Journal of Hydrology 530 (2015) 117–126

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Hypothesis tests for hydrologic alteration

Charles N. Kroll [a,*], Kelly E. Croteau [a,b], Richard M. Vogel [c]

[a] Environmental Resources Engineering, SUNY ESF, Syracuse, NY, United States
[b] Windward Environmental, Syracuse, NY, United States
[c] Civil and Environmental Engineering, Tufts University, Medford, MA, United States

ARTICLE INFO

Article history:
Received 26 June 2015
Received in revised form 30 August 2015
Accepted 23 September 2015
Available online 28 September 2015
This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Felix Frances, Associate Editor

Keywords:
Flow duration curves
Hydrologic alteration
Hypothesis testing

SUMMARY

Hydrologic systems can be altered by anthropogenic and climatic influences. While there are a number of statistical frameworks for describing and evaluating the extent of hydrologic alteration, here we present a new framework for assessing whether statistically significant hydrologic alteration has occurred, or whether the shift in the hydrologic regime is consistent with the natural variability of the system. Four hypothesis tests based on shifts of flow duration curves (FDCs) are developed and tested using three different experimental designs based on different strategies for resampling of annual FDCs. The four hypothesis tests examined are the Kolmogorov–Smirnov (KS), Kuiper (K), confidence interval (CI), and ecosurplus and ecodeficit (Eco). Here 117 streamflow sites that have potentially undergone hydrologic alteration due to reservoir construction are examined. 20 years of pre-reservoir record is used to develop the critical value of the test statistic for type I errors of 5% and 10%, while 10 years of post-alteration record is used to examine the power of each test. The best experimental design, based on calculating the mean annual FDC from an exhaustive jackknife resampling regime, provided a larger number of unique values of each test statistic and properly reproduced type I errors. Of the four tests, the CI test consistently had the highest power, while the K test had the second highest power; KS and Eco always had the lowest power. The power of the CI test appeared related to the storage ratio of the reservoir, a rough measure of the hydrologic alteration of the system.

1. Introduction

River systems provide an array of services to humans and the environment. These systems are used to meet human needs, including domestic, industrial, and agricultural water supplies, waste disposal, hydropower, and recreational activities. They also provide critical habitat for many aquatic and non-aquatic species. Anthropogenic activities and climatic influences can alter the hydrology of river systems. Activities such as damming a river, water discharges and withdrawals, and regional variations in climate can alter the hydrologic system and impact freshwater biodiversity and ecosystem services (Bunn and Arthington, 2002; Magilligan and Nislow, 2005; Gao et al., 2009). There is clearly tremendous interest in understanding ecological responses to altered flow regimes as evidenced by the hundreds of citations to the recent review article on this topic by Poff and Zimmerman (2010).

A wide variety of metrics has been developed to assess changes in hydrologic systems (Olden and Poff, 2003; Gao et al., 2009). A common set of metrics is the Nature Conservancy's Indicators of Hydrologic Alteration (IHA), which describe changes in 33 hydrologic statistics that characterize a wide array of hydrologic function (Richter et al., 1996). The IHA are often used to assess the impact of human activities on hydrology and to determine environmental flow recommendations for water managers. The ecological limits of hydrologic alteration (ELOHA) provide a framework for linking statistics such as those in the IHA to critical ecological responses (Poff et al., 2010). In the ELOHA framework, relationships between altered flow and ecological characteristics are empirically developed using existing and newly collected field data (Arthington et al., 2006). Similarly, numerous empirical multivariate relationships have been developed which characterize the impact of various anthropogenic influences on streamflow regimes ranging from flood regimes (Fitzhugh and Vogel, 2011) and low flow regimes (Homa et al., 2013) to the entire flow regime (McManamay, 2014). Fitzhugh (2014, page 826) reviews numerous recent studies which have sought to characterize alteration of a streamflow regimes over regions of the U.S.

Even with the broad suite of metrics of hydrologic change which have been introduced, as well as numerous empirical multivariate statistical models of the relationship between streamflow

* Corresponding author.
E-mail addresses: cnkroll@esf.edu (C.N. Kroll), kmetz878@gmail.com (K.E. Croteau), richard.vogel@tufts.edu (R.M. Vogel).

http://dx.doi.org/10.1016/j.jhydrol.2015.09.057
0022-1694/

regimes and anthropogenic factors, it is often difficult to assess whether changes to the hydrologic system are significant or are instead simply a result of the natural variability of streamflow under stationary conditions (Burn and Hag Elnur, 2002). While one could perform a hypothesis test to determine if there were significant changes in particular IHA statistics (Magilligan and Nislow, 2005) or other relevant hydrologic statistics, it is more challenging to assess changes to the complete streamflow regime. While many previous studies have examined tests of trends (e.g. Douglas et al., 2000; Burn and Hag Elnur, 2002) and shifts (e.g. Salas and Boes, 1980; Buishand, 1984) in hydrologic series, here we explore tests to assess the significance of an alteration to the complete hydrologic series.

One tool utilized in this experiment is the river's flow duration curve (FDC) (Foster, 1924; Searcy, 1959). The FDC is a plot of typically mean daily streamflow versus the probability of exceeding that streamflow. The FDC covers the entire range of streamflow magnitudes, and thus integrates the complete streamflow regime into a single tool. FDCs have been employed for a wide range of applications, including hydropower design, habitat assessment, flood abatement, water quality evaluation and for comparative hydrologic assessments (Vogel and Fennessey, 1995; Castellarin et al., 2013). An FDC is a convenient tool for observing and understanding hydrologic change. For instance, if a reservoir was placed in a river, one might expect a flattening of the FDC, where the higher flows are reduced (due to flood storage) and the lower flows are increased (by augmenting low flows with reservoir releases). Water withdrawals or reduction in precipitation would generally result in a lowering of the entire FDC. Botter et al. (2008) provide a theoretical linkage between the structure of FDC's and underlying ecohydrological, climatic and other watershed processes. Castellarin et al. (2013) provide a detailed review of the influence of a variety of natural and anthropogenic influences on FDC's and associated hydrologic processes.

FDCs and FDC statistics can be employed for assessing hydrologic alteration. Using shifts in FDCs, Vogel et al. (2007) defined the ecodeficit and ecosurplus as the percent loss or gain in streamflow due to flow regulation. Gao et al. (2009) employed a principal component analysis to examine how IHA statistics were related to ecodeficit and ecosurplus. Homa et al. (2013) developed regional regression models for quantiles of an FDC at altered streamflow sites in Massachusetts. Similarly, Mejia et al. (2014) derived a stochastic model of FDC's suitable for 11 urbanizing Washington, DC–Baltimore basins based on the stochastic properties of rainfall and various watershed properties.

Here we develop four new hypothesis tests of hydrologic alteration based on shifts in FDCs. Developing a hypothesis test based on the complete FDC, instead of a single hydrologic statistic, poses a unique challenge requiring a sampling strategy to implement each hypothesis test. Here we are faced with evaluating both the effectiveness of a number of different sampling strategies in addition to the power of the resulting hypothesis tests. Each hypothesis test is defined by its 'test statistic', whereas the sampling strategies are held fixed across all hypothesis tests considered. The hypothesis tests are developed similarly to common tests of a change in the probability distribution of a series. Two of the tests are based on deviations between the cumulative distribution functions (cdfs) of the FDCs. The test statistics for these tests are similar to those employed in the well-known Kolmogorov–Smirnov (Smirnov, 1948) and Kuiper's (Kuiper, 1960) tests. One of the other two tests is based on exceeding confidence interval-type bounds on the FDC, while the final test is based on the combined ecosurplus and ecodeficit, which has been termed ecochange.

To develop such tests, we could develop a hydrologic model for a basin, perturb a parameter or management scheme in the model, and assess the significance of the change in the FDCs due to the magnitude of the perturbation. While this would be controlled experiment, it would be reliant on how well the model represents reality, and how well the model perturbation represented a change in the hydrologic system. Instead we choose a method similar to Burn and Hag Elnur (2002) and Douglas et al. (2000), where measured streamflow sequences are employed along with a resampling strategy to assess the significance of the test. Unlike those studies which examined the significance of regional trends, here we develop a test to assess the significance of hydrologic alteration at a single streamflow site. Our tests rely upon having a period of record (here 20 years) during which it is assumed there is no alteration in the hydrologic series, and some periods of record (here assumed 5 years) after which a potential alteration has occurred to evaluate the significance of the alteration.

For each hypothesis test considered, 20 years of unaltered daily streamflow are used to develop the critical value of the test statistic corresponding to type I error probabilities of 5% or 10%. This is done by using either annual FDCs, or median or mean annual FDCs (see Vogel and Fennessey, 1994, for definitions of mean and median annual FDC's) obtained via an exhaustive jackknife resampling of the 20 year record in 5 year increments. Once each test is developed, the power of the test $(1-\beta)$ is assessed, where β is the probability of a type II error. Power is assessed by an exhaustive jackknife resampling of 10 years of annual or jackknifed median or mean FDCs of potentially altered streamflows in 5 year increments. The significance of the alteration could be assessed with one 5 year post-alteration sequence. In practice, both type I and type II errors are of concern. Type I errors correspond to overprotecting the environment; type II errors, which are potentially worse than type I errors, correspond to not protecting the environment when we really should have.

For a case study, streamflow alteration in this experiment is due to the construction of a reservoir (Magilligan and Nislow, 2005). A subset of the reservoir sites employed by Poff et al. (2007) and Gao et al. (2009) that have a relatively long historic daily streamflow record both before and after construction of the reservoir are analyzed. It is assumed that the construction of a reservoir will produce a significant alteration of the streamflow record (which may not be true at all sites), and that no other forms of hydrologic alteration are impacting these records. The proposed hypothesis test framework provides a framework to assess hydrologic alteration, which could then inform water management decisions.

2. Development of test statistics

In this section, we describe four FDC-based hypothesis tests of hydrologic alteration. The resampling scheme to determine the critical values of the test statistic, significance and power of each test is held fixed across tests. For each test, the null hypothesis (H_o) is that there is no hydrologic alteration, and the alternative hypothesis (H_a) is that there is hydrologic alteration. Two of the tests are based on common hypothesis tests of distributional change (Kolmogorov–Smirnov and Kuiper), while the other two are based on observed shifts in the FDC. In this section (Section 2), the test statistic for each hypothesis test is described. In the following section (Section 3), the methodology to develop the critical values of each test statistic and the power of each test is discussed, and the reservoir sites employed in this analysis are presented.

2.1. Test 1: Kolmogorov–Smirnov test (KS)

The 2-sample KS test is a non-parametric hypothesis test where the null hypothesis is that two samples are drawn from the same distribution. The KS test compares the cumulative distribution function (cdf) of two data sets, and computes a test statistic based