Decision trees for incorporating the likelihood of excessive hydrologic alteration into hydropower-ecosystem tradeoffs

Jory S. Hecht[1,2*], Richard M. Vogel[1], Ryan A. McManamay[3], Charles N. Kroll[4], J. Michael Reed[5]

[1]Department of Civil and Environmental Engineering, Tufts University, Medford, MA

[2]Vermont EPSCoR, University of Vermont, Burlington, VT

[3]Urban Dynamics Institute, Oak Ridge National Laboratory, Oak Ridge, TN

[4]Department of Civil and Environmental Engineering, State University of New York – College of Environmental Science and Forestry, Syracuse, NY

[5]Department of Biology, Tufts University, Medford, MA

[*]Corresponding author

Final report to Hydro Research Foundation

Article to be submitted to the *Journal of Hydrology*

Adapted from dissertation chapter submitted in May 2017

Abstract

Detrimental ecological impacts have often been observed or anticipated when changes in streamflow indicators exceed percent deviation thresholds believed to be ecologically critical. Yet, short pre- and post-impact flow records often make it difficult to determine whether changes exceeding tolerable thresholds are due to dam operations or natural variability. Through a hypothetical reservoir operations example, we incorporate the uncertainty of dam-induced streamflow changes into a Bayesian decision tree framework that evaluates tradeoffs between expected regrets associated with hydropower and ecology. The likelihood of over-protection (type I) and under-protection (type II) errors associated with hypothesis tests are used to compute expected hydropower and ecosystem regrets associated with dam operation decisions. We examine changes to high (annual Q5) and low flows (annual Q95) in typical years using a modified and extended nonparametric ranked-sum test that accounts for percent deviation thresholds. A multiple comparison test is then used to determine the likelihood of at least one threshold violation. An example shows that our decision-theoretic approach can lead to different dam operation recommendations than do other common methods, and highlights limitations that arise when the type I error rate is selected *a priori*. While we illustrate a hydropower-ecosystem tradeoff, our approach can also be applied to other multi-stakeholder reservoir and river management conflicts.

Key Words: Bayesian decision-making, environmental flows, flow duration curve, reservoir operation, type I error, type II error

1. Introduction

Reservoirs provide storage for water supply, hydroelectric production, flood control, irrigation, recreation, and other conservation needs. The retention and selective release of water downstream modifies the flow regimes that sustain riverine ecosystems. As dams alter pre-existing flow regimes in myriad ways (McManamay *et al.*, 2013), their effects on downstream ecosystems are diverse (e.g. Carlisle *et al.*, 2011). The composition of species in riverine ecosystems often changes following the commissioning of dams with large storage capacities that regulate high- and low-flow extremes (e.g., Poff *et al.*, 2007; Mims and Olden, 2013). Prescribing operating rules that sustain pre-dam ecosystems has become an increasingly recognized challenge, especially when flow manipulation benefits off-stream human interests (e.g., Cardwell *et al.*, 1996; Suen and Eheart, 2006; Poff *et al.*, 2009). Short streamflow records make it more difficult to determine whether ecosystem changes are due to dam impacts or natural variability (e.g. Kennard *et al.*, 2010; Nikghalb *et al.*, 2016). Moreover, there are few guidelines for incorporating uncertainty stemming from short streamflow records into these tradeoffs.

Streamflow has been identified as the primary driver in riverine ecosystems (Power *et al.*, 1995; Walker *et al*, 1995) since it influences ecological conditions through changes in velocity, depth, temperature, water quality and substrate (e.g. Poff *et al.* 1997; Jager, 2014; McManamay *et al.*, 2015). Indeed, many interconnections between streamflow and habitat conditions are indirect, which can moderate any apparent correlation between hydrologic alteration and ecological responses (e.g. Jager, 2014; McManamay, 2015). Yet, the availability of streamflow data relative to other proximate indicators of ecological degradation has generated tremendous interest in characterizing ecological responses to flow alteration, including reviews by Poff and

Zimmerman (2010) and Webb *et al.* (2013). Flow-based approaches to managing environmental flows can also be implemented in data-poor regions where hydrologic and ecological monitoring can be more challenging (McKay, 2015; Lamouroux, 2015; Eriyagama *et al.*, 2016). Institutions representing ecological interests, such as The Nature Conservancy, have recently promoted expert-elicited flow-ecology relationships in stakeholder negotiations (e.g., Kendy *et al.*, 2012; Steinschneider *et al*., 2015). In this paper, we introduce a tool suited for assessing impacts to species and communities whose overall welfare (e.g. abundance, diversity) is well correlated with deviations in flows from pre-impact conditions.

Determining the extent to which dam operations change a flow regime is not always straightforward. Indeed, dam operations can be deduced through water balance equations if there are flow gauges both up- and down-stream of the dam, or reservoir water level measurements and an up- or down-stream gauge. However, in other cases, only a downstream gauge may exist. In such situations, how does one evaluate the likelihood that differences in exceeding perceived ecological thresholds between pre- and post-dam periods arise from dam operations alone, rather than due solely to the natural random variability of streamflow between the two periods? These uncertainties, which are especially large when there are short pre- and post-dam streamflow records (e.g. Kennard *et al*., 2010; Williams, 2017), confound efforts to develop relationships between flow alteration and ecological responses.

To date, only a handful of studies have examined hypotheses of dam-induced hydrologic alteration within a statistical framework (Botter *et al.*, 2010; Kennard *et al.*, 2010; FitzHugh, 2014; Taylor *et al.*, 2014; Kroll *et al.*, 2015). Importantly, these studies evaluated changes in the natural flow regime without considering thresholds of alteration beyond which species may be adversely impacted (see Poff and Zimmerman, 2010; Kendy *et al*., 2012; Steinschneider *et al*., 2014), a unique feature considered in this study. The strict preservation of the natural flow regime is often impractical or infeasible (Kendy *et al.*, 2012; Kopf *et al.*, 2015). Even if an ecosystem indicator worsens with flow alteration, stakeholders may still consider a maximum allowable degree of alteration. Thresholds expressed in terms of percent deviations from pre-dam conditions can be applied across sites exhibiting similar ecosystem and hydrologic behavior without the detailed site-specific information necessary for identifying flow magnitude thresholds. They have also been increasingly advocated because they preserve natural flow variability better than alternative policies, such as those requiring constant or seasonally varying minimum flows (e.g. Smakhtin *et al.*, 2004; Vogel *et al.*, 2007; Richter, 2009; Richter *et al.*, 2012; Razurel *et al.*, 2015; Rheinheimer *et al*., 2016).

While hypothesis testing has been increasingly applied to characterize changes in flow statistics deemed important for managing riverine ecosystems (e.g. FitzHugh, 2014; Taylor *et al*., 2014; Kroll *et al*., 2015), reservoir operators are left with little guidance regarding how to use such tests for prescribing reservoir release operating rules. Hypothesis test results have potential societal consequences, as they can lead to unnecessary changes in operating rules that reduce reservoir benefits or avoidable ecological consequences when release rules are not changed. We apply a Bayesian decision-tree framework based on statistical decision theory (e.g. Wald, 1939; Berger, 1993) to incorporate the uncertainty associated with decisions concerning violations of alteration thresholds into evaluations of tradeoffs between flow benefits for off-stream (hydropower) and in-stream (ecosystem) users. Decision trees have been applied to other water resources problems concerned with uncertain environmental changes, including the potential ecological effects of effluent discharge (Mapstone, 1995), planning hydraulic infrastructure under uncertain future lake levels (e.g. Hobbs *et al.*, 1997), building barriers for possible stormsurge increases (Rosner *et al.*, 2014) and optimization of removal of barriers to fish passage (O'Hanley and Tomberlin, 2005). To highlight the value of our probabilistic framework, we compare our findings with results obtained using (i) non-probabilistic methods and (ii) null hypothesis significance testing (NHST), a widely-used approach that examines only the chance of falsely concluding a threshold violation, i.e. a type I error. For illustrative purposes, we compare flows at a downstream station with and without the reservoir in place using a Before-After impact analysis. This experimental design ensures that differences between the two flow regimes arise from either random sampling variability or dam operations, thus avoiding the need to consider confounding factors that may explain differences between them.

Our decision-oriented hypothesis testing approach can be applied to reservoirs and dams with a range of off-stream benefits, including water supply, hydropower, and flood control. However, in this study, we demonstrate the effects of seasonal flow alteration from a baseload hydropower dam. Over the next two decades, many hydropower dams are slated to be built in highly biodiverse basins with short streamflow records (Winemiller *et al.*, 1996; Zarfl *et al.*, 2015) and little transparency about operating rules (e.g. Lauri *et al.*, 2012). While studies examining hydropower-ecological tradeoffs under different dam operating rules have proliferated (e.g. Halleraker *et al.*, 2007; Renöfalt *et al.*, 2010; Yang and Cai, 2011; Jager, 2014), including ones revising operating rules when renewing licenses (e.g. Pearsall *et al.*, 2005), few studies share our focus on the extent to which the uncertainty of differences between pre- and post-dam flows affect dam operating decisions.

The remainder of the paper is structured as follows. First, we formulate a hypothesis test for examining the likelihood that changes in high- (annual Q5) and low-flow indicators (annual Q95) indicators exceed percent-deviation thresholds – and are not due to random sampling variability. Next, we integrate a Bayesian decision tree framework, which considers the likelihood of under- and over-design interpreted from Type I and II hypothesis testing errors, respectively. We then describe the hypothetical baseload hydropower reservoir with which we demonstrate our method, before discussing limitations and possible extensions of our decision-theoretic approach and offering concluding remarks.

2. Testing for violations of hydrologic alteration thresholds

This section describes our approach for examining whether changes to ecologically important AFDC quantiles violate percent-deviation thresholds. First, we present a hypothesis test that detects violations for individual AFDC quantiles. Then, we present a multiple comparison (field significance) test to determine the likelihood of one or more threshold violations associated with an ecologically critical AFDC quantile.

2.1 Annual flow duration curves

Another challenge with implementing hydrologic methods is choosing indicators of alteration that recognize the distinct ways in which different riverine species respond to hydrologic alteration, yet can easily be incorporated into reservoir operation rules. Flow duration curves (FDCs) indicate the probability that a daily flow of a given magnitude will be exceeded. An FDC may be thought of as a graphical signature which summarizes a river's overall hydrologic behavior. They have been used in an extremely wide range of water resources applications, including hydropower design, habitat assessment, flood abatement, and water quality evaluation (Vogel and Fennessey, 1995; Castellarin *et al.*, 2013). They underpin environmental flow management in data-poor regions, including India (Jain, 2015) and Sri Lanka (Eriyagama *et al.*, 2016) and are useful in regions where other indicators inform environmental flows management (e.g., Tennant *et al.*, 1976; Renöfalt *et al.*, 2010). While period-of-record FDCs computed from daily flows in pre- and post-dam periods offer a signature of the flow variability over an entire station record, they cannot assess changes in typical years between pre- and post-dam periods. In contrast, annual FDCs (AFDC's), introduced by Vogel and Fennessey (1994), represent the variability of flows within a single water year. Sets of AFDC's depict the within- and between-year hydrologic variability and can be used to construct confidence intervals for FDCs (Vogel and Fennessey, 1994). In addition, the median AFDC can reveal high- and low-flow conditions representative of a typical year. Figure 1 shows pre- and post-dam AFDCs from the stylized example presented in Section 4, in which we examine long-term decreases in typical annual values of high in-channel flows (Q5) and long-term increases in typical low flows (Q95) due to the flow homogenization effects of baseload hydropower. High in-channel flows are essential for flushing sediment and pollutants and are often correlated with ecologically critical flood flows. Meanwhile, low-flow increases can cause drought-tolerant, native species to be replaced with generalist species that favor conditions created by less seasonally variable flows (Carlisle *et al*., 2011; Mims and Olden; 2013, Nikghalb *et al.*, 2016). The acute flat-lining effect visible in the post-dam flow plot indicates the turbine discharge capacity that constrains reservoir releases when storage is adequate. While AFDC's do not indicate temporal flow sequences, hypothesis tests of AFDC's can address some ecologically important timing issues if the intervals over which FDCs are computed match the timescale of a problem, e.g., using seasonal FDCs for changes in hydrologic conditions during spawning periods (Gao *et al*., 2009).

[INSERT FIGURE 1]

2.2 Threshold violations of individual AFDC quantiles

Our hypothesis test determines the likelihood that differences in an AFDC quantile between pre- and post-dam periods in excess of a tolerable percent deviation threshold are not due to random sampling variability, but rather to dam operations, i.e. a proof by contradiction (Cohn and Lins, 2005). Conventional hypothesis tests with a null hypothesis of no change, i.e. no threshold violation, and an alternative hypothesis of change, i.e. a threshold violation, accommodate this dichotomy of “acceptable” and “unacceptable” changes well (Figure 2). If a threshold violation is not detected, then it implies we should keep existing dam operating rules. Conversely, a violation implies a need to change reservoir operations to ‘protect the environment’.

Figure 2 shows that a type I error corresponds to the likelihood of detecting a threshold violation when, in fact, there is no violation. We denote this type I error probability $\alpha = P(CA|NA)$, where $P(CA)$ is the probability of concluding alteration, and $P(NA)$ is the probability of no alteration. A type I error amounts to an overdesign error and implies that hydropower production would be reduced unnecessarily if a more ecosystem-friendly operating rule were chosen. The probability of a type II error indicates the likelihood of not detecting a threshold violation when there is one. It amounts to an under-design error from an ecosystem protection perspective and signals the possibility of adverse ecosystem impacts. We denote this probability □ $= P(CNA|A)$, where $P(CNA)$ is the probability of concluding no alteration beyond a percent-deviation threshold from the test and $P(A)$ is the probability of a violation of an alteration threshold. Of perhaps greatest interest is the power of the hypothesis test 1-□□ which reflects our ability to detect alteration when present.

[INSERT FIGURE 2]

A two-sample hypothesis test comparing pre- and post-dam AFDC quantiles is needed to evaluate the likelihood of making decisions concerning violations of various alteration thresholds. For instance, this test could compare the distributions of *m* different annual Q95 values in the pre-dam period with *n* different annual Q95 values from the post-dam period. The overarching objective is to classify the difference between the pre- and post-dam AFDCs as either: (i) alteration within a tolerable threshold, i.e., "no alteration", or (ii) alteration exceeding a tolerable threshold, i.e., "alteration".

Kroll *et al.* (2015) found two promising tests for testing for changes of any magnitude in annual flow duration curves. These two tests offer a viable starting point for devising a hypothesis test that can assess changes in thresholds of stakeholder-identified flow indicators. However, these two tests pose problems for assessing hydropower-ecosystem tradeoffs. First, their confidence interval (CI) test assumes that AFDC quantiles are normally distributed, which is often not the case downstream of hydropower dams due to turbine release constraints and other operational caveats (Botter *et al*., 2010; FitzHugh, 2014). Their nonparametric Kuiper test (Kuiper, 1960), which accommodates non-normal distributions, identifies the maximum positive and negative differences between the cumulative probabilities of pre- and post-dam flows to assess the likelihood of distributional change. While accounting for these two differences make it suitable for analyzing reservoir-induced annual flow regulation, it does not examine changes in pre-determined flow durations, indicators that are often used to assess flow alteration in environmental flows management (Kendy *et al*., 2012).

In contrast, the nonparametric Mann-Whitney-Wilcoxon test can be used to evaluate differences between individual AFDC quantiles of interest without concern over distributional hypothesis of AFDC quantiles. A multiple comparisons test can then be used to draw conclusions about the overall likelihood of hydrologic alteration from tests applied to individual quantiles. Similar field significance approaches for estimating the overall likelihood of type I and

II errors from a series of hypothesis tests assessing changes in individual indicators of concern have been advocated for other infrastructure design problems recently (Reiff *et al.*, 2016).

2.3 Mann-Whitney-Wilcoxon test

We first illustrate the Mann-Whitney-Wilcoxon (MWW) test for a simplified case in which any change in flow for a given AFDC exceedance probability constitutes alteration. This test assesses the likelihood that the difference between pre- and post-dam distributions of a given AFDC quantile, Q_{pre} and Q_{post}, belongs to either a "no alteration" class corresponding to a null hypothesis of no change in distribution, or an "alteration" class corresponding to an alternative hypothesis of changes in distribution. First, to compute the type I error, we can express the null hypothesis as $H_0: Q_{pre} = Q_{post}$ and the alternative hypothesis as either $H_A: Q_{pre} > Q_{post}$ or $H_A: Q_{pre} < Q_{post}$ depending on the direction of hypothesized change. In Section 2.4, we describe an adjustment to the test that enables us to examine threshold exceedance hypotheses. While our Before-After experimental design may suggest the need for a paired sample hypothesis test, such as the Wilcoxon signed-rank test, the MWW test can also be applied to independent, unpaired samples representing pre- and post-dam records of different lengths. See Yue and Wang (2002) for a detailed appraisal of this test for different sample distributions and properties.

[INSERT FIGURE 3]

Figure 3 shows an example of the distribution of AFDC quantiles corresponding to a possible post-dam decrease in a flow indicator. Two possible AFDC quantile outcomes, "no alteration" and "alteration", each have probability distributions with locations defined using the MWW test statistic *U*. *U* describes the difference between the AFDC quantiles corresponding to the two samples of lengths *m* and *n*, respectively. To compute it, each flow value in a sample of length *m* corresponding to an AFDC quantile of interest *m* is paired with every flow corresponding to the same AFDC quantile in the sample of length *n*, yielding $m*n$ pairs. The test statistic *U* summarizes the number of pairs for which the alternative hypothesis is true, which, in turn, indicates the extent to which a given AFDC is stochastically greater than the other. In contrast, a value of zero means that all post-dam flows corresponding to a given AFDC quantile are greater than all pre-dam observations, and is the strongest evidence possible against an alternative hypothesis of lower pre-dam flows. A value of $(m*n)/2$ would signal that the two samples cannot be distinguished from each other. Mathematically, the *U* test statistic is computed as follows:

$$U = \sum_{i=1}^{m} \sum_{j=1}^{n} \varphi \left(Q_{pre_i} - Q_{post_j} \right) \tag{1}$$

where $\varphi \left(Q_{pre_i} - Q_{post_j} \right) = 1$ if $Q_{pre_i} - Q_{post_j} > 0$ and is equal to 0 if $Q_{pre_i} - Q_{post_j} \leq 0$. Another unique feature of this test is that Shieh *et al.* (2006) derived the probability distribution of *U* under both the null and alternative hypotheses. When both samples are larger than eight, a standardized *U* statistic, termed *Z*, may be approximated by a standard normal distribution under the null hypothesis:

$$Z = \frac{(U - \mu_0)}{\sigma_0} \tag{2}$$

where μ_0 and σ_0 are the mean and standard deviation of the *U* statistic under the null hypothesis, and are both functions of only the known sample sizes *m* and *n*:

$$\mu_0 = \frac{m * n}{2} \tag{3}$$

$$\sigma_0 = \frac{m * n * (m + n + 1)}{12} \tag{4}$$

When the size of both samples is less than eight, an exact empirical distribution of *U* under the null hypothesis should be computed from the ranks using formulae given by Mann and

Whitney (1947). Bellera *et al.* (2010) describes applications of this test to small samples of unequal length. Type I error probability estimates from this test may be distorted if the variances of $Q_{pre}$and Q_{post} differ significantly and corrections to the null hypothesis assumption of equal variances are not made (Kasuya, 2001).

In contrast with conditions under the null hypothesis, determining the probability distribution associated with U under the alternative hypothesis is much less straightforward because the distribution of U is unknown. Unless a Markov chain Monte Carlo method is applied (Lee, 2014), one must define an alternative hypothesis based on a given distributional assumption (Blair and Higgins, 1980; Shieh *et al.,* 2006) or run Monte Carlo simulations (e.g. Neave and Granger, 1968; Yue and Wang, 2002; Kroll *et al*., 2015) to determine type II errors. Kroll *et al.* (2015) demonstrate that a normal distributional hypothesis associated with the estimated quantiles from an AFDC could not be rejected across 80 percent of all exceedance probabilities considered in a set of 20-year pre-dam and 10-year post-dam daily flow records at 117 United States Geological Survey (USGS) stations. We have found that this assumption is especially suitable for quantiles near the median flow (Q50). In this initial study, we assume alternative hypotheses in which AFDC quantiles are well approximated by standard normal distributions and standardize both Q_{pre} and Q_{post} with the pre-dam mean and standard deviation. In practice, transformations could normalize AFDC quantile distributions, if needed. We apply an analytical large-sample method from Shieh *et al.* (2006) based on standard normal distributional assumptions to estimate the probability of type II errors associated with a MWW test. Shieh *et al.* (2006) express the mean of the test statistic U under the alternative hypothesis, which we term μ_A , as a function of the two sample sizes and the difference between their sample means $\hat{\theta}$:

$$\mu_A = mn * \Phi\left(\frac{\hat{\theta}}{\sqrt{2}}\right) \tag{5}$$

where $\Phi()$ denotes the cdf of a standard normal variate. Shieh *et al*. (2006) give the standard deviation of *U* under the alternative hypothesis as:

$$\sigma_A^2 = m * n * \left\{\Phi\left(\frac{\hat{\theta}}{\sqrt{2}}\right) * \left[1 - \Phi\left(\frac{\hat{\theta}}{\sqrt{2}}\right)\right] + (m + n - 2) * \left(E\left[\left\{\phi\left(Z + \hat{\theta}\right)\right\}^2\right] - \left[\Phi\left(\frac{\hat{\theta}}{\sqrt{2}}\right)\right]^2\right)\right\} \tag{6}$$

where Z is a standard normal random variable defined in (2). To evaluate σ_A^2, we approximate the term $E\left[\left\{\Phi\left(Z + \hat{\theta}\right)\right\}^2\right]$ by numerically computing the average value of $\left\{\Phi\left(Z + \hat{\theta}\right)\right\}^2$ using the *Z* scores for quantiles ranging from 0.0005 to 0.9995. Using μ_A and σ_A, we then estimate the power of the test $(1 - \beta)$ as follows:

$$1 - \beta = P\{U > \mu_0 + z_\alpha \sigma_0\} = \phi\left(\frac{\mu_A - \mu_0 - z_\alpha \sigma_0}{\sigma_A}\right) \tag{7}$$

where α is the type I error probability and z_α is the $100(1 - \alpha)$ percentile of the standard normal distribution. For a given value of z_α, the power $1 - \beta$ rises as the difference between the means of the alternative and null distributions of *U* increases. Figure 4 illustrates the tradeoff between the likelihood of type I and type II errors, and shows that, if one is more concerned with designing a test with a low false positive (type I error) rate, there is a greater chance of obtaining false negatives (type II errors). In other words, the likelihood of over-protection and the likelihood of under-protection are inversely related. The concavity of this tradeoff curve increases with (i) the difference between pre- and post- dam flows and (ii) the length of the station record before and after dam is commissioned.

[INSERT FIGURE 4]

2.4 Adapting the MWW test for use with percent deviation thresholds

Next, we describe how to modify the MWW test to determine the likelihood that a typical pre-dam flow deviates from a typical post-dam flow by more than a given threshold. The MWW test cannot directly evaluate hypotheses regarding percent deviation thresholds because the values of the observations are transformed into ranks. Thus, for the MWW test to account for the percent deviation thresholds being examined, we must scale the pre-dam flows Q_{pre} with the tolerated percent deviation before testing hypotheses that pre-dam flows differ from their post-dam counterparts so that:

$$Q_{pre-scaled} = Q_{pre} * (1 + d) \quad (8)$$

where *d* indicates the percent deviation threshold tolerated. The probability of type I and II errors of the threshold-adjusted MWW test indicate the likelihood that (i) the threshold will not be violated when the test suggests it will be and (ii) it will be violated when the tests suggests it will not be, respectively. With (8), we can test a hypothesis that the post-dam flows are greater than the pre-dam flows without specifying a given shift. Monotonic transformations can be applied to ensure $Q_{pre-scaled}$ and Q_{post} have similar variances without changing the order of ranked observations from the two samples.

[INSERT FIGURE 5]

To illustrate the modification of the MWW test for percent deviation thresholds, we assume the post-dam values for a hypothetical high-flow AFDC quantile follow the same distribution as its pre-dam counterpart, except that they have been reduced by 30%. If we evaluate the presumptive flow standard from Richter *et al*. (2012), which specifies that a decrease exceeding 20% could cause adverse ecological impacts, we first scale the pre-dam flows by 0.8, i.e. d = - 0.2. Then, we perform a MWW test in which the null hypothesis states that the scaled pre-dam flows are less than or equal to the post-dam flows, i.e. no violation of an alteration threshold, and the alternative hypothesis is that the scaled pre-dam high flows are greater than the post-dam high flows. In other words, we hypothesize a violation of the alteration threshold because the post-dam flows are still lower than the scaled pre-dam flows. The location of the mode of the post-dam flow distribution to the left of the pre-dam ones in Figure 5 indicates that the probability of a threshold violation not due to sampling uncertainty exceeds 50%. Again, monotonic transformations can be applied to ensure $Q_{pre-scaled}$ and Q_{post} have similar variances without changing the order of ranked observations from which the test draws conclusions.

2.5 Hypothesis tests of overall change in AFDCs

Next, we assess the likelihood that dam operations cause violations of alteration thresholds for *at least one* ecologically critical AFDC quantile. Multiple comparison procedures assess the overall, or field, significance associated with the repeated application of a hypothesis test applied to independent sub-samples of a phenomenon (e.g. Thompson *et al*., 2011; Reiff *et al*., 2016). Assuming the high (Q5) and low flows (Q95) are statistically independent, the following test determines the likelihood of a violation for at least one of these two ecologically critical AFDC quantiles as follows:

$$\mathrm{H_0}\text{: No threshold violations}$$

$$\mathrm{H_A}\text{: At least one threshold violation}$$

We then compute the probability of hypothesis testing errors for *K* independent AFDC quantiles:

$$\alpha_{\mathrm{overall}} = 1 - \prod_{\mathrm{k}=1}^{\mathrm{K}} (1 - \alpha_{\mathrm{k}}) \tag{9}$$

$$\beta_{\text{overall}} = 1 - \prod_{k=1}^{K} (1 - \beta_k) \tag{10}$$

Type II errors β_k are conditional upon the type I errors α_k selected for each AFDC quantile hypothesis test, and vice versa, since a value of one uniquely determines the other. Future work may consider the impact of cross-correlations among individual hypothesis test results when determining the overall field significance (Douglas *et al*., 2000).

3. Informing dam operation decisions with hypothesis test results

In this section, we introduce our Bayesian decision-tree framework, which accounts for both type I and II errors associated with hypothesis tests regarding the exceedance of percent deviation thresholds when evaluating potential hydropower and ecological consequence of reservoir operation decisions. We also contrast this approach with common deterministic and probabilistic decision-making methods.

3.1 Approaches for incorporating Type I and II errors in decisions

In our *post-hoc*, or observed, power analysis, we determine the power $1 - \beta$ corresponding to a given α, estimated effect size $\hat{\theta}$, and sample sizes m and n. We show the importance of carefully choosing among numerous methods for incorporating hypothesis test results into decisions. First, we apply null hypothesis significance testing (NHST), a common practice for making decisions in which null hypotheses of no change are rejected if the likelihood of a type I error falls below a critical probability set *a priori*, commonly 0.05 (e.g. Ioannidis, 2005). If the type I error probability exceeds this statistical threshold, one concludes insufficient evidence for rejecting the null hypothesis. This need for evidence of change may lead to situations in which flow alteration only slightly exceeding a threshold is deemed insignificant for declaring a violation. Such a low type I error acceptance rate implies that over-protection regrets, i.e. hydropower losses, are much more consequential than under-protection ones, an implied valuation which does not necessarily reflect the relative benefits of hydropower and ecosystems. Moreover, type II errors are not considered in this approach.

A second approach involves setting the type I (or type II) error to an arbitrary *a priori* value that reflects stakeholder risk tolerances and valuations (e.g., Mapstone, 1995; Field *et al*., 2004), and then determining the type II (or type I) error associated with it. Both errors are then used in a decision-tree framework. While some authors have called for a reversal of the "burden of proof" when the potential environmental damage is greater than the potential cost overruns stemming from protective actions (e.g. Field *et al.,* 2004), we employ the conventional formulation in which a type I error signals overprotection.

A third approach involves maximizing the overall tradeoff between two competing objectives by optimizing the values of type I and II error probabilities. This is especially suitable when a central decision-maker has a vested interest in both objectives or an external party aims to negotiate tradeoffs between stakeholders with competing interests. Examples of such a criterion include minimizing the overall expected cost (Field *et al*., 2004) and minimizing expected regret, i.e. the expected consequences of incorrect decisions (Rosner *et al*., 2014). The latter criterion is especially applicable when a decision-maker has a budget sufficient for changing the reservoir operating rules but wants to avoid unnecessary costs. In contrast, cost minimization may be more appropriate for minimizing the sum of hydropower and ecological "damage" costs (see Field *et al.* (2004) for an ecological conservation example). Here, we focus on minimizing the expected regrets associated with incorrect inferences.

3.2 Linking hypothesis testing errors to decision regrets

Statistical decision theory provides an avenue for incorporating type I and II errors into

Bayesian decision tree approaches for evaluating infrastructure design and operation decisions made when changes in environmental conditions are uncertain (e.g. Hobbs *et al.*, 1997). A Bayesian approach is necessary because type I and II error probabilities express the likelihood of a decision conditional upon an unknown true state of nature. However, for decision-makers knowing the probability of a consequence conditioned upon a decision, is imperative. For instance, the probability of a type I error expresses the likelihood of concluding a threshold violation if there is not actually one, i.e. P(CA|NA) = P(Conclude Alteration | No Alteration). This is different from the probability of not having a threshold violation if we conclude significant alteration and thus decide to change dam operating rules. i.e. P(NA|CA) = P(No Alteration | Conclude Alteration). The decision tree in Figure 6 further illustrates this concept. The square node indicates a dam operator decision and the circular nodes represent chance nodes, which reflect the likelihood of type I and II errors associated with the MWW test regarding the violation of alteration thresholds due to a previous dam-building decision. The ensuing branches identify the probability of making decisions leading to subsequent regrets associated with hydropower or ecological aspects of the project. Thus, Figure 6 integrates the hypothesis test outcomes with the consequences, or regret, associated with hydropower and ecological outcomes.

[INSERT FIGURE 6]

Figure 6 also shows that Bayes Theorem allows us to specify a prior probability of violating an alteration threshold and integrate it into a decision tree to obtain the final posterior probabilities. While prior probabilities based on stakeholder and expert beliefs can be incorporated using Bayes Theorem (Webb *et al.*, 2015), we demonstrate this method with an arbitrary non-informative prior probability in which there is a 50% chance of violating an alteration threshold. Bayes Theorem yields the probability of not violating an alteration threshold when deciding to change dam operating rules based on a conclusion of alteration $P(NA|CA)$:

$$P(NA|CA) = \frac{P(CA|NA)P(NA)}{P(CA)} \tag{11}$$

where the probability of concluding alteration is:

$$P(CA) = P(CA|NA)P(NA) + P(CA|A)P(A) \tag{12}$$

P(NA|CA) can be interpreted as the hydropower regret probability because it reflects the likelihood of unnecessarily changing dam operating dams based on an incorrect conclusion of alteration. Substituting (12) into (11), we obtain:

$$P(NA|CA) = \frac{P(CA|NA)P(NA)}{P(CA|NA)P(NA) + P(CA|A)P(A)} \tag{13}$$

Since we are assuming $P(NA) = P(A) = 0.5$, $P(CA|NA)$ and $P(CA|A)$ can be removed from (13) and then we can use α and β from Figure 2 to solve for $P(NA|CA)$:

$$P(NA|CA) = \frac{\alpha}{\alpha + (1-\beta)} \tag{14}$$

We also use Bayes Theorem to estimate probabilities for the other three possible combinations of flow alteration violation outcomes conditional upon conclusions from the MWW test.

With regret probabilities, we can compute the expected regrets (e.g. Rosner *et al.*, 2014) of dam operations decisions. First, we compute the expected hydropower regret ER_{HP} in terms of the difference in hydropower production between a "fraction-of-inflow" operating rule HP_{FOI}, which ensures that the outflow does not excessively deviate from the inflow (see Section 4), and a reference run-of-river operating rule HP_{ROR} requiring daily releases to equal daily inflows:

$$ER_{HP} = P(NA|CA) * (HP_{FOI} - HP_{ROR}) \tag{15}$$

Substituting in the expression derived from the hypothesis test result in Figure 7 leads to:

$$ER_{HP} = \frac{\alpha}{\alpha + (1 - \beta)} * (HP_{FOI} - HP_{ROR}) \tag{16}$$

Hydropower production can be quantified in terms of revenue, energy generation, reliability or other relevant performance indicators.

Next, we compute the expected ecological regret R_{ECO}. The probability that a decision will lead to an undesirable ecological state is given by *P(A|CNA)* in Figure 7. When dam operation changes a measurable ecological indicator, $P(A|CNA)$ serves as a weight for determining ER_{ECO} so that:

$$ER_{ECO} = \frac{\beta}{(1 - \alpha) + \beta} * (E_{ECO} - E_{HP}) \tag{17}$$

Ecological indicators may include measures of ecological health, such as species abundance and diversity, or monetary values of a fishery or ecosystem services. We assume flow alteration uniformly affects all species and that a single stakeholder represents all ecological interests, even though species and ecosystem functions (and the stakeholders representing them) often have competing hydrologic interests (e.g. Szemis *et al.*, 2012; Railsback *et al.*, 2015; Kozak *et al.*, 2015). When hydropower and ecological objectives are not commensurate, the impacts of dam operations on each objective can be measured relative to maximum possible values, e.g. hydropower under HP_{ROR} as a percentage of the hydropower production under HP_{FOI}. This way, we can identify the values of α and β that minimize the total expected regret ER_{TOT} for a given threshold set:

$$ER_{TOT} = \frac{\alpha}{\alpha + (1-\beta)} * (HP_{FOI} - HP_{ROR}) + \frac{\beta}{(1-\alpha)+\beta} * (E_{ROR} - E_{FOI}) \tag{18}$$

This approach differs from classic hydropower optimization models in which production is maximized given a set of constraints, including environmental flows (e.g. Cardwell *et al.*, 1996). Our approach is equivalent to selecting an optimal point on a receiving (relative) operating characteristic curve (ROC), a graphical technique for selecting thresholds for diagnostics based on binary classification systems (Swets, 1992). For an earth sciences example of ROC curves, see Figure 11 in Oommen *et al*. (2010). We apply this optimization routine to each individual AFDC quantile separately and then combine the optimal values of α and β to determine the overall probabilities of type I and II errors.

[INSERT FIGURE 8]

Figure 8 illustrates the relationship between the hypothesis test errors and regret probabilities when the prior probability of alteration is 0.5. These complex tradeoffs, which stem from having inadequate information concerning the likelihood of alteration, show that the likelihood of regret due to either hydropower (panel A), or ecological (panel B) or both (panel C) generally decreases as both type I and II error likelihoods decrease. Consider the case of fixing the type I error probability at 5%, a common assumption in NHST. The hydropower regret will always be very low, though the likelihood of ecological regret (panel B) or the total regret likelihood (panel C) will be much higher. Making a type II error has a much greater impact on the ecosystem than on hydropower when the type I error is low.

4. Stylized baseload hydropower dam example

4.1 Reservoir operations simulation model

We illustrate our decision-tree framework for comparing the hydropower-ecosystem tradeoffs resulting from different reservoir operating rules using a stylized hydropower reservoir example to avoid accounting for systematic differences between pre- and post-dam periods other than dam operations (see the Supporting Information for details). We use a 37-year inflow series (1913-1949) from the USGS station (02080500) on the Roanoke River at Roanoke Rapids, North Carolina for daily inflows. The reservoir can store 22% of the mean annual inflow during this period. Water is released downstream from the reservoir via: (i) turbine outflows, (ii) an environmental flow bypass and (iii) spills during high-flow periods (Figure 9). Turbines are situated in an integral powerhouse built into the dam, and release water into the main channel below the dam. To avoid accounting for the operation of individual turbines, we assume they can release between 20% and 100% of the mean annual discharge (239 m^3/s). Turbine releases are permitted whenever the sum of available storage in the conservation pool and the daily inflow exceeds 20% of the mean annual discharge. When low inflow and storage prevent power generation, a low-flow outlet releases 28.6 m^3/s, the annual seven-day minimum discharge with a ten-year recurrence interval (7Q10), provided that sufficient storage remains. When the conservation storage pool is full, inflow passes downstream via a spillway with an infinite discharge capacity without any gates for controlling releases. Our hypothetical dam has an installed energy generating capacity of 49.4 MW, which can power as many as 49,400 homes in an industrialized region (Electricity Power Supply Association, 2017). However, its annual average generation is expected to be substantially lower since energy is not always produced at the maximum rate due to water shortages, environmental flow constraints, and other operational issues.

[INSERT FIGURE 9]

4.2 Operating rules and flow alteration thresholds analyzed

We compare a fraction-of-inflow (FOI) operating rule, which requires daily turbine releases to be between 40% and 180% of the inflow on the same day, with a run-of-river (RR) scheme that does not alter daily flows. Although it is unlikely that a large storage reservoir would be converted into a run-of-river facility, this comparison provides a valuable reference point for assessing the hydropower and ecosystem impacts of flow alteration. The percent deviations in daily inflows permitted are based on percent deviation thresholds for fish that Carlisle *et al*. (2011) detected in annual maximum and seven-day low flows collected at 237 stations in the contiguous United States. We first evaluate flow alteration for a set of percent deviation thresholds (Threshold Set 1) based on Carlisle *et al.* (2011) in which the annual Q5 flows cannot decrease by more than 60% and the annual Q95 AFDC flows cannot increase by more than 80%. While we recognize that changes to ecosystems from low and high flows are not equivalent to ones stemming from alterations to extreme high and low flows, we use them to illustrate the incorporation of empirical thresholds into our decision-making framework. Then, we replace these thresholds with a set in which the annual Q5 values cannot decrease by more than 30% and the annual Q95 values cannot increase by more than 50% (Threshold Set 2).

5. Results

5.1 Threshold Set 1

The average annual hydropower production declines by just 13% when the reservoir switches from HP_{FOI} (330 GWh, 3.30 x 10^6 kWh) to HP_{ROR} (286 GWh, 2.86 x 10^6 kWh). The very low interannual flow variability (annual Cv = 0.22) at this site explains this mild reduction (see Vogel *et al*. (1998) for the Cv of annual flows in the U.S.). HP_{FOI} reduces the average Q5 by 37% and elevates the average Q95 by 66%, changes that are less severe than the 60% decrease in high flows and 80% increase in low flows that Threshold Set 1 tolerates. Both the large turbine discharge capacity and spills prevent the annual Q5 from decreasing below 40% of its pre-dam average. While Threshold Set 1 permits Q95 values to increase by 80%, reservoir storage is often insufficient for releases equal to 180% of the inflow to be made during these low-flow periods. These results clearly illustrate that changes in high and low flows under HP_{FOI} do not exceed the deviations stipulated in Threshold Set 1. These results also highlight the extent to which turbine constraints affect the flow alteration impacts of reservoirs, and demonstrate that they supplement the storage and generation capacity metrics commonly used to appraise the ecological performance of hydropower dams (e.g. Kibler and Tullos, 2014).

5.2 Threshold Set 2

Next, we illustrate a case with stricter thresholds, a 30% high-flow decrease and a 50% low-flow increase. Even though these percent differences between HP_{ROR} (-37%) and HP_{FOI} (66%) are both greater than permitted in Threshold Set 2, we must rule out the possibility that these violations arise from random sampling variability. When stakeholders agree to tolerate a 20% type I error probability for both high- and low-flow alteration, i.e. $\alpha = 0.2$, the type II error probabilities for violating high- and low-flow alteration thresholds are 0.38 and 0.26, respectively. Next, we compute the probability of at least one violation of an alteration threshold at either the Q5 or Q95 AFDC quantile. The probability of at least one type I error is 0.36, and the probability of at least one Type II error is 0.54. Using (18), this translates to overall hydropower and ecological regret probabilities of 0.44 and 0.46, respectively. Even though the average changes in the two AFDC quantiles exceed these posited thresholds, a total regret probability close to one indicates insufficient evidence of a threshold violation.

Next, we compute the expected hydropower and ecological regrets with the hypothesis test error probabilities. Since the difference in hydropower production between HP_{FOI} and HP_{ROR} is 44 GWh per year, the expected hydropower regret is 0.44 * 44 GWh = 19 GWh, less than 6% of the average annual hydropower output under HP_{FOI}. In contrast, the likelihood of inducing adverse ecological impacts by not modifying dam operations is 46%. (Recall that we interpret ecological regret probability as an indicator of the likelihood of adverse ecological impacts in this hypothetical example.) Figure 10 shows that, if we assume a one-percent decrease in hydropower is of equal value to a one-percent increase in the likelihood of adverse ecological impacts, under-protection errors become much more consequential. In other words, our *a priori* type I error probability tolerance of 0.2 is too strict since HP_{ROR} only reduces hydropower production mildly (13%). The lighter gray points in Figure 10 indicate values for which $\alpha = 0.2$, whereas the darker "optimal" points denote the combination of type I and II error probabilities that minimize the percent regret for each objective. Each point in the expected percent regret plots corresponds to a point of the same color in the hypothesis test error tradeoff plots. The tradeoff curve optima indicate that we should accept very high probabilities of type I errors to minimize the total decision regret, as the ecological consequence of an incorrect decision to maintain hydropower production is much greater than the hydropower consequence of changing to a run-of-river rule. This example clearly illustrates the problem with arbitrarily setting type I error probability requirements.

Table 1 below shows the importance of carefully selecting among the different hypothesis testing approaches presented in Section 3.1 as well as the overall importance of hypothesis testing. Most notably, NHST applied with a critical type I error probability of 0.05 suggests that these changes to high (p = 0.09) and low flows (p = 0.10) are not significant enough to warrant any action. (The overall probability of at least type I error of 0.18 is even higher). In contrast, the two approaches considering both type I and II errors indicate a need to protect the river, although the degree to which they each suggest protection varies considerably.

[INSERT TABLE 1]

Next, we examine the difference between expected regrets when different thresholds are tested. Figure 11 shows the range of optimal expected regrets for percent decrease thresholds of 0-30% for high flows and for percent increase thresholds of 0-60% for low flows. The actual percent changes in high (-37%) and low flows (66%) exceed all these thresholds. Figure 11a shows that optimal solutions for the other thresholds examined also have expected ecological regrets much lower than their corresponding expected hydropower regrets because the potential ecological consequences are much greater on a percent-loss basis. In contrast, Figure 11b shows the optimal regrets for a scheme in which a 1% probability of a hydropower regret is weighted the same as a 1% probability of an ecological regret, i.e., a 13% hydropower loss is equal to a 100% chance of incurring adverse ecological impacts. In other words, hydropower production is assumed to be worth nearly eight times as much as the ecosystem. Under this valuation, one can see that some high- and low-flow percent deviation thresholds have optimal points with a greater expected ecological regret and, hence, a larger type II error probability. As expected, Figure 11 also shows that the regret probabilities rise as the percent difference between pre- and post-dam AFDC quantiles approach the stakeholder-identified percent deviation thresholds.

[INSERT FIGURE 11]

Since our one-tailed hypothesis test only examines the likelihood of thresholds being exceeded, we exclude cases in which the sum of the two regret probabilities exceeds one. For this reason, the contours in Figure 11 may exceed 100%. When thresholds are not exceeded and hypotheses of threshold exceedance are tested, this test yields values of α and β where $\alpha + \beta > 1$. A value greater than one indicates that the test has been applied in the wrong direction, e.g. the alternative hypothesis is that scaled pre-dam flows are greater than post-dam flows when they are, in fact, less than them. Yet, tests for which $\alpha + \beta > 1$ indicate that the violation of an alteration threshold is very unlikely. Instead, they demonstrate a higher likelihood of hydropower regret, which has equally important implications for dam operation decisions.

6. Discussion

This paper makes an initial contribution toward the incorporation of the uncertainty associated with over- and under-design regrets associated with decisions regarding reservoir release rules which impact hydropower-ecosystem tradeoffs. Most importantly, we highlight the decision implications of different probabilistic approaches for incorporating the uncertainty of long-term hydrologic alteration in evaluations of tradeoffs between hydropower and ecosystem benefits of different operating rules for a large baseload hydropower reservoir. Our approach focuses on minimizing the regrets associated with reservoir release decisions made in the face of hydrologic uncertainty as opposed to the total cost-minimization approach introduced by Field *et al.* (2004). One of the key differences between the regret- and cost-minimization approaches is that cost-minimization also considers hydropower costs when changing to $HP_{ROR}$is the correct decision. This implies that a cost-minimization approach would be more likely to recommend operating rules conducive to hydropower production. Yet, further research is needed to compare the implications of these two approaches for evaluating hydropower-ecosystem tradeoffs and other environmental management problems, including the introduction of constraints on the distribution of costs among stakeholders.

Other nonparametric tests offer possibilities for extending our analysis. While we analyze changes in intra-annual variability through our assessments of changes to typical values of high and low flows, the nonparametric Siegel-Tukey test (Siegel and Tukey, 1960; see FitzHugh (2014) for an environmental flows application) could examine changes in interannual variability using AFDCs. In fact, interannual flow variability, especially extremely wet and dry years, controls the composition of riverine ecosystems in many settings (e.g. Nislow *et al.*, 2002; Rivaes, 2015). Other more specialized nonparametric approaches could detect simultaneous changes in the central tendency and variability of environmental flow indicators (see Marozzi, 2013). In addition, nonparametric tests that evaluate changes in distributions, such as the Kuiper test that Kroll *et al.* (2015) apply to changes in AFDCs, could be modified to determine whether changes in AFDC quantiles have exceeded ecologically critical thresholds d.

Our test could easily be adapted to examine changes in flows on ecologically relevant dates, though serial correlations among flows on different dates may need to be addressed. While AFDCs describe the entire range of daily flows in each water year, many ecological functions depend on the timing and sequence of flows (Stewart-Koster *et al.*, 2015), factors not considered within the context of AFDCs. Other studies of highly gauged rivers have recommended ecological flow targets based on percent deviations from sequential daily flow hydrographs (e.g. Steinschneider *et al.*, 2014; McKay, 2015). Our framework can also assess changes to many other flow statistics, such as flashiness indices (e.g. Baker *et al*., 2004) describing the rate of change of sub-daily peaking flows (Haas *et al*., 2014), and even relevant non-flow indicators, such as habitat suitability indices (Bovee and Cochnauer, 1977).

One might question the purpose of testing for hydrologic alteration many years after a dam is built if the flows have already been altered to an extent that might adversely affect species with lifespans of no more than a few years. However, excessive hydrologic alteration may signal the ongoing or potential decline of an ecosystem since the post-dam ecological equilibrium may take some time to become established. Perkin *et al.* (2016) observed fewer native opportunistic species and more non-native generalist species downstream of a reservoir approximately a decade after its impoundment compared to the first few post-dam years. Taylor *et al. (*2014) detected fewer changes in pre-dam fish assemblages during a six-year post-impoundment period than they did during the ensuing seven years. While more research on the rates of these ecological transitions is needed, these two studies illuminate the value of performing hypothesis tests on post-dam flow records of approximately one decade. Moreover, even if dam operations adversely affect a riverine ecosystem within a few years, our hypothesis testing framework could determine whether dam operations must be changed to provide flow conditions enabling ecological restoration, or if other causes should be addressed instead. In fact, this hypothesis testing framework could be inverted to examine the achievement of pre-dam flow conditions, an increasing challenge given the proliferation of dam removal projects (O'Connor *et al*., 2015). Also, in data-poor regions, these hypothesis tests can be used to screen sites where more intensive ecological reconnaissance may be necessary. Ideally, such screening-level studies would also determine when flow alteration is the limiting factor constraining riverine ecosystems (e.g. McManamay *et al*., 2013; Knight *et al*., 2014).

Finally, our testing procedure motivates efforts to estimate the economic value of environmental flows. In some cases, the value of fisheries provides a more objective measure of the benefits of environmental flows (e.g., Kozak *et al*., 2015) while other benefits, such as the aesthetic and intrinsic values of riverine ecosystems, are more challenging to monetize. Most importantly, our framework can provide useful results regarding tradeoffs between hydropower and ecosystem management objectives even if they cannot be assessed commensurately. Decision trees can also produce tradeoff curves that decision-makers can consult subsequently, an increasingly advocated form of environmental planning decision support (e.g. Quinn *et al.*, 2017).

7. Conclusions

There is a growing interest in protecting riverine ecosystems downstream of dams using percent-deviation thresholds of hydrologic alteration (e.g. Poff *et al.*, 1997; Vogel *et al.*, 2007; Richter *et al.*, 2012; McKay, 2015) since they require less field reconnaissance than flow magnitude-based river and reservoir management policies. When evaluating changes between pre- and post-dam periods, one wishes to consider the possibility that the exceedance of thresholds is partly due to sampling variability alone, which stems from o natural streamflow variability and the limited samples available contributes to the exceedance of thresholds. Our decision-theoretic approach removes the effects of sampling variability from reservoir operation decisions made for the purposes of maintaining hydropower production and/or conserving ecosystem health thus enabling us to focus exclusively on uncertain hydropower and ecosystem outcomes resulting from reservoir release decisions. We modify and extend a nonparametric hypothesis test to examine whether differences between pre- and post-dam flow of an annual FDC (AFDC) quantile exceed allowable percent deviation thresholds. While we apply this approach to AFDCs, we have also created a general framework for incorporating the uncertainty of environmental threshold exceedances into tradeoffs between off-stream reservoir benefits and in-stream water uses. Our stylized example highlights differences between our decision-theoretic approach and conventional decision-making methods, which, in turn, motivates future efforts to consider the uncertainty of hydrologic alteration assessments carefully when addressing river basin conflicts.

Acknowledgments

We are indebted to the Hydro Research Foundation for providing a research award through a grant from the United States Department of Energy (DOE). The National Science Foundation's Integrative Graduate Education and Research Traineeship (IGERT) program in water diplomacy at Tufts University also provided in-kind support. Ryan McManamay was supported by the Water Power Technologies Office with the US DOE. This manuscript has been authored by UT-Battelle, LLC under Contract No. DE-AC05-00OR22725 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan).

References

Arias, M.E., Cochrane, T.A., Kummu, M., Lauri, H., Holtgrieve, G., Koponen, J., Piman, T. 2014. Impacts of hydropower and climate change on drivers of ecological productivity of Southeast Asia's most important wetland. *Ecol. Model.*, 272, 252-263.

Baker, D.B., Richards, P., Loftus, T.T., Kramer, J.W. 2004. A new flashiness index: characteristics and applications to Midwestern rivers and streams. *J. Am. Water Assoc.*, 40(2), 503-522.

Bellera, C.A., Julien, M., Hanley, J.A. 2010. Normal approximation to the distributions of the Wilcoxon statistics: Accurate to what N? Graphical insights. *J. Stat. Educ*., 18(2), 1-17.

Berger, J.O. 1993. Statistical decision theory and Bayesian analysis. New York: Springer-Verlag, Inc. 2nd Edition.

Blair, R.C., Higgins, J.J. 1980. A comparison of the power of Wilcoxon's rank-sum statistics to that of Student's t-statistic under various nonnormal distribution. *J. Educ. Behav. Stat.*, 5(4), 309-335.

Botter, G., Basso, S., Porporato, A., Rodriguez-Iturbe, I., Rinaldo, A. 2010. Natural streamflow regime alterations: Damming of the Piave river basin (Italy). *Wat. Resour. Res.*, 46, W06522, doi:10.1029/2009WR008523.

Bovee, K.D., Cochnauer, T. 1977. Development and evaluation of weighted criteria probability-of-use curves for Instream flow assessments: fisheries. Instream Flow Information Paper 3. United States Fish and Wildlife Service FWS/ORS-77/63.

Carlisle, D.M., Wolock, D.M., Meador, M.R. 2011. Alteration of streamflow magnitudes and potential ecological consequences: A multiregional assessment. *Front. Ecol. Environ*., doi:10.1890/100053.

Cardwell, H., Jager, H.I., Sale, M.J. 1996. Designing instream flows to satisfy fish and human water needs. *J. Wat. Resour. Plann. Manage.*, 122(5), 356-363.

Castellarin, A., Botter, G., Hughes, D.A., Liu, S., Ouarda, T.B.M.J., Parajka, J., Post, D.A., Sivapalan, M., Spence, C., Viglione, A., Vogel, R.M. 2013. Prediction of flow duration curves in ungauged basins, Chapter 7 in Prediction in Ungaged Basins, Cambridge University Press, 496 pages.

Cioffi, F., Gallerano, F. 2012. Multi-objective analysis of dam release flows in rivers downstream from hydropower reservoirs. *Appl. Math. Modell.*, 36, 2868-2889.

Cohn, T.A., Lins, H.F. 2005. Nature's style: Naturally trendy. *Geophys. Res. Lett.*, 32, L32402, doi:10.1029/2005GL024476.

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

Electricity Power Supply Association (EPSA). 2017. Electricity Primer: The Basics of Power and Competitive Markets. Available here: https://www.epsa.org/industry/primer/?fa=wholesaleMarket

Eriyagama, N., Smakhtin, V., Jinapala, K. 2016. The Sri Lanka environmental flow calculator: a science-based tool to support sustainable national water management. *Water Policy*, 18(2), 480-492.

Field, S.A., Tyre, A.J., Jonzén, N., Rhodes, J.R., Possingham, H.P. 2004. Minimizing the cost of environmental management decisions by optimizing statistical thresholds. *Ecol. Lett.*, 7, 669-675.

FitzHugh, T.W. 2014. EFCAM: A method for assessing alteration of environmental flow components. *River Res. Applic.*, 30, 825-844.

Gao, Y., Vogel, R.M., Kroll, C.N., Poff, N.L., Olden, J.D. 2009. Development of representative indicators of hydrologic alteration. *J. Hydrol.*, 374, 136-147.

Haas, N.A., O'Connor, B.L., Hayse, J.W., Bevelheimer, M.S., Endreny, T.A. 2013. Analysis of daily peaking and run-of-river operations with flow variability metrics, considering subdaily to seasonal time scales. *J. Amer. Water. Resour. Assoc.*, 50(6), 1622-1640.

Halleraker, J.H., Sundt, H., Alfredsen, K.T., Dangelmaier, G. 2007. Application of multiscale environmental flow methodologies as tools for optimized management of a Norwegian regulated national salmon watercourse. *Riv. Res. Appl.*, 23, 493-510.

Hobbs, B.F., Chao, P.T., Venkatesh, B.N. 1997. Using decision analysis to include climate change in water resources decision making. *Clim. Change*, 37, 177-202.

Ioannidis, J.P.A. 2005. Why most published research findings are false. *PLOS Med.*, 2(8), 124-129.

Jager, H.I. 2014. Thinking outside the channel: Timing pulse flows to benefit salmon via indirect pathways. *Ecol. Modell.*, 273, 117-127.

Jager, H.I., Smith, B.T. 2008. Sustainable reservoir operation: can we generate hydropower and preserve ecosystem values? *River Res. Applic.*, 24, 340-352.

Jain, S.K. 2015. Assessment of environmental flow requirements for hydropower projects in India. *Current Science*, 108(10), 1815-1825.

Kasuya, E. 2001. Mann-Whitney *U* test when variances are unequal. *Anim. Behav.*, 61, 1247-1249.

Kendy, E., Apse, C., Blann, K., Smith, M.P, Richardson, A. 2012. Practical guide to environmental flows for policy and planning. Available at https://www.conservationgateway.org.

Kennard, M.J., Mackay, S.J., Pusey, B.J., Olden, J.D., Marsh, N. 2010. Quantifying uncertainty in estimation of hydrologic metrics for ecohydrological studies. *River Res. Applic.*, 26,137-156.

Knight, R.R., Murphy, J.C., Wolfe, W.J., Saylor, C.F., Wales, A.K. 2014. Ecological limit functions relating fish community response to hydrologic departures of the ecological flow regime in the Tennessee River basin, USA. *Ecohydrology*, 7, 1262-1280.

Kopf, R.K., Finlayson, C.M., Humphries, P., Sims, N.C., Hladyz, S. 2015. Anthropocene baselines: assessing change and managing biodiversity in human-dominated aquatic ecosystems. *BioScience*, doi: 10.1093/biosci/biv092.

Kozak, J.P., Bennett, M.G., Hayden-Lesmeister, A., Fritz, K.A., Nickolotsky, A. 2015. Using flow-ecology relationships to evaluate ecosystem service trade-offs and complementarities in the nation's largest river swamp. *Environ. Manage.*, doi:10.1007/s00267-015-0474-4

Kroll, C.N., Croteau, K.E., Vogel, R.M. 2015. Hypothesis tests for hydrologic alteration. *J. Hydrol.* 530, 117-126.

Kuiper, N.H. 1960. Tests concerning random points on a circle. *Pro. Koninklijke Nederlandse Akademie van Wetenschappen, Series A*, 63, 38-47.

Lamouroux, N., Gore, J.A., Lepori, F., Statzner, B. 2015. The ecological restoration of large rivers needs science-based, predictive tools meeting public expectations: an overview of the Rhone project. *Freshwater Biol.*, doi:10.1111/fwb.12563.

Lauri, H., de Moel, H., Ward, P.J., Rasanen, T.A., Keskinen, M., Kummu, M. 2012. Future changes in Mekong River hydrology: impact of climate change and reservoir operation on discharge. *Hydrol. Earth Syst. Sci.*, 16, 4603-4619.

Lee, W.-C. 2014. Joint distribution of rank statistics considering the location and scale parameters and its power study. *J. Stat Dist. Applic.*, 1-6.

Mapstone, B.D. 1995. Scalable decision rules for environmental impact studies: effect size, type I, and type II errors. *Ecol. Applic.*, 5(2), 401-410.

Marozzi, M. 2013. Nonparametric simultaneous tests for location and scale testing: a comparison of several methods. *Comm. Stat. Sim. Comp.*, 42(6), 1298-1317.

McKay, K.S. 2015. Quantifying tradeoffs associated with hydrologic environmental flow methods. *J. Am. Wat. Resour. Assoc.*, 51(6), 1508-1518.

McManamay, R.A. 2015. Isolating causal pathways between flow and fish in the regulated river hierarchy. *Can J. Fish Aquat. Sci.,* 72, 1-18.

McManamay, R.A., Orth, D.J., Dolloff, C.A., Mathews, D.C. 2013. Application of the ELOHA framework to regulated rivers in the Upper Tennessee River basin: a case study. *Environ. Manage.*, doi:10.1007/s00267-013-0055-3.

Mims, M.C., Olden, J.D. 2013. Fish assemblages respond to altered flow regimes via ecological filtering of life history strategies. Freshwater Biology, 58, 50-62.

Neave, H.R., Granger, C.W.J. 1968. A Monte Carlo study comparing various two-sample tests for differences in mean. *Technometrics,* 10(3), 509-522.

Nikghalb, S., Shokoohi, A., Singh. V.P., Yu, R. 2016. Ecological regime versus minimum environmental flow: comparison of results for a river in a semi-Mediterranean region. *Water. Resour. Manage.*, doi:10.1007/s11269-016-1488-2.

Nislow, S., Magilligan, F.J., Fassnacht, H., Bechtel, D., Ruesink, A. 2002. Effects of dam impoundment on the flood regime of natural floodplain communities in the upper Connecticut River. *J. Am. Water Resour. Assoc.*, 38(6), 1533-1548.

O'Connor, J.E., Duda, J.J., Grant, G.E. 2015. 1000 dams down and counting. *Science*, 348(6234), 496-497.

O'Hanley, J.R., Tomberlin, D. 2005. Optimizing the removal of small fish passage barriers. *Environ. Model. Assess.*, 10(2), 85-98.

Oommen, T., Baise, L.G., Vogel, R.M. 2011. Sampling bias and class imbalance in maximum likelihood logistic regression. *Math. Geosci.,* 43 (1), 99-120.

Pearsall, S.H., McCrodden, B.J., Townsend, P.A. 2005. Adaptive management of flows in the Lower Roanoke River, North Carolina, USA. *Environ. Manage.*, 35(4), 353-367.

Peñas, F.J., Barquín, J., Álvarez, C. 2016. Assessing hydrologic alteration: Evaluation of different alternatives according to data availability. *Ecol. Ind.*, 60, 470-482.

Perkin, J.S., Knorp, N.E., Boersig, T.C., Gebhard, A.E., Hix, L.A., Johnson, T.C. 2016. Life history theory predicts long-term fish assemblage response to stream impoundment. *Can J. Fish Aquat. Sci.*, 3

Poff, N.L., Zimmerman, J.K.H. 2010. Ecological responses to altered flow regimes: a literature review to inform the science and management of environmental flows. *Freshwater Biol.,* 55, 194-205.

Poff, N.L. and many others. 2010. The ecological limits of hydrologic alteration (ELOHA): a new framework for developing regional environmental flow standards. *Freshwater Biol.*, 55, 147-170.

Poff, N.L. 2009. Managing for variability to sustain freshwater ecosystems. *J. Water Res. Plann. Manage.*, 135, 1-4.

Poff, N.L., Olden, J.D., Merritt, D.M., Pepin, D.M. 2007. Homogenization of regional river dynamics by dams and global biodiversity implications. *Proc. Nat. Acad. Sci.*, 104(14), 5722-5727.

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

Power, M.E., Sun, A., Parker, G., Dietrich, W.E., Wootton, J.T. 1995. Hydraulic food-chain models. *BioScience*, 45, 159-167.

Quinn, J.D., Reed, P.M., Keller, K. 2017. Direct policy search for robust multi-objective management of deeply uncertain socio-ecological tipping points. *Environ. Model. Softw.*, 92, 125-141.

Railsback, S.F., Harvey, B.C., Kupferberg, S.J., Lang, M.M., McBain, S., Welsh Jr., H.H. 2015. Modeling potential river management conflicts between frogs and salmonids. *Can. J. Fish Aquat. Sci.*

Reiff, A.J., Sanayei, M., Vogel, R.M. 2016. Statistical bridge damage detection using girder distribution factors. *Engineering Structures*, 109, 139-151.

Renöfalt, B.M., Jansson, R., Nilsson, C. 2010. Effects of hydropower generation and opportunities for environmental flow management in Swedish riverine ecosystems. *Freshwater Biol.* 55, 49-67.

Rheinheimer, D.E., Liu, P., Guo, S. 2016. Re-operating the Three Gorges Reservoir for environmental flows: a preliminary assessment of trade-offs. *River Res. Applic.*, doi:10.1002/rra.2866.

Richter, B.D. 2010. Re-thinking environmental flows: from allocations and reserves to sustainable boundaries. *River Res. Applic.*, 26, 1052-1063.

Richter, B.D., Baumgartner, J.V., Powell, J., Braun, D.P. 1996. A method for assessing hydrologic alteration within ecosystems. *Conserv. Biol.*, 10(4), 1163-1174.

Richter, B.D., Davis, M.M., Apse, C., Konrad, C. 2012. A presumptive standard for environmental flow protection. *River Res. Applic.*, 28, 1312-1321.

Richter, B.D., Thomas, G.A. 2007. Restoring environmental flows by modifying dam operations. *Ecol & Soc.*, 12(1), 12.

Rivaes, R., Boavida, I., Santos, J.M., Pinheiro, A.N., Ferreira, M.T. 2015. The inbuilt long-term unfeasibility of environmental flows when disregarding riparian vegetation requirements. *Hydrol. Earth Syst. Sci. Discuss.*, 12, 10701-10737.

Rosner, A., Vogel, R.M., Kirshen, P.H. 2014. A risk-based approach to flood management decisions in a nonstationary world. *Water Resour. Res.*, 50, 10.1002/2013WR014561.

Siegel, S., Tukey, J.W. 1960. A nonparametric sum of ranks procedure for relative spread in unpaired samples. *J. Am. Stat. Assoc.*, 55(291), 429-445.

Smakhtin, V. Revenga, C., Döll, P. 2004. A pilot global assessment of environmental water requirements and scarcity. *Water International*, 29(3), 307-317.

Steinschneider, S., Bernstein, A., Palmer, R., Polebitski, A. 2014. Reservoir management optimization for basin-wide ecological restoration in the Connecticut River. *J. Wat. Resour. Plann. Manage.*, doi:10.10161/(ASCE)WR.1943-5452.0000399.

Stewart-Koster, B., Olden, J.D., Gido, K.B. 2014. Quantifying flow-ecology relationships with functional linear models. *Hydrol. Sci. J.*, 59(3-4), 1-16.

Suen, J.-P., Eheart, J.W. 2006. Reservoir management to balance ecosystem and human needs: incorporating the paradigm of the ecological flow regime. *Water Resour. Res.*, 42, W03417, doi:10.1029/2005WR004314.

Swets, J. 1992. The science of choosing the right decision threshold in high-stakes diagnostics. *Am. Psych.*, 47(4), 522-532.

Szemis, J. M., Maier, H.R., Dandy, G.C. 2012. A framework for using ant colony optimization to schedule environmental flow management alternatives for rivers, wetlands, and floodplains. *Wat. Resour. Res.*, 48, W08502, doi:10.1029/2011WR011276.

Taylor, J.M., Seilheimer, T.S., Fisher, W.L. 2014. Downstream fish assemblage response to river impoundment varies with degree of hydrologic alteration. *Hydrobiologia*, 728, 23-39.

Tennant, D.L. 1976. Instream flow regimens for fish, wildlife, recreation, and related environmental resources. *Fisheries*, 1(4), 6-10.

Tilmant, A., Beevers, L., Muyunda, B. 2010. Restoring a flow regime through the coordinated operation of a multireservoir system: The case of the Zambezi River basin. *Water Resour. Res.*, 46, W07533, doi:10.1029/2009WR008897.

USACE (United States Army Corps of Engineers). 1992. Water control plan for John H. Kerr Dam and Reservoir. Wilmington, North Carolina, 24 pp.

Vogel, R.M., Fennessey, N.M. 1995. Flow duration curves II: A review of applications in water resources planning. *Water Resources Bulletin*, 31(6), 1029-1039.

Vogel, R.M., Fennessey, N.M. 1994. Flow-duration curves I: New interpretation and confidence intervals. *J. Wat. Resour. Plann. Manage.,* 120(4), 485-504.

Vogel, R.M., Rosner, A., Kirshen, P.H. 2013. Likelihood of Societal Preparedness for Global Change – Trend Detection, *Nat. Hazards and Earth Sys.,* Brief Communication, 13, 1-6.

Vogel, R.M., Sieber, J., Archfield, S.A., Smith, M.P., Apse, C.D., Huber-Lee, A. 2007. Relations among storage, yield and instream flow. *Water Resour. Res.*, 43, doi:10.1029/2006WR005226.

Vogel, R.M., Tsai, Y., Limbrunner, J.F. 1998. The regional persistence and variability of annual streamflow in the United States. *Water Resour. Res.*, 34(12), 3445-3459.

Wald, A. 1939. Contributions to the Theory of Statistical Estimation and Testing Hypotheses. *Ann. Math. Stat.*, 10(4), 299-326.

Walker, K.F., Sheldon, F., Puckridge, J.T. 1995. A perspective on dryland river ecosystems. *Regulated Rivers: Research and Management*, 11, 85-104.

Webb, J.A., Miller, K.A., King, E.L., de Little, S.C., Stewardson, M.J., Zimmerman, J.K.H., Poff, N.L. 2013. Squeezing the most out of existing literature: a systematic re-analysis of published evidence on ecological responses to altered flows. *Freshwater Biol.*, 58, 2439-2451.

Webb, J.A. and many others. 2015. A general approach to predicting ecological responses to environmental flows: making best use of the literature, expert knowledge, and monitoring data. *River. Res. Applic.*, doi:10.1002/rra.2832.

Williams, J.G. 2017. Building hydrologic foundations for applications of ELOHA: how long a record should you have? *River Res. Applic.*, doi:10.1002/rra.3143.

Winemiller, K.O. and many others. 2016. Balancing hydropower and biodiversity in the Amazon, Congo, Mekong. *Science*, 321(6269), 128-129.

Yang, Y.C.E., Cai, X. 2011. Reservoir re-operation for fish ecosystem restoration using daily inflows. *J. Wat. Resour. Plann. Manage.*, 137(6), 470-480.

Yue, S., Wang, C.Y. 2002. Power of the Mann-Whitney test for detecting a shift in median or mean of hydro-meteorological data. *Stoch. Env. Res. Risk A.*, 16, 307-323.

Zarfl, C., Lumsdon, A., Berlekamp, J., Tydecks, L., Tockner, K. 2015. A global boom in hydropower dam construction. *Aquat. Sci.*, 77, 161-170.

Supporting Information: Simulation model for stylized reservoir example

We illustrate our decision-tree framework for comparing the hydropower-ecosystem tradeoffs resulting from different reservoir operating rules using a stylized example inspired by the John H. Kerr Reservoir situated on the Roanoke River on the border of the U.S. states of North Carolina and Virginia. Streamflow information is available from the United States Geological Survey (USGS) station (02080500) on the Roanoke River at Roanoke Rapids, North Carolina downstream from the dam. A 37-year pre-dam record is available from the 1913-1949 water years (Oct 1 – Sep 30), as construction began altering the flow in 1950 (Richter *et al.*, 1996). Dam-induced changes in streamflow at this station have also been profiled in studies evaluating the observed impact of the dam (e.g. Richter *et al.*, 1996). While there are numerous ecological concerns downstream of the dam, the preservation of seasonally flooded hardwood forests has formed the crux of many conservation efforts (Richter *et al.*, 1996/7). The potential impacts of revised operating rules proposed during a stakeholder-driven Federal Energy Regulatory Commission dam relicensing process required for privately operated dams in the United States (e.g. Pearsall *et al.*, 2005) have also been investigated. To simulate the impact of the reservoir on post-dam flows, we treat this streamflow record as reservoir inflow. This approach enables us to avoid accounting for other possible systematic differences between pre- and post-dam periods. However, we apply the hypothesis test as if we did not know that we had done this.

We make numerous assumptions and simplifications to the actual reservoir system to elucidate key features of our decision-making framework. The reservoir stores 1.27 x 10^9 m^3 (1,027,000 acre-feet) of water between the top of the dead storage zone at 268 ft (81.7 m) above sea level and the top of the conservation storage pool at 300 ft (91.4 m) above sea level (USACE, 1992). The dead storage pool, from which hydropower releases cannot be made, stores an additional 5.69 x 10^8 m^3 (461,600 acre-feet). The conservation storage, which can be released for hydropower, is equal to approximately 20 percent of the mean annual inflow. In the conservation storage pool, the reservoir surface area rises from 19,700 acres (79.7 km^2) at the top of the dead storage pool to 48,900 acres (197.9 km^2) at the top of the conservation storage pool. To estimate the daily reservoir water level, we relate changes in storage with changes in depth using a power-law relationship, which considers the storage the conservation pool provides as well as the reservoir surface area at the bottom and top of the conservation storage zone). While we include many design parameters from the actual reservoir reported in USACE (1992), the operations we simulate differ from the dam's actual operations, which are driven by flood control and diurnal energy price variability (Pearsall *et al.*, 2005). We also do not consider evaporation from the reservoir and assume that its storage capacity does not decrease over time due to sedimentation.

Figure 8 in the main text displays a schematic diagram of releases through three mechanisms: (i) turbine outflows, (ii) an environmental flow bypass and (iii) spills during high-flow periods. We express the minimum and maximum discharge capacity of the turbines as a percentage of the mean daily discharge to avoid accounting for individual turbine operation decisions. The John H. Kerr Dam powerhouse, which operates as a peaking facility, has nine turbines, whose collective discharge capacity is nearly four times the mean annual inflow at this site. However, such a large discharge capacity is unrealistic for a baseload hydropower plant. Instead, we assume that the turbines, which are situated in an integral powerhouse built into the dam, can release water at a rate up to the mean annual discharge. The turbines can be operated when the daily inflow plus the storage available in the conservation pool exceeds 20% of the turbine daily discharge capacity, since the dam may only operate just one or two turbines at a given instant if storage in the reservoir is low. This minimum turbine discharge parameter enables us to avoid accounting for the operation of each individual turbine. We assume that all turbine outflows immediately re-enter the main channel of the river without any ramping rate restrictions.

We also consider reservoir releases made via the spillway and an environmental flow bypass, i.e. a low-flow outlet that releases environmental flows equal to the annual seven-day minimum discharge with a ten-year recurrence interval (7Q10) when there is an insufficient combination of inflow and storage for hydropower generation. Unlike the actual John H. Kerr Reservoir, at which flood storage up to 320 ft (97.5 m) is released through a set of controlled gates with different discharge capacities, we assume that all water above the conservation storage pool (300 ft) passes downstream via a spillway with an infinite discharge capacity. We estimated the 7Q10 flow from the pre-dam record using the kappa distribution with the lmomco package in R (R Core Team, 2016). We computed seven-day low flows for each calendar year instead of Oct 1 – Sep 30 water years to include flows recorded during periods spanning the months of September and October.

The gross elevation head that indicates the potential energy available for hydropower is computed as the difference between the reservoir water surface elevation on a given day and the tailwater elevation below the dam. We assume this elevation to be constant and set it to 62.2 m (204 ft), the midpoint of the 199 ft -209 ft range reported in USACE (1992). We assume that the net elevation head is 10% lower than this elevation difference to account for friction losses. The hydropower generated on a given day HP_t, measured in kilowatt-hours (kWh), is computed as follows:

$$HP_t = 24 * 9.807 * \varepsilon * Q_{turb,t} * h_{net,t} \tag{A1}$$

where ε denotes the efficiency of hydropower production, which is assumed to be 80 percent at all times, $Q_{turb,t}$ is the turbine outflow on each day (m^3/s), and $h_{net,t}$ indicates the net head (m) on each day. The coefficients of 24 and 9.807 represent the number of hours in a day and the rate of gravitational acceleration at the Earth's surface (m/s^2), respectively.

Tables

Decision Making Approach	High flows (Q5)	Low flows (Q95)	Overall alteration
Deterministic	**Alteration.** Exceeds the 30% decrease threshold, an adverse ecological impact will occur if operations are not changed.	**Alteration.** Exceeds the 50% increase threshold, an adverse ecological impact will occur if operations are not changed.	**Alteration.** Thresholds at both sites are exceeded, an adverse ecological impact will occur if operations are not changed.
Null hypothesis significance testing ($p < 0.05$)	**No alteration.** $\alpha = 0.09$. Alteration insignificant, no significant risk of an adverse ecological impact.	**No alteration.** $\alpha = 0.10$, Alteration insignificant, no significant risk of an adverse ecological impact.	**No alteration.** $\alpha = 0.18$, Alteration insignificant, no significant risk of an adverse ecological impact.
Fixing the type I error probability α at 0.20, using decision tree	**Hydropower and ecological regret probabilities are similar.** Even though α is much higher than 0.05, β (0.38) indicates a high likelihood that test will not recommend changing the reservoir operating rule when necessary.	**Hydropower and ecological regret probabilities are similar.** Even though α is much higher than 0.05, β (0.26) indicates a high likelihood that test will not recommend changing the reservoir operating rule when necessary.	**Substantial regret probabilities, decision depends on relative hydropower and ecological values.** While the hydropower and ecological regret probabilities are 0.44 and 0.46, respectively, the greater potential ecological consequences of HP_{FOI} make changing to HP_{ROR} worthwhile.
Optimal α and β (minimizing total regret)	**The hydropower regret dominates because hydropower losses are less important than ecosystem losses.** α is much higher (> 0.99) than β (< 0.01). The lower consequence of hydropower regrets makes a higher α much	**The hydropower regret dominates because hydropower losses are less important than ecosystem losses.** α is much higher (> 0.99) than β (< 0.01). The lower consequence of hydropower regrets makes a higher α much	**The hydropower regret dominates because hydropower losses are much less important than ecosystem losses.** Since α is very high for both high and low flows, it is also very high for overall alteration. Similarly, since β is

	more tolerable.	more tolerable.	very low for both flows, it is very low for overall alteration.

Table 1: Management implications of different decision-making approaches.

Figures

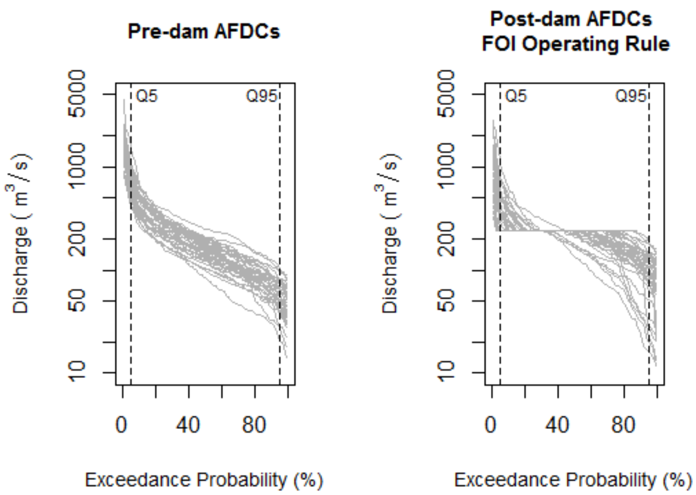

Figure 1: Pre- and post-dam annual flow duration curves from the stylized example in Sec 5.

Decision Rule	Unknown Truth	
	Alteration not above threshold **No Alteration, NA**	**Alteration above threshold** **Alteration, A**
Keep reservoir operation rules **(Conclude No Alteration)** **CNA**	$1-\alpha$ P(CNA|NA)	Under-protection error (Type II error) β P(CNA|A)
Change reservoir operation rules **(Conclude Alteration)** **CA**	Over-protection error (Type I error) α P(CA|NA)	$1-\beta$ P(CA|A)

Figure 2: Confusion matrix for testing hypotheses of hydrologic change, defining unknown true outcomes, decision rules with table entries showing the likelihood of the various possible outcomes (i.e. type I and type II errors).

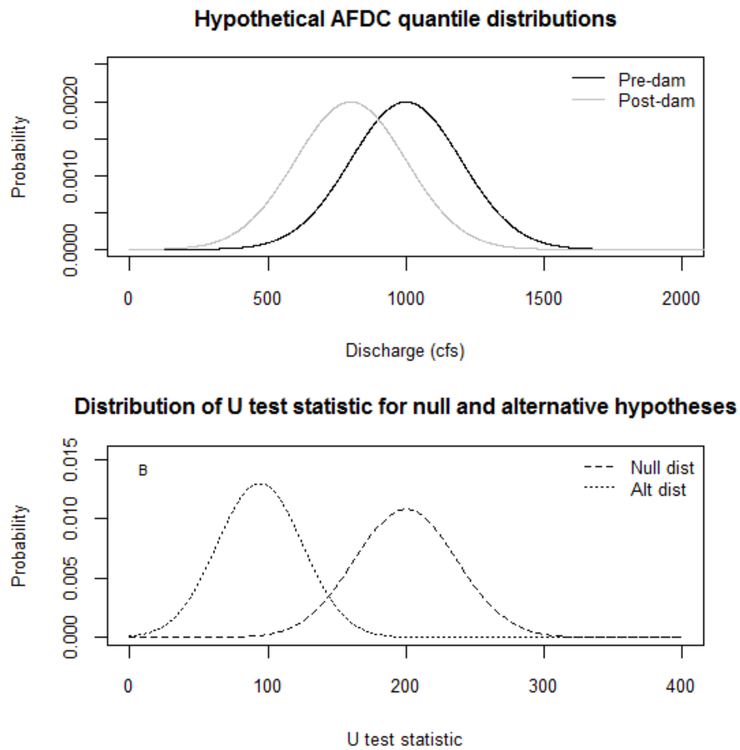

Figure 3: Mann-Whitney-Wilcoxon test: (a) Distributions of hypothetical pre- and post-dam AFDC quantiles/flow statistics and (b) null and alternative hypothesis distributions

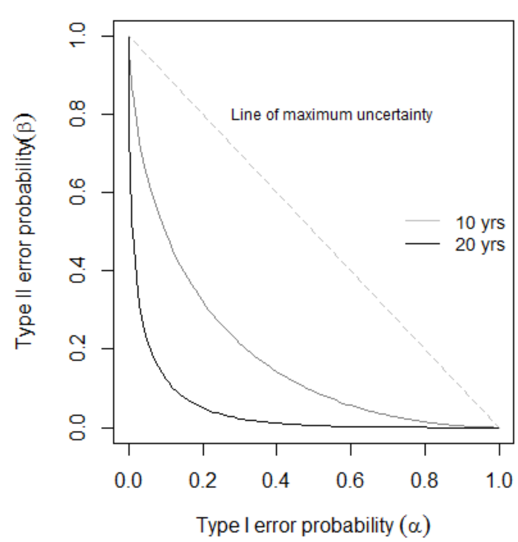

Figure 4: Tradeoff between type I and II errors for hypothetical records of different lengths.

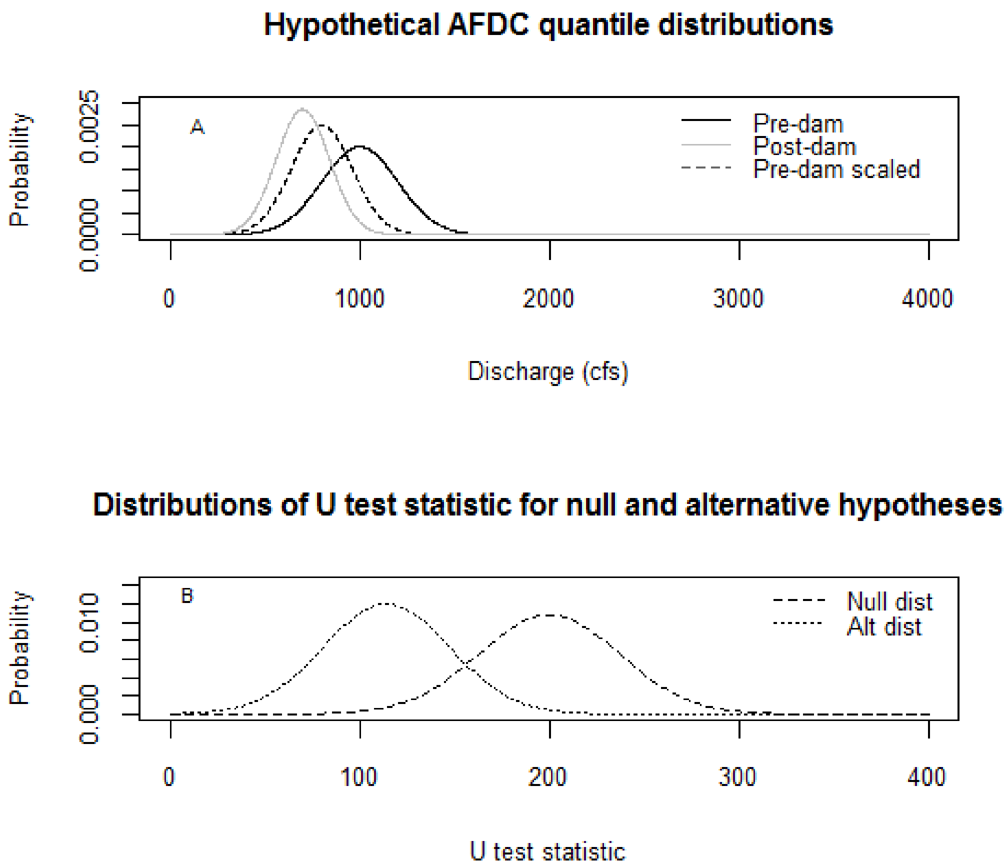

Figure 5: Mann-Whitney-Wilcoxon test with a percent deviation threshold.

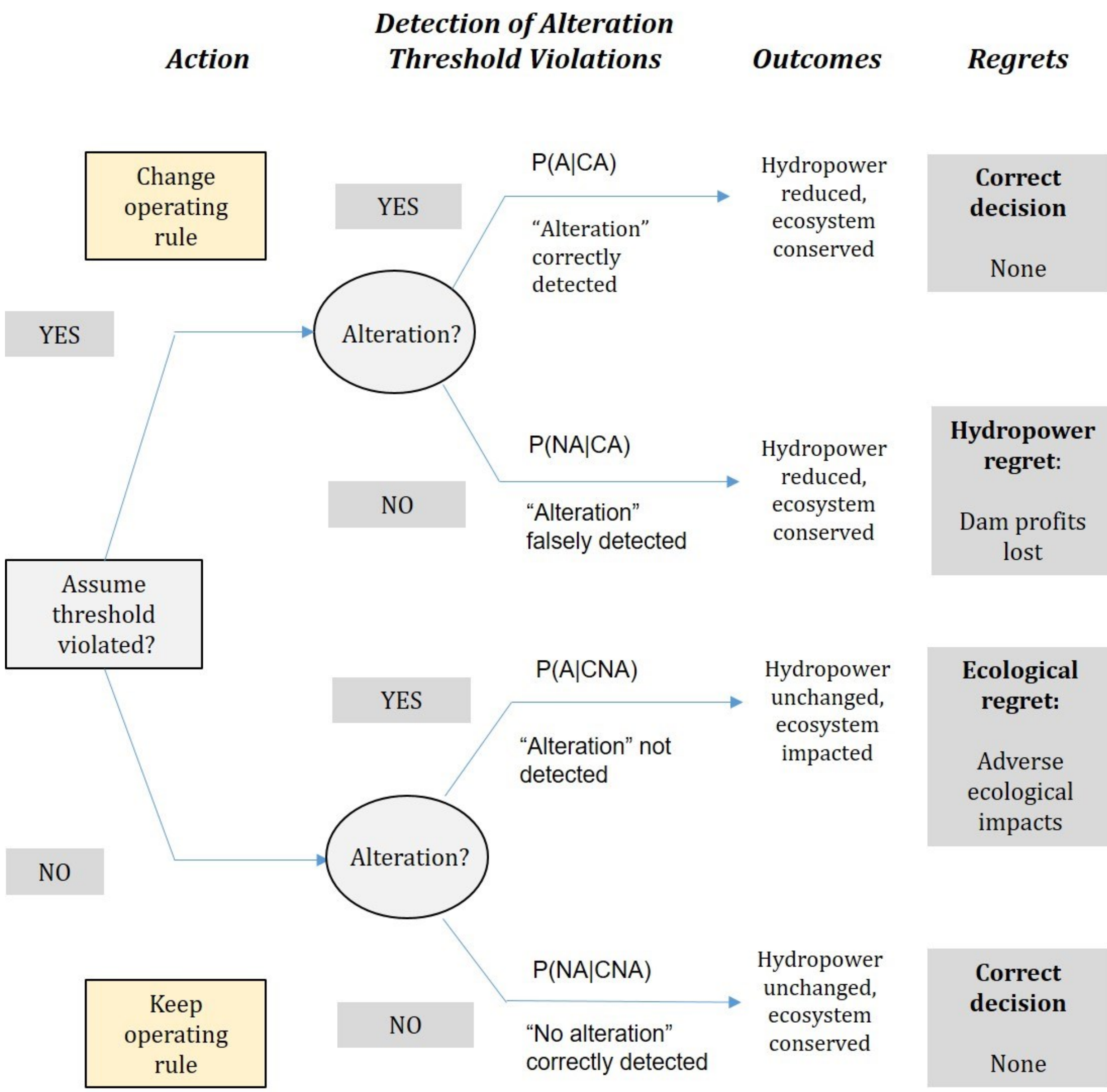

Figure 6: Bayesian decision tree for incorporating the uncertainty of hydrologic alteration into dam operating decisions

		Unknown Truth	
		No alteration threshold violation **P(NA)**	**Alteration threshold violation** **P(A)**
Decision Rule	**No protection implemented** **P(CNA)**	P(NA|CNA) $\frac{(1-\alpha)}{(1-\alpha)+\beta}$	Ecosystem regret probability P(A|CNA) $\frac{\beta}{(1-\alpha)+\beta}$
	Protection Implemented P(CA)	Hydropower regret probability P(NA|CA) $\frac{\alpha}{\alpha+(1-\beta)}$	P(A|CA) $\frac{(1-\beta)}{\alpha+(1-\beta)}$

Figure 7: Regret probabilities for decision analysis in Figure 6 based on a non-informative prior probability (0.5) of alteration.

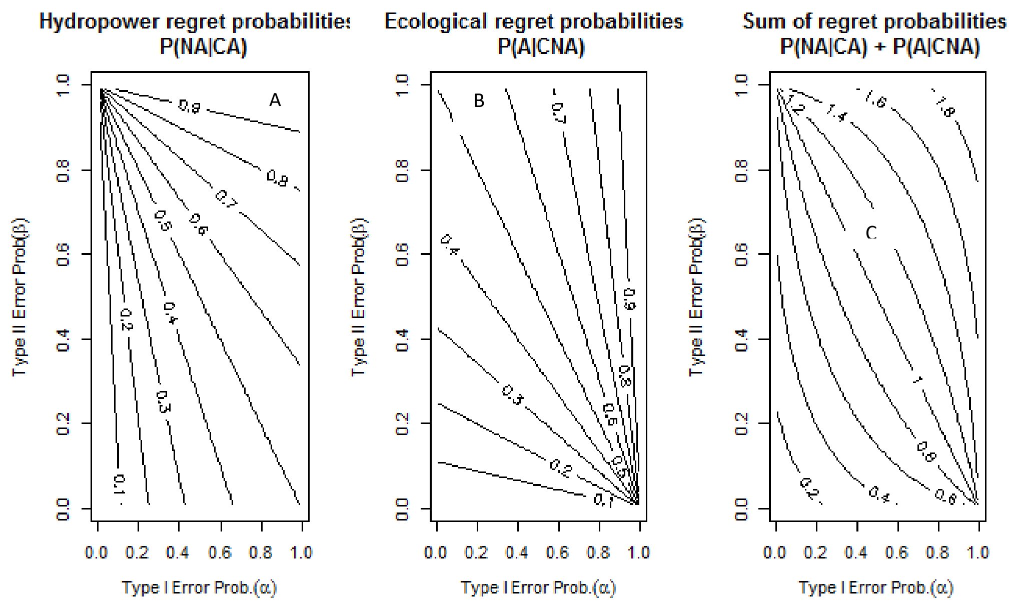

Figure 8: Relationship between hypothesis test error probabilities and decision regret probabilities (contours)

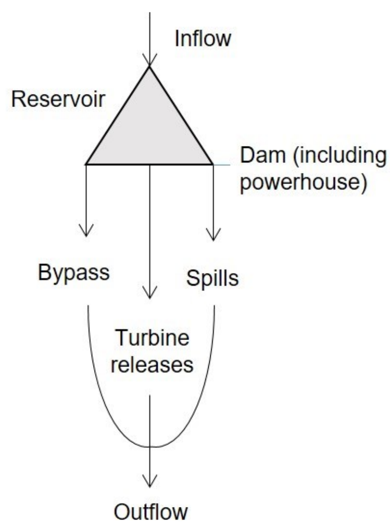

Figure 9: Inflows and outflows of hypothetical reservoir

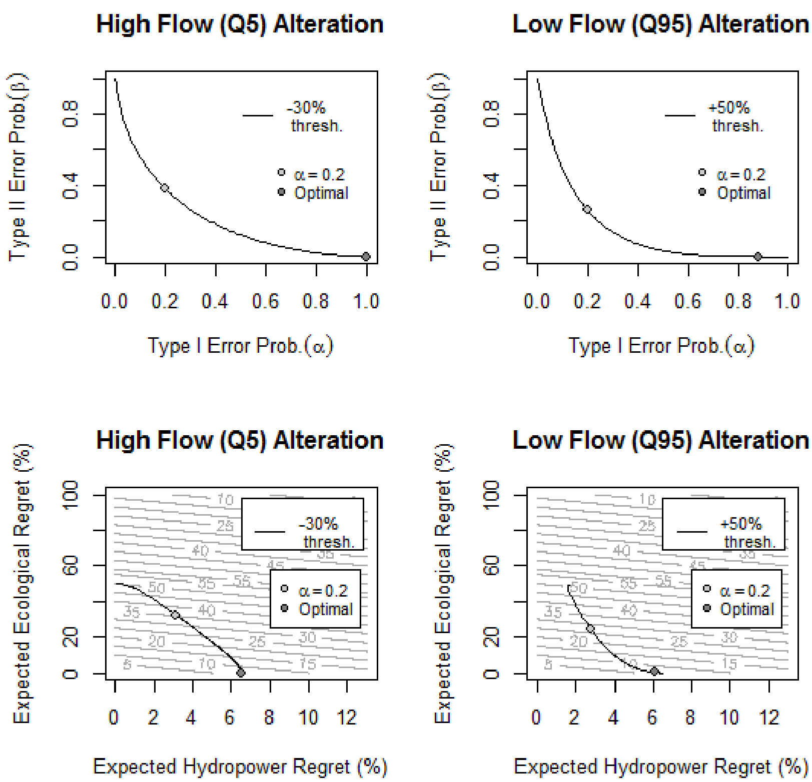

Figure 10: Tradeoffs between type I and II errors and regret probabilities for high- and low-flow alteration, Threshold Set 2.

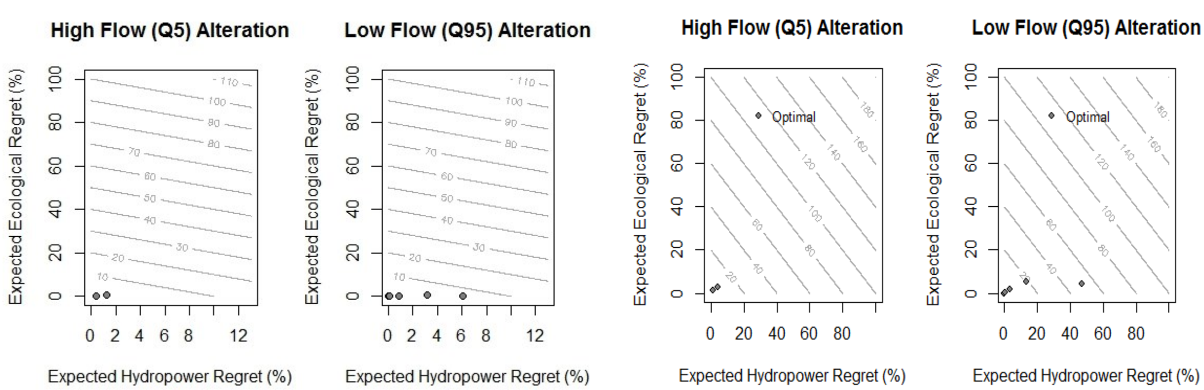

Figure 11: Optimal regret combinations when (a) a one-percent decrease in hydropower is weighted equally to a 1% increase in the likelihood of adverse ecological impacts, and (b) hydropower and ecological regret probabilities are weighted equally.