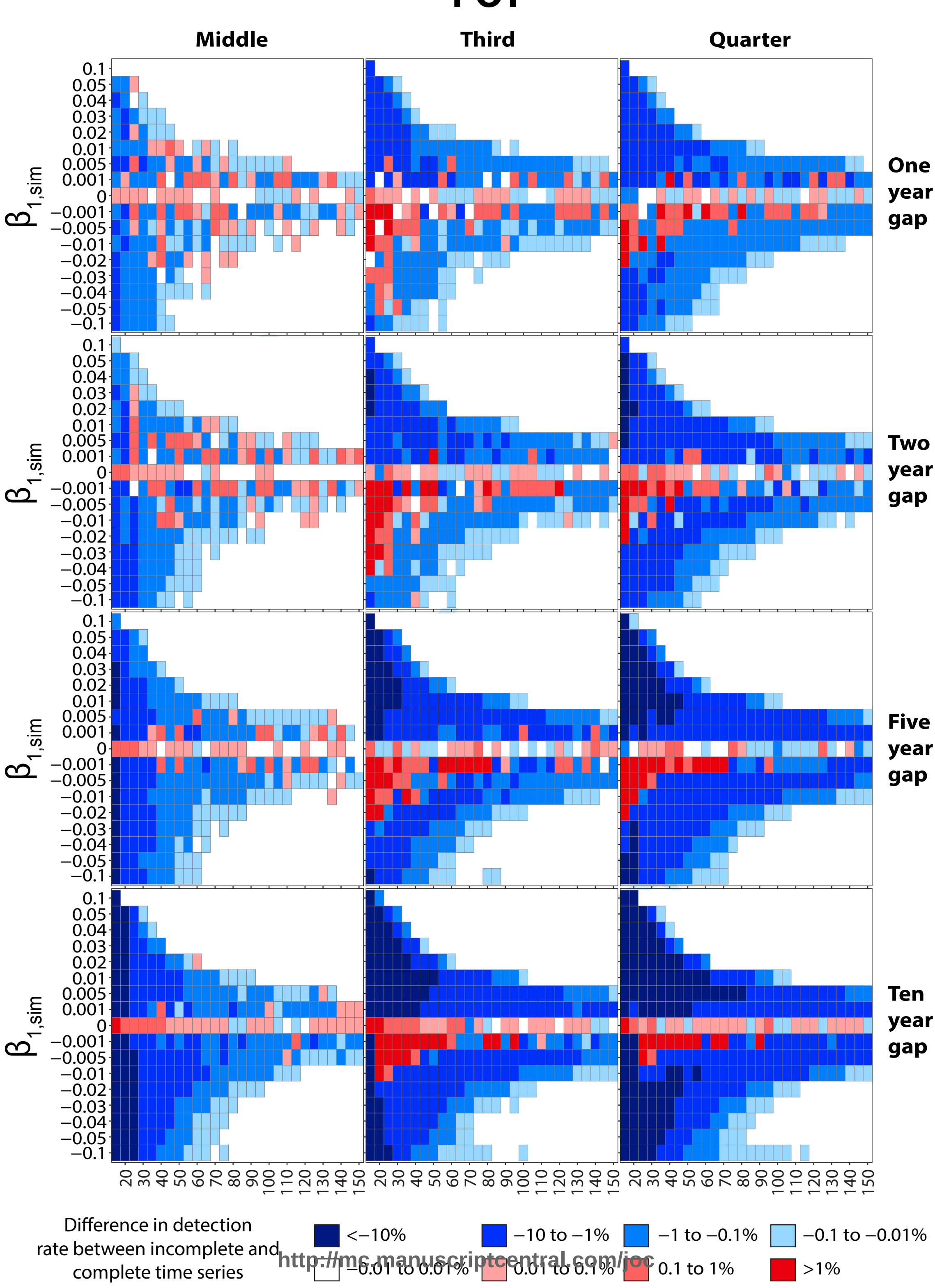
POT
Middle
Third
Quarter
$\beta_{1,sim}$
0.1
0.05
0.04
0.03
0.02
0.01
0.005
0.001
0
−0.001
−0.005
−0.01
−0.02
−0.03
−0.04
−0.05
−0.1
One year gap
Two year gap
Five year gap
Ten year gap
20 30 40 50 60 70 80 90 100 110 120 130 140 150
Difference in detection rate between incomplete and complete time series
<−10%
−10 to −1%
−1 to −0.1%
−0.1 to −0.01%
−0.01 to 0.01%
0.01 to 0.1%
0.1 to 1%
>1%

Figure 2

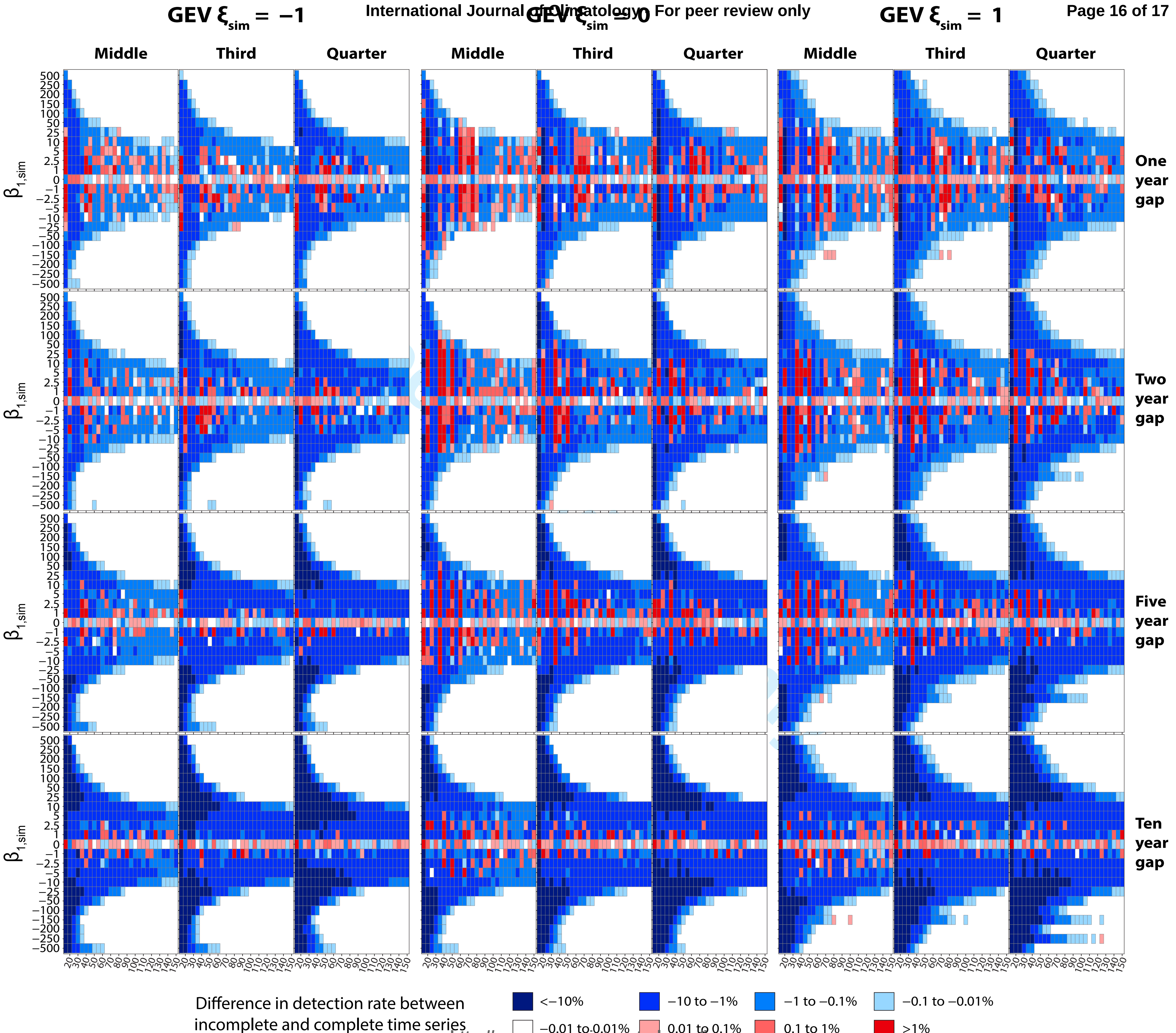
GEV ξ_{sim} = −1
GEV ξ_{sim} = 0
GEV ξ_{sim} = 1
Middle
Third
Quarter
$\beta_{1,sim}$
500 250 200 150 100 50 25 10 5 2.5 1 0 −1 −2.5 −5 −10 −25 −50 −100 −150 −200 −250 −500
20 30 40 50 60 70 80 90 100 110 120 130 140 150
One year gap
Two year gap
Five year gap
Ten year gap
Difference in detection rate between incomplete and complete time series
<−10%
−10 to −1%
−1 to −0.1%
−0.1 to −0.01%
−0.01 to 0.01%
0.01 to 0.1%
0.1 to 1%
>1%

Figure 3

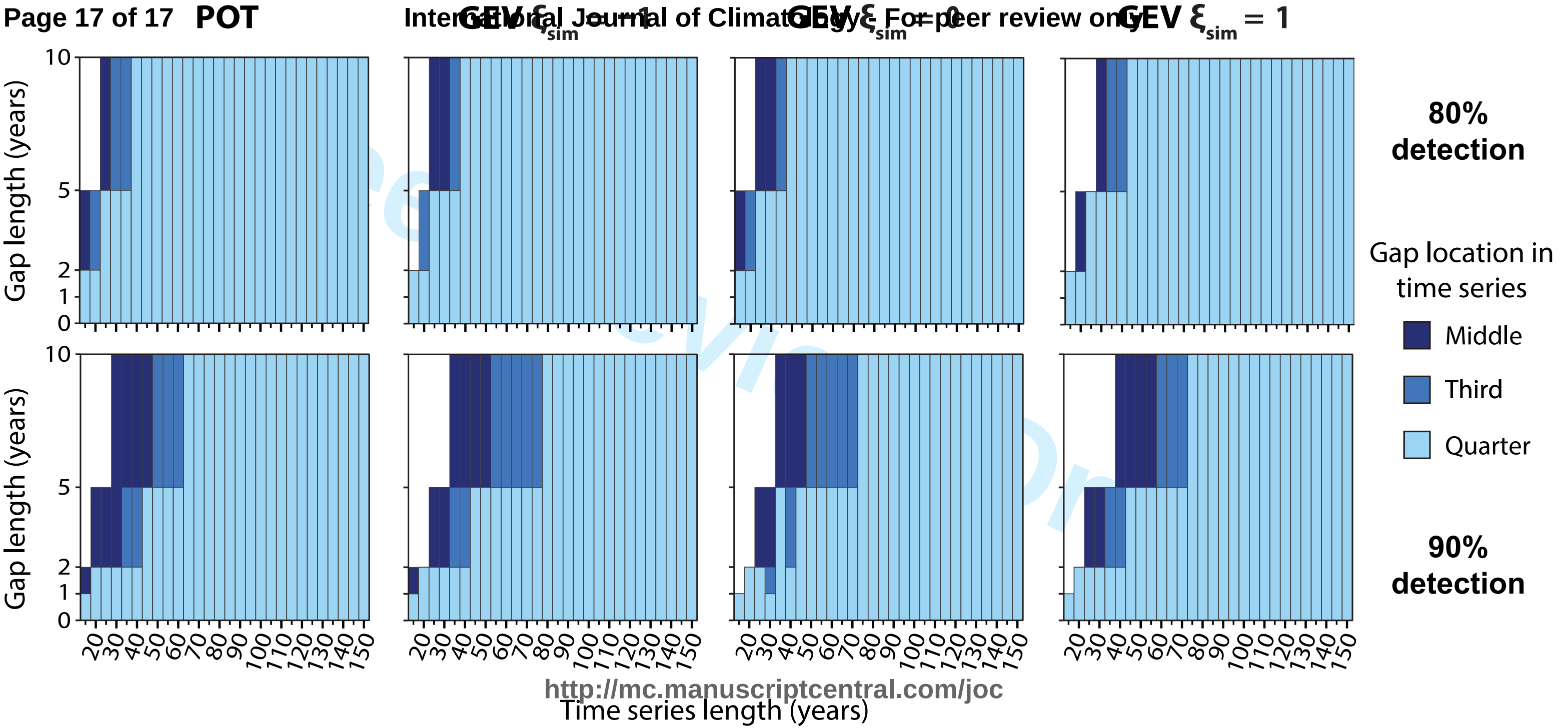
POT
GEV ξ_{sim} = −1
GEV ξ_{sim} = 0
GEV ξ_{sim} = 1
Gap length (years)
10
5
2
1
0
20
30
40
50
60
70
80
90
100
110
120
130
140
150
Time series length (years)
80% detection
90% detection
Gap location in time series
Middle
Third
Quarter

Figure 4