# Decision Trees for Incorporating Hypothesis Tests of Hydrologic Alteration into Hydropower–Ecosystem Tradeoffs

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

**Abstract:** Short streamflow records make it difficult to determine the extent to which discharge changes in excess of ecological thresholds are due to dam operations or natural variability. Unnecessary changes to reservoir operating rules can reduce off-stream benefits, whereas no changes to rules when thresholds are exceeded can degrade downstream riverine ecosystems. We introduce a Bayesian decision tree approach to a hypothetical hydropower–ecosystem decision problem that compares expected in-stream and off-stream losses resulting from incorrect decisions. Expected losses are computed using loss probabilities derived using Bayes' theorem, type I and II errors, and prior probabilities of alteration. Decision-tree recommendations compared with those from deterministic and null hypothesis significance testing under a variety of conditions illuminate the benefits of including valuations of hydropower and ecological losses as well as type II error probabilities in reservoir operation decisions. This is the first study to both introduce and demonstrate the value of Bayesian decision trees for addressing tradeoffs between hydropower and ecosystem benefits and losses. **DOI: 10.1061/(ASCE)WR.1943-5452.0001184.** 

## Introduction

Prescribing reservoir operating rules that sustain predam riverine ecosystems has become an increasingly recognized challenge (e.g., Suen and Eheart 2006; Poff et al. 2007; Vogel et al. 2007; Jager and Smith 2008). Short streamflow records make it difficult to distinguish whether long-term hydrologic changes of ecological relevance are due solely to dam impacts or are also influenced by natural variability between two periods (e.g., Nikghalb et al. 2016). How can this sampling uncertainty, which may be especially large for short pre- and postdam streamflow records (e.g., Kennard et al. 2010; Williams 2018), be considered when making contentious decisions to change reservoir operations after dams have been constructed? What is the likelihood and impact of not changing operations when it is ecologically necessary to do so? Conversely, what is the likelihood and impact of unnecessarily reducing the off-stream benefits of a reservoir? Currently, there are few guidelines for incorporating sampling uncertainty into evaluations of tradeoffs between off-stream and in-stream benefits. Under data-limited circumstances, general flow alteration guidelines that have been advocated often do not distinguish among the flow requirements of different riverine species (Smakhtin et al. 2004; Richter et al. 2012; Eriyagama et al. 2016). It can be especially difficult to disentangle flow alteration effects of dam operations from other differences between pre- and postdam periods when there is only a gauging station downstream of a dam, and no reservoir water level measurements (or satellite-based estimates) from which to deduce flow alteration through a water balance.

Since many efforts to monitor ecological impacts of flow alteration are hypothesis driven (e.g., Downes et al. 2002; Mudge et al. 2012b), statistical decision theory (Wald 1939) offers a potential framework for integrating sampling uncertainty, expressed as the likelihood of type I and II errors, into tradeoff evaluations. Indeed, such statistical decision-theoretic approaches have been adopted in numerous water resources applications (e.g., Peterman 1990; Mapstone 1995; Hobbs et al. 1997; Mudge et al. 2012b; Rosner et al. 2014) and have also been recommended for environmental flows (Downes et al. 2002; Growns 2004; Hering et al. 2010; Bark et al. 2013; Kroll et al. 2015; Gillespie et al. 2015). While many of these prior studies have noted that type I and II errors can correspond to over- and underprotection errors associated with environmental flow prescriptions, they have not demonstrated the integration of the likelihood of such errors into dam-operation decision problems.

To create a statistical, decision-theoretic screening tool that supports dam operation decisions, we must consider some prior methodological recommendations. First, given the common legal and regulatory infeasibility of completely preserving natural flows (Kopf et al. 2015) as well as the insensitivity of some species to mild flow alteration (e.g., Poff and Zimmerman 2010; Ceola et al. 2018), we must account for thresholds of alteration beyond which a riverine ecosystem adapted to historical flow conditions may be adversely affected (Kendy et al. 2012). We use hypothesis testing

[1]Dept. of Civil and Environmental Engineering, Tufts Univ., Medford, MA 02155; Hydrologist, USGS, Reston, VA 20192 (corresponding author). ORCID: https://orcid.org/0000-0002-9485-3332. Email: jhecht@usgs.gov

[2]Research Professor, Dept. of Civil and Environmental Engineering, Tufts Univ., Medford, MA 02155. ORCID: https://orcid.org/0000-0001-9759-0024

[3]Assistant Professor, Urban Dynamics Institute, Oak Ridge National Laboratory, Oak Ridge, TN 37830. ORCID: https://orcid.org/0000-0002-5551-3140

[4]Professor, Dept. of Environmental Resources Engineering, State Univ. of New York—College of Environmental Science and Forestry, Syracuse, NY 13210.

[5]Professor, Dept. of Biology, Tufts Univ., Medford, MA 02155. ORCID: https://orcid.org/0000-0002-3571-2652

Note. This manuscript was submitted on October 10, 2018; approved on September 24, 2019; published online on February 18, 2020. Discussion period open until July 18, 2020; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Water Resources Planning and Management***, © ASCE, ISSN 0733-9496.

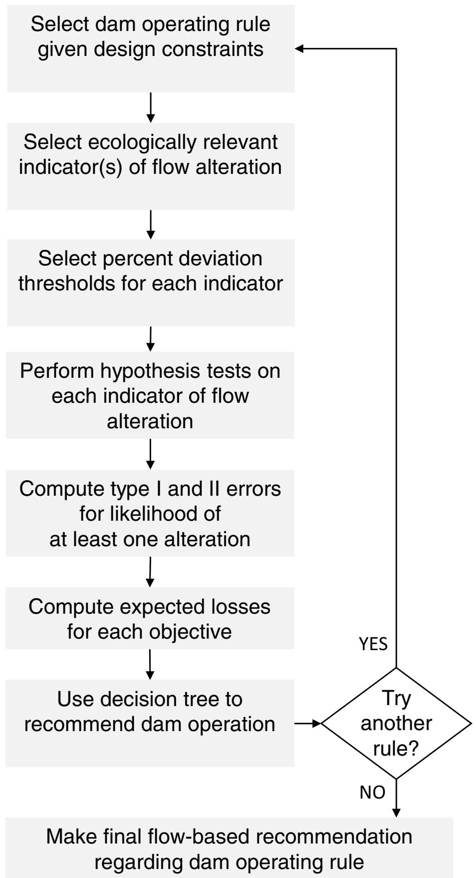


**Fig. 1.** Flow diagram showing sequence of methods leading to decision tree recommendations.

errors to determine the likelihood of incorrect decisions regarding threshold exceedances. However, one distinguishing feature of our approach is our use of Bayes' theorem to derive expected loss probabilities using type I and II errors from a frequentist hypothesis test assessing changes in flow. Prior efforts that have employed hypothesis testing errors to estimate costs of incorrect decisions (e.g., Reitsch 1976; Field et al. 2004; Mudge et al. 2012b; Rosner et al. 2014) have not distinguished the likelihood of hypothesis testing errors from the more important likelihood of making a decision resulting in regret. Importantly, the probability of a hypothesis test error constitutes the likelihood of a conclusion from a statistical test conditional upon an unknown truth, whereas the probability of a regretful decision is the likelihood of an unknown truth conditioned upon a statistical test outcome. We show that Bayes' theorem can relate type I and type II errors to the probabilities of changing hydropower operating rules when ecologically unnecessary and not changing them when ecologically necessary, respectively. We henceforth refer to these probabilities of regretful decisions as loss probabilities. Another distinguishing feature of our work is that we employ a field significance (a.k.a. multiple comparison) test to determine the type I and II errors associated with threshold-exceeding changes to multiple flow indicators relevant for riverine ecosystems.

To address these needs, we introduce a Bayesian statistical decision tree that considers situations in which flow alteration exceeding hypothesized thresholds may be the most critical impact that dams have on riverine ecosystems (as opposed to other factors, such as passage barriers). We demonstrate our approach using a hypothetical baseload hydropower dam that reduces high flows and increases low flows while aiming to produce as much as energy as possible for a large regional energy grid. The goal is to determine whether we can maintain a hydropower-friendly operating rule or if we should switch to one that poses fewer ecological risks. While many recent studies have sought rules that optimize hydropower-ecosystem tradeoffs (e.g., Jager 2014), in practice, existing rules are often compared to a single proposed alternative. However, our method can be used iteratively to evaluate a variety of flow indicators, percent deviation thresholds, and operating rules (Fig. 1). It is especially useful for detecting flow alterations to which ecosystems may exhibit gradual or lagged responses, along with identifying sites for monitoring or flow restoration activities. While we will demonstrate our approach with a hydropower example, it is also intended for reservoirs offering other off-stream benefits. After introducing this hypothetical baseload hydropower example, we formulate a hypothesis test for examining the likelihood that changes in flow indicators exceeding percent deviation thresholds are due to dam operations alone. Next, we apply the Bayesian decision tree to deduce expected losses associated with decisions to maintain or change an existing operating rule. Finally, we compare statistical decision-tree recommendations with those derived from a deterministic approach and null hypothesis significance testing (NHST), before discussing limitations and possible extensions.

## Setting: A Hypothetical Baseload Hydropower Dam

### *Reservoir Simulation Model and Operating Rules*

We introduce our decision-theoretic approach using an example featuring a hypothetical yet realistic hydropower reservoir. We employ a *with versus without dam* experiment to avoid having to account for natural hydrologic variability and other changes upstream of the dam between pre- and postdam periods. The daily inflows we use for this hypothetical reservoir, inspired by the John H. Kerr Reservoir in North Carolina, USA, come from a 37-year daily discharge time series (1913–1949) from the USGS station (02080500) on the Roanoke River at Roanoke Rapids, North Carolina (USGS 2019). The reservoir stores 24.3% of its mean annual inflow during this period.

While we will demonstrate our hypothesis testing approach using paired records, this same approach can be applied to unpaired pre- and postdam records if one believes that other changes in watershed conditions may not lead to a misleading conclusion of dam-induced alteration. To focus on hydropower-induced flow changes, we do not consider net evaporation or seepage or sedimentation-induced changes in storage. We also assume the reservoir is full at the outset of the postdam period we simulate given that the reservoir stores much less than 1 year of inflow. While we utilize some design parameters from the John H. Kerr reservoir, such as storage capacity, our reservoir operations model coded using R statistical software (R Core Team 2019) simulates operations entirely differently from its actual releases. In reality, its releases are also driven by flood control and diurnal energy price variability, as well as water supply, recreation, and fish and wildlife objectives [see USACE (2012), Kern et al. (2012) and references cited therein]. Appendix S1 compares actual and stylized parameters. Downstream releases from the reservoir can be made via (1) outflows from turbines situated in an integral powerhouse built into the dam, (2) a low-flow outlet for sustaining minimum flow during droughts, and (3) spills during high-flow periods (Fig. 2). The turbines can release between 20% and 110% of the mean annual discharge (239 $m^3/s$) whenever storage exceeds 32.3% of the reservoir's capacity (the dead storage fraction of the total reservoir's storage capacity below the spillway elevation). We assume that all

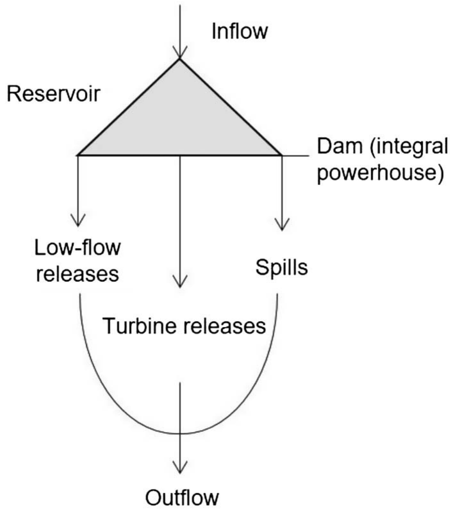


**Fig. 2.** Inflows and outflows of hypothetical reservoir.

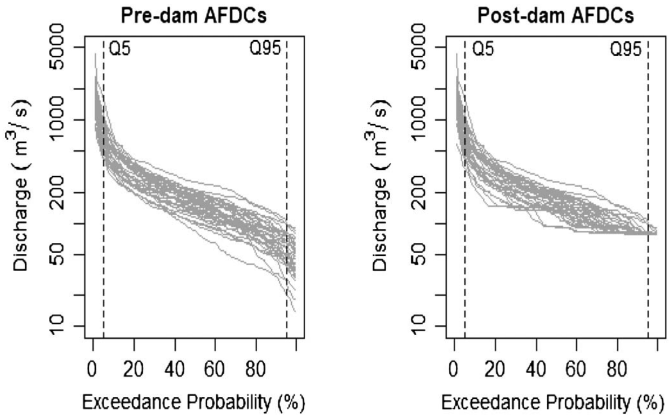


**Fig. 3.** Pre- and postdam annual flow duration curves from hypothetical example.

turbine outflows immediately reenter the main channel of the river without any ramping rate restrictions. The hypothetical dam has an installed generating capacity of 49.4 MW, which can power as many as 49,400 homes in an industrialized region (EPSA 2017). We compute the hydropower generated on a given day $HP_t$, measured in kilowatt-hours (kWh), as follows:

$$HP_t = 24 \times 9.807 \times \varepsilon \times Q_{\mathrm{turb},t} \times h_{\mathrm{net},t} \tag{1}$$

where $\varepsilon$ = efficiency of production (fixed at 80%); $Q_{\mathrm{turb},t}$ is the daily mean turbine outflow ($\mathrm{m^3/s}$); $h_{\mathrm{net},t}$ indicates the daily mean net head (m); 24 is the number of hours per day; and 9.807 is the rate of gravitational acceleration at sea level ($\mathrm{m/s^2}$).

When turbine releases do not meet the minimum outflow requirement under a hydropower-maximizing operating rule $HP_{max}$, environmental flow releases through a low-flow outlet can supplement turbine discharges, as allowed by available reservoir storage. Based on a variability-preserving operating rule similar to the one from Pastor et al. (2014), the dam must release (1) 40% of the mean monthly discharge during months whose mean discharge is at least 80% of the annual mean, (2) 50% of the mean monthly discharge during months whose mean discharge is at least 40% of the annual mean, and (3) 60% of the mean discharge when the mean monthly discharge is less than 40% of the annual mean. We assume that these releases are made daily during these months and that the dam gates can convey these flows downstream. Pastor et al. (2014) showed that this method correlated better with local environmental flow requirements than other methods based on predam discharge records. Since we assume our dam supplies power to a large regional grid with both hydropower and nonhydropower sources of energy, we do not consider hedging rules that reduce turbine releases in anticipation of droughts. Meanwhile, a gateless spillway with an infinite discharge capacity conveys excess inflow downstream.

### *Choosing Flow Alteration Indicators and Thresholds*

One challenge with implementing flow-based approaches for managing riverine ecosystems is choosing indicators of flow alteration that can easily be incorporated into reservoir operation rules while also recognizing that multiple aspects of flow alteration can affect riverine ecosystems. Flow duration curves (FDCs), which indicate the probability of exceeding a daily flow of a given magnitude, can be generated wherever continuous flow records are available or where they can be estimated reliably (e.g., Archfield et al. 2013). Consequently, they have been used in many water resources applications, including hydropower design, habitat assessment, flood abatement, and water quality evaluation (Vogel and Fennessey 1995), and have underpinned environmental flow management in data-poor regions (e.g., Jain 2015; Eriyagama et al. 2016). While FDCs offer a signature of flow variability over an entire station record, they cannot assess changes in typical years between pre- and postdam periods. In contrast, sets of annual FDCs (AFDCs) depict both within- and between-year hydrologic variability (e.g., Vogel and Fennessey 1994) and can be useful for evaluating changes in the annual distributions of flow indicators between pre- and postdam periods (Kroll et al. 2015). We examine postdam decreases in annual Q5 values and increases in annual Q95 values due to the flow homogenization effects of baseload hydropower (Fig. 3). The annual Q5 and Q95 flows are the mean daily discharges that are exceeded 5% and 95% of the time during a given year, respectively. (Unlike period-of-record FDCs, these curves may vary from year to year and a set of annual FDCs provides a signature of interannual flow variability.) These two flow indicators have been used in environmental flow assessments and guided reservoir operation decisions (e.g., Acreman et al. 2009). In the United States, indicators of alteration describing high-flow depletion and flow homogenization often have a strong association with indices of biological integrity (Carlisle et al. 2017). High in-channel flows, such as annual Q5, are essential for flushing sediment and pollutants and are often correlated with ecologically critical annual floods. Meanwhile, low-flow increases can cause drought-tolerant, native species to be replaced with generalists that favor less seasonally variable flows (Carlisle et al. 2011; Mims and Olden 2013). The acute flat-lining effects in Fig. 3 result from (1) the fixed turbine discharge capacity that constrains reservoir releases when the water level exceeds the minimum elevation of the conservation storage pool and (2) low-flow outlet releases made when turbine discharges do not fully meet monthly minimum flow requirements. We illustrate our test using two sets of hypothetical ecological thresholds. Threshold set 1 ($-50\%$ change in Q5 and $+50\%$ change in Q95) has thresholds exceeded by simulated changes under $HP_{max}$ ($-51\%$ change in Q5 and $+84\%$ change in Q95). Meanwhile, Threshold set 2 ($-60\%$ change in Q5 and $+90\%$ change in Q95) consists of more relaxed thresholds that are not exceeded by the changes under $HP_{max}$. These thresholds fall within the ranges of ones observed for various high- and low-flow indicators (e.g., Poff and Zimmerman 2010; Carlisle et al. 2011). While comparing AFDCs cannot account for changes in timing, FDCs computed over seasons or other problem-relevant

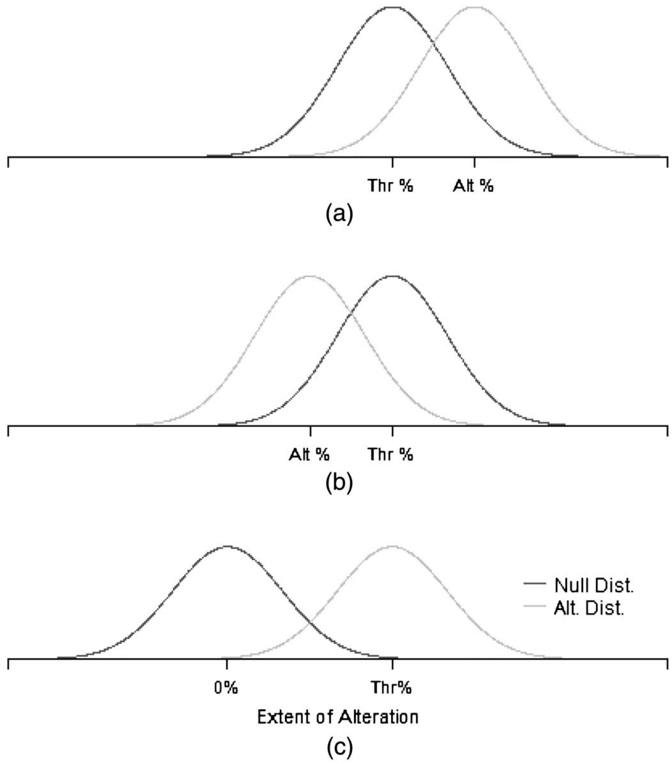


**Fig. 4.** Conceptual illustration of null and alternate hypothesis distribution locations relative to percent deviation thresholds: (a) for decision trees when the percent alteration (Alt %) exceeds a threshold (Thr %); (b) for decision trees when percent alteration does not reach a threshold; and (c) when using an effect size determined a priori.

timescales may be used instead, e.g., using seasonal FDCs for spawning periods (Gao et al. 2009).

## Testing for Violations of Hydrologic Alteration Thresholds

In this section, we present (1) a hypothesis test that detects when changes in typical values of flow indicators, such as AFDC quantiles, exceed percent-deviation thresholds, and (2) a multiple comparison (field significance) test that determines type I and II probabilities associated with the likelihood of one or more threshold violations when multiple indicators are examined.

### *Individual Indicators of Flow Alteration*

To obtain the probability of type I and II errors associated with decisions concerning violations of individual percent-deviation alteration thresholds, we sought a two-sample hypothesis testing framework that could (1) yield equal type I and II error probabilities when a flow alteration threshold is reached exactly, (2) recommend changing dam operating rules when a threshold is not exceeded but consequences of ecological losses are greater than hydropower ones, and (3) recommend not changing operating rules when a threshold is exceeded but hydropower consequences are greater than ecological ones. These objectives can be achieved by centering the distribution of the test statistic under the null hypothesis at the percent deviation threshold and then using the observed absolute effect size (defined as the difference between the observed mean percent deviation and the percent deviation threshold) to locate the alternative hypothesis distribution (Fig. 4). The effect size provides a measure of the degree to which the hypothesis test reveals departures from the null hypothesis. With this configuration, the null and alternative hypothesis distributions overlap if the threshold is exactly met, i.e., $\alpha = \beta = 0.5$. A one-tailed, two-sample hypothesis test determines whether percent changes in typical flow values are greater or less than a given percent deviation threshold. If the threshold is exceeded, then the type I error probability reflects the likelihood of concluding "alteration" when no alteration occurs. The type II error probability indicates the likelihood of incorrectly concluding no threshold exceedance when there is indeed "alteration" (Table 1). Conversely, if a threshold is not exceeded, type I errors reflect the likelihood of incorrectly concluding "no alteration" when a threshold is met or exceeded, and type II errors indicate the likelihood of incorrectly concluding that a threshold is met or exceeded when the truth is "no alteration."

Importantly, the two-sample testing framework we devise differs from a power analysis with effect sizes determined *a priori* [see environmental examples in the works of Downes et al. (2002), Field et al. (2004), and Mudge et al. (2012b)], in which the null hypothesis distribution is centered around 0% alteration and the alternative hypothesis distribution is centered around the percent deviation threshold, i.e., the hypothesized effect size. However, if (1) the null and alternative hypotheses have equal variances and (2) hydropower and ecological losses are equally consequential, this procedure suggests that operating rules causing alteration exceeding just *half* the hypothesized percent deviation threshold should be changed [Fig. 4(c)]. While $\beta$ is often close to 0.5 when the critical effect size is set equal to the observed effect size, stakeholders can choose any pair of hypothesis testing errors ($\alpha$, $\beta$) that falls on the test's type I–type II error tradeoff curve [see plots (a) and (b) in Fig. 5]. For instance, if they require specific type I or II error probabilities, such as a minimum of 80% power (a maximum type II error probability of 20%), they can reduce the *critical alteration value*, which would reduce the type II error at the expense of the type I error. Numerous two-sample hypothesis tests could be applied to obtain the type I and II error probabilities that our decision-theoretic framework requires [see the work of Kroll et al. (2015) for a comparative study]. Here, we use a nonparametric Mann-Whitney-Wilcoxon (MWW) test (a.k.a. a rank-sum test) to obtain values of hypothesis testing errors regarding differences in central tendency between pre- and postdam flow indicators. A Wilcoxon signed-rank test for paired data may seem more appropriate for our controlled *with or without dam* experiment. However, we apply an MWW test instead because pre- and postdam records

**Table 1.** Confusion matrix for testing hypotheses of flow alteration showing possible combinations of test decision rules and unknown true outcomes for cases when flow alteration exceeds percent deviation thresholds; table entries include likelihoods of type I and type II errors

| | Unknown truth | |
|---|---|---|
| Decision rule | Alteration not above threshold (no alteration, NA) | Alteration above threshold (alteration, $A$) |
| Keep reservoir operation rules (conclude no alteration) CNA | $1 - \alpha$ | $\beta$ (type II error probability) |
| Change reservoir operation rules (conclude alteration) CA | $\alpha$ (type I error probability) | $1 - \beta$ |

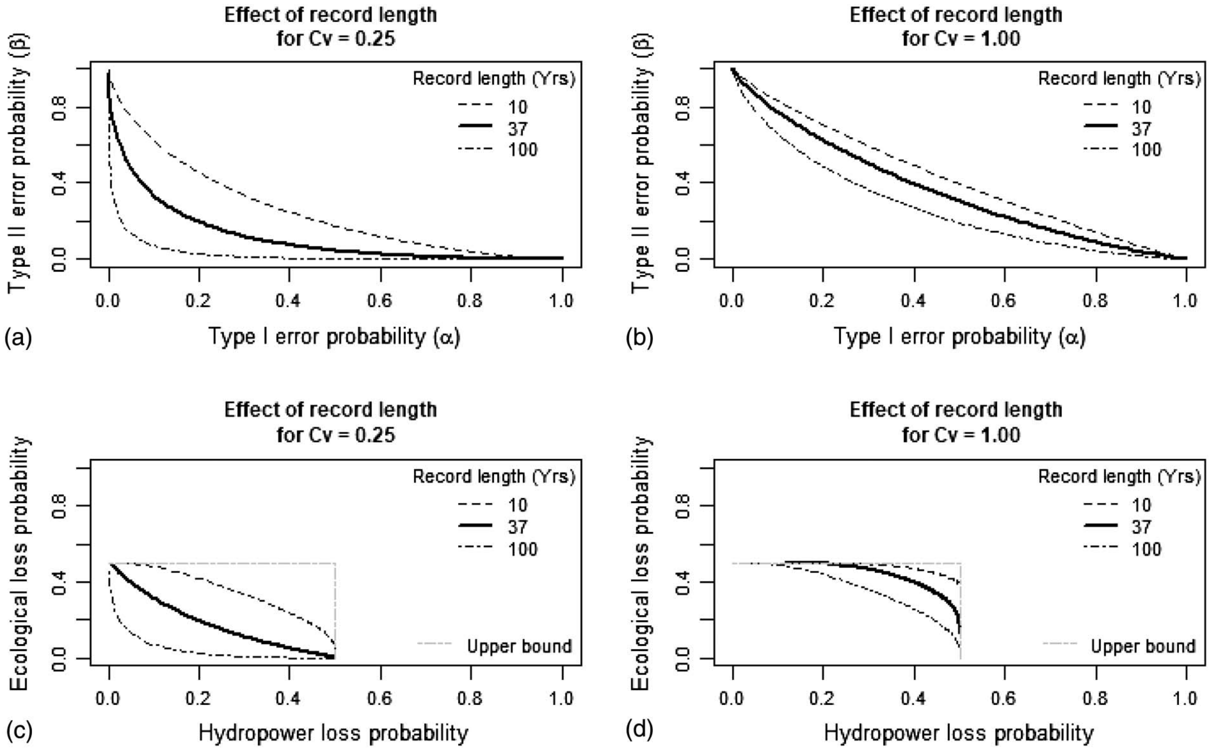


**Fig. 5.** Effects of record length and Cv on hypothesis testing errors and loss probabilities for a hypothetical threshold exceedance of 10%. Noninformative prior probabilities of no alteration and alteration (both 0.5) comprise upper limits of hydropower and ecological loss probabilities. Loss probability tradeoffs become concave downward when $\alpha + \beta > 0.5$.

are not paired in practice. The MWW test has been recommended for nonnormal distributions of flow indicators frequently observed downstream of dams (FitzHugh 2014) and is well suited for highly skewed data (e.g., Fay and Proschan 2010). We use a normal approximation of the $U$ test statistic, which requires pre- and postdam samples of 8 years each (Mann and Whitney 1947). See the work of Bellera et al. (2010) for guidance on using this approximation with pre- and postdam records of unequal length, and Fay and Proschan (2010) for sample sizes required to reject the null hypothesis using the exact MWW test statistic (four at $\alpha = 0.05$). Appendix S2 presents the MWW test in greater detail, including a power analysis approach based on assumed probability distributions for alternative hypotheses (Shieh et al. 2007). Appendix S3 describes our adaptation of the rank-based MWW test for examining exceedances of percent deviation thresholds.

## The Likelihood of At Least One Threshold Violation

We also assess the likelihood that dam operations violate 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 individual hypothesis tests applied to independent subsamples (e.g., Douglas et al. 2000). For demonstration purposes, we assume that high (Q5) and low flows (Q95) are independent; in practice one would expect some degree of correlation between high and low AFDC quantiles since large storm events often provide groundwater recharge that increases base flow during subsequent dry periods. We apply a hypothesis test where $H_0$ denotes no threshold violations and $H_A$ denotes at least one threshold violation to determine the likelihood of violating thresholds for *at least* one AFDC quantile. In other words, we are assuming that just one violation can impact a riverine ecosystem. To falsely conclude at least one alteration from a set of tests on two individual AFDC quantiles, both tests must result in type I errors. Thus, if we have $K$ independent indicators of flow alteration, the overall probability of a type I error $\alpha_{\text{overall}}$ is equal to the probability of falsely concluding at least one alteration, which is

$$\alpha_{\text{overall}} = \prod_{k=1}^{K} \alpha_k \tag{2}$$

In contrast, we want to know the overall likelihood of a type II error $\beta_{\text{overall}}$ resulting from falsely concluding no alteration when there is, in fact, alteration to *at least one* flow indicator. Thus, for $K$ independent AFDC quantiles, $\beta_{\text{overall}}$ is computed as follows:

$$\beta_{\text{overall}} = 1 - \prod_{k=1}^{K} (1 - \beta_k) \tag{3}$$

where $\beta_k$ = probability of a type II error for the $k$th AFDC quantile. The procedure in Eq. (2) differs from the familywise error rates (FWER) often used to assess field significance, which state the likelihood of *at least* one type I error (e.g., Hochberg and Tamhane 1987). For instance, let us say we were to use FWER for $\alpha_{\text{overall}}$ and had individual test results for two flow indicators, $\alpha_1 = 0.01$ and $\alpha_2 = 0.5$. It would suggest that a conclusion of at least one violation has a nearly 50% chance of being incorrect, i.e., $1 - [(1 - 0.01) \times (1 - 0.5)] = 0.495$, despite there being one individual test whose low type I error probability indicates a violation. In contrast, Eq. (2) indicates that the probability of two false positives is approximately 0.005, which reasonably suggests that unnecessary hydropower losses are much less likely (see the next section for computation of loss probabilities). When applying multiple comparison tests to flow indicators whose thresholds are not exceeded, Eqs. (2) and (3) are used to compute $\alpha_{\text{overall}}$ and $\beta_{\text{overall}}$, respectively.

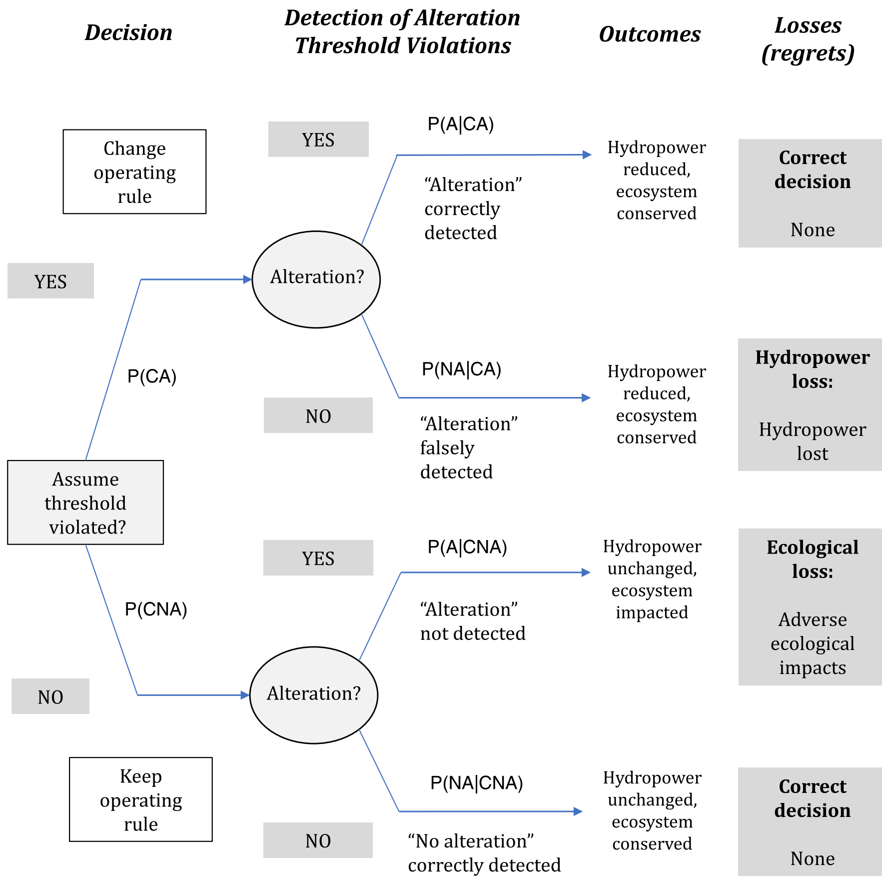


**Fig. 6.** Bayesian decision tree for incorporating the uncertainty of exceeding thresholds of hydrologic alteration into dam operating decisions.

## Bayesian Decision Trees for Informing Dam Operation Decisions

Another unique aspect of our work is a statistical decision tree that compares expected hydropower and ecosystem losses associated with over- and underprotection decisions. We demonstrate this tree for a case where the percent flow alteration is slightly greater than the hypothesized percent deviation threshold (Fig. 6). First, we apply Bayes' theorem to deduce loss probabilities from type I and II errors for which we must specify prior probabilities of alteration. While contention surrounding expert-elicited prior probabilities can make Bayesian applications challenging (e.g., Gelman 2008), their increasing use in river monitoring applications requiring expert judgment (Mudge et al. 2012b; Webb et al. 2015) motivates our approach. We illustrate this method with noninformative prior probabilities, i.e., a 50% chance of violating at least one threshold, an accepted assumption in multistakeholder environmental management applications with insufficient prior information (e.g., Field et al. 2004; Mudge et al. 2012a; Webb et al. 2015) and one that stakeholders with competing needs may be more likely to perceive as fair. Bayes' theorem yields the probability of modifying dam operating rules based on an erroneous conclusion of alteration exceeding the percent deviation threshold $P(\mathrm{NA}|\mathrm{CA})$:

$$P(\mathrm{NA}|\mathrm{CA}) = \frac{P(\mathrm{CA}|\mathrm{NA})P(\mathrm{NA})}{P(\mathrm{CA})} \tag{4}$$

where $P(\mathrm{NA})$ = probability of no alteration; and $P(\mathrm{CA})$ = probability of concluding alteration with a statistical test, the latter of which can also be expressed as follows:

$$P(\mathrm{CA}) = P(\mathrm{CA}|\mathrm{NA})P(\mathrm{NA}) + P(\mathrm{CA}|A)P(A) \tag{5}$$

$P(\mathrm{NA}|\mathrm{CA})$ in (4) can be interpreted as the hydropower loss probability because it indicates the likelihood of unnecessarily changing an operating rule given an incorrect conclusion of alteration. Substituting Eq. (5) into Eq. (4), we obtain

$$P(\mathrm{NA}|\mathrm{CA}) = \frac{P(\mathrm{CA}|\mathrm{NA})P(\mathrm{NA})}{P(\mathrm{CA}|\mathrm{NA})P(\mathrm{NA}) + P(\mathrm{CA}|A)P(A)} \tag{6}$$

Since we assume $P(\mathrm{NA}) = P(A) = 0.5$, and $P(\mathrm{CA}|\mathrm{NA}) = \alpha$ and $P(\mathrm{CA}|A) = 1-\beta$, $P(\mathrm{NA})$ and $P(A)$ can be removed from Eq. (9). Then, we can use $\alpha$ and $\beta$ to obtain $P(\mathrm{NA}|\mathrm{CA})$:

$$P(\mathrm{NA}|\mathrm{CA}) = \frac{\alpha}{\alpha + (1-\beta)} \tag{7}$$

We can also use Eqs. (4)–(7) to derive the ecological loss probability for concluding no alteration (CNA) beyond a threshold when, in fact, a threshold is exceeded:

$$P(A|\mathrm{CNA}) = \frac{\beta}{\beta + (1-\alpha)} \tag{8}$$

Table 2 shows all four possible combinations of unknown outcomes conditional upon hypothesis test conclusions of alteration or no alteration, including those associated with correct decisions. Importantly, these expressions for loss probabilities differ from previously used decision-theoretic procedures that would call for multiplying the type I and II errors with the prior probabilities of no alteration and alteration, respectively (Reitsch 1976; Field et al. 2004). Fig. 5 shows the relationship between loss probabilities and

**Table 2.** Loss probabilities based on a noninformative prior probability (0.5) of alteration

| Decision rule | Unknown truth: No alteration threshold violation $P(\mathrm{NA})$ | Unknown truth: Alteration threshold violation $P(A)$ |
|---|---|---|
| No protection implemented $P(\mathrm{CNA})$ | $P(\mathrm{NA}\vert\mathrm{CNA})\ \frac{(1-\alpha)}{(1-\alpha)+\beta}$ | Ecosystem loss probability $P(A\vert\mathrm{CNA})\ \frac{\beta}{(1-\alpha)+\beta}$ |
| Protection implemented $P(\mathrm{CA})$ | Hydropower loss probability $P(\mathrm{NA}\vert\mathrm{CA})\ \frac{\alpha}{\alpha+(1-\beta)}$ | $P(A\vert\mathrm{CA})\ \frac{(1-\beta)}{\alpha+(1-\beta)}$ |

hypothesis testing errors. The prior probabilities set the upper bounds on each of the loss probabilities. Also, the relationship between the hydropower and ecological probabilities becomes concave downward when the sum of type I and II error probabilities exceeds 0.5. Appendix S4 and Fig. S3 further illustrate the relationship between hypothesis testing errors to decision-relevant loss probabilities.

Next, we calculate the expected losses of dam operations decisions using loss probabilities, an approach which differs from classical efforts to maximize hydropower production subject to fixed environmental constraints. First, we compute the expected hydropower loss $\mathrm{EL_{HP}}$ from the perspective of dam operators in terms of (1) the capital costs associated with any dam retrofit needed to achieve new flow targets and (2) the difference in the value or amount of hydropower production (HP) between a hydropower maximizing operating rule $\mathrm{HP_{max}}$ and a reference run-of-river (ROR) operating rule $\mathrm{HP_{ROR}}$:

$$\mathrm{EL_{HP}} = P(\mathrm{NA}|\mathrm{CA}) \times [C_{\mathrm{DamRetro}} + (\mathrm{HP_{max}} - \mathrm{HP_{ROR}})] \tag{9}$$

Substituting Eq. (7) into Eq. (9) yields

$$\mathrm{EL_{HP}} = \frac{\alpha}{\alpha + (1-\beta)} \times [C_{\mathrm{DamRetro}} + (\mathrm{HP_{max}} - \mathrm{HP_{ROR}})] \tag{10}$$

One can quantify hydropower losses in terms of costs, energy generation, or other relevant performance indicators from both producer and consumer perspectives. In our example, we focus on comparing differences in energy generation and ignore retrofitting costs that producers might bear as well as changes in energy prices for consumers. While operating a large storage reservoir as an ROR facility is unrealistic and could require the construction of additional flow bypass infrastructure, ROR operating rules still provide an instructive lower-bound reference (e.g., Kern et al. 2012). In practice, stakeholders could elect to evaluate another operating rule whose ecological impact is of much less concern as the alternative to $\mathrm{HP_{max}}$ in the decision tree.

Next, we compute the expected ecological losses $\mathrm{EL_{eco}}$. When dam operation changes a measurable ecological flow indicator, the probability that a decision will lead to an undesirable ecological state $P(A|\mathrm{CNA})$ can serve as a weight for determining $\mathrm{EL_{eco}}$ so that

$$\begin{aligned}\mathrm{EL_{eco}} &= P(A|\mathrm{CNA}) \times (E_{\mathrm{eco}} - E_{\mathrm{HP}}) \\ &= \frac{\beta}{(1-\alpha)+\beta} \times (E_{\mathrm{eco}} - E_{\mathrm{HP}})\end{aligned} \tag{11}$$

Alternatively, ecological indicators may include measures of species or ecosystem health, or monetary values of ecosystem services, such as fisheries or costs of ecological restoration. Here, we assume that 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., Railsback et al. 2016). As the responses of species and ecosystems to flow alteration vary widely (e.g., Poff and Zimmerman 2010), users can value losses associated with threshold exceedances as they deem appropriate for a given management context. In a study that associated flow alteration thresholds with riverine ecosystem disturbances, Carlisle et al. (2011) considered disturbed basins to be ones where indices of biological diversity registered values less than 90% of ones expected at appropriate reference sites (a lower-end estimate). To illustrate our approach, we assume losses of 10% when flow alteration thresholds are exceeded. We also consider changes in recommended operating rules when threshold exceedances are expected to cause greater losses. As our example shows, noncommensurate objectives can be compared when both are expressed in relative percent terms, e.g., the hydropower produced under $\mathrm{HP_{ROR}}$ as a percentage of production under $\mathrm{HP_{max}}$.

## Results

### *General Effects of Threshold Exceedances and Costs of Decision Consequences*

First, we compare decisions that the deterministic (difference in means), NHST, and decision tree methods recommend under different combinations of (1) observed effect size relative to the percent-deviation threshold for one flow indicator and (2) differences between the costs associated with hydropower and ecological losses. The plots in the bottom and middle rows of Fig. 7 show that both deterministic and NHST methods are insensitive to these cost differences, whereas cost differences play a critical role in our approach. The deterministic method indicates that a threshold exceedance of any magnitude induces greater ecological losses than hydropower ones, whereas NHST indicates that the effective threshold is the percent flow alteration at which $\alpha = 0.05$ (FitzHugh 2014). This effective threshold decreases gradually as records become longer.

To examine the decision tree's performance (top row) further, we computed averages of 1,000 runs with two random samples drawn from normal distributions with effect sizes (differences in means) measured in standard deviations. The percent alteration at which changing operating rules is recommended becomes lower (higher) as the relative cost of ecological losses increases (decreases). Importantly, the decision tree can recommend changing an operating rule if a threshold is not exceeded when potential ecological losses are greater than hydropower ones. Conversely, it can suggest keeping a rule if a threshold is exceeded but hydropower costs are greater. The reduced slope of the boundary demarcating the two decisions around an effect size of 0.0 and hydropower loss fraction of 0.5 stems from differences between zero and the mean of small samples, as its prominence decreases when sample sizes are increased from 10 to 37. A comparison of results for these two sample sizes illustrates that shorter records make decisions more



J. Water Resour. Plann. Manage., 2020, 146(5): 04020017

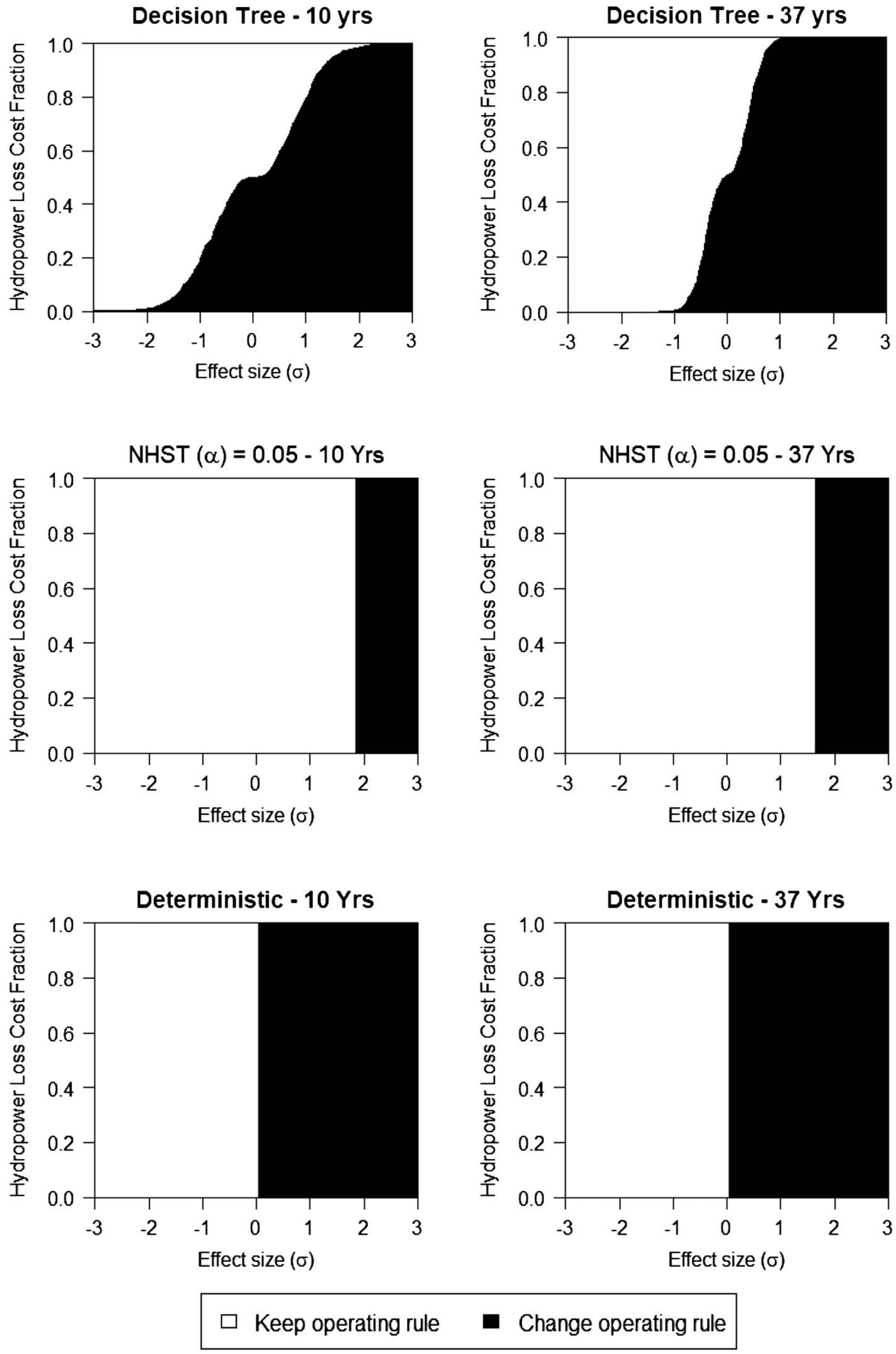


**Fig. 7.** Decision recommendations for different combinations of effect size (measured in standard deviations) and ratio of costs associated with hydropower and ecological losses.

sensitive to the relative costs of losses. Fig. 5 also shows that this uncertainty affects longer records, especially when flow indicator values vary substantially from year to year.

### *Hypothetical Reservoir Simulation Results*

Our simulations also demonstrate that decision trees can suggest different operating rules than conventional approaches (Table 3). First, we examine Threshold set 1 in which Q5 cannot decrease by more than 50% and Q95 cannot increase by more than 50%. Since the mean changes in annual Q5 (−51%) and Q95 (+84%) values exceed these thresholds, a deterministic test indicates excessive alteration. Note that, in some cases, differences in means may slightly exceed a threshold when the MWW test has a type I error probability greater than 50% since the $U$ statistic of central tendency difference is not equivalent to a difference in means.

Next, while the one-tailed NHST tests applied to Q5 and Q95 are both insignificant ($\alpha_{Q5} = 0.239$, $\alpha_{Q95} = 0.093$), together, they indicate a low likelihood of falsely concluding at least one alteration ($\alpha_{overall} = 0.022$). Meanwhile, type II error probabilities for annual Q5 and Q95 values are 0.731 and 0.016, respectively. Since we are concerned with the probability of missing one or more violations, the high type II error probability for annual Q5 produces a high value of $\beta_{overall}$ (0.736). These hypothesis testing errors lead to

**Table 3.** Recommendations (in bold) obtained with decision tree and conventional decision methods

| Decision method | Deterministic (difference in means) | Null hypothesis significance testing ($p < 0.05$) | Decision tree |
|---|---|---|---|
| Threshold set 1: (Q5 = −50%, Q95 = +50%) | **Switch to $\mathbf{HP_{OR}}$**. Alteration (Q5 = −51%; Q95 = 84%) exceeds both thresholds. Even one violation would prompt a switch to $HP_{ROR}$. | **Switch to $\mathbf{HP_{OR}}$**. Although Q95 and Q95 alteration are both insignificant ($\alpha_{Q5} = 0.239$; $\alpha_{Q95} = 0.093$), the probability of *at least* one threshold is significant ($\alpha_{overall} = 0.022$). | **Switch to $\mathbf{HP_{ROR}}$**. Hydropower loss probability is 47.0%, while ecological loss probability is 48.8%. Expected ecological losses (48.8% × 10% = 4.9%) are greater than hydropower ones (47.0% × 6.8% = 3.1%). |
| Threshold set 2: (Q5 = −60%, Q95 = +90%) | **Keep $\mathbf{HP_{max}}$**. Alteration does not exceed either threshold. | **Keep $\mathbf{HP_{max}}$**. Alteration is highly insignificant ($\alpha_{Q5} = 0.946$; $\alpha_{Q95} = 0.984$; $\alpha_{overall} = 0.931$). | **Switch to $\mathbf{HP_{ROR}}$**. Even though the hydropower loss probability is higher (9.3%) than the ecological one (6.8%), expected ecological losses (6.8% × 10% = 0.7%) exceed expected hydropower ones (9.3% × 6.8% = 0.6%). |



hydropower and ecological loss probabilities of 7.7% and 42.9%, respectively, which indicate that hydropower losses must be nearly six times larger to warrant keeping $HP_{max}$. Meanwhile, the overall type I and II errors suggest that much greater hydropower losses would be necessary to keep the operating rule.

When we assess losses in percent terms, the decision tree recommends changing the operating rule because using $HP_{ROR}$ (297.2 GWh) instead of $HP_{max}$ (319.0 GWh) only reduces the mean annual hydropower output by 6.8% due to the very low interannual flow variability at this site (mean annual flow CV = 0.22). In percent terms, the expected hydropower loss is 0.5% (7.7% × 6.8%), whereas the expected ecological loss is 4.3% (42.9% × 10%). While changing the operating rule is preferable in this example, the marginal hydropower benefits of reservoir storage may be greater in places with greater seasonal and interannual flow variability. In contrast, threshold exceedances often reduce some ecological indicators by more than 10% (Poff and Zimmerman 2010), which would lead to a stronger recommendation to change the rule.

Next, Threshold set 2 (changes of −60% and +90% for Q5 and Q95, respectively) shows that the decision tree can recommend operating rule changes even when thresholds are not exceeded. In this case, the alternative hypothesis is that alteration does not meet or exceed a threshold (see the "Testing for Violations of Hydrologic Alteration Thresholds" section), we obtain $\alpha_5 = 0.052$ and $\alpha_{95} = 0.016$ and $\beta_5 = 0.096$ and $\beta_{95} = 0.984$, with the wide range of type II error probabilities stemming from the nonnormality of the postdam flow distribution. Since the test is being applied in the opposite direction, we use Eqs. (10) and (11) to compute the hydropower and ecological loss probabilities, respectively. While the hydropower loss probability (9.3%) exceeds the ecological one (6.8%), the difference between ecological losses (10%) and hydropower ones (6.8%) is slightly greater. Thus, expected hydropower losses (0.63%) are slightly lower than ecological ones (0.68%), which causes the tree to recommend changing the dam operating rule. The small difference between these expected losses further underscores the importance that the relative costs can have on decision tree outcomes.

## Discussion and Conclusions

To address a growing interest in protecting ecosystems downstream of dams using percent-deviation flow-alteration thresholds, we modified a nonparametric hypothesis test to distinguish dam-induced changes exceeding thresholds from changes arising from natural variability. Statistical decision theory enables us to integrate this sampling uncertainty into evaluations of tradeoffs between hydropower and ecological losses. It bridges the gap between hypothesis-driven monitoring studies and the plethora of simulation and optimization studies that evaluate multiobjective tradeoffs of various reservoir operation rules. Our two analyses reveal important differences between conventional statistical decision-making approaches and our Bayesian decision tree, which highlights the value of incorporating type II errors and stakeholder valuations of hydropower production and ecosystem services into decisions. Our hypothetical reservoir simulation also elucidates the benefits of incorporating sampling uncertainty in tradeoff analyses even when examining alteration over several decades, as private US hydropower producers are required to get relicensed every 30–50 years (FERC 2016). Moreover, potential impacts at new dam sites with much shorter records make accounting for sampling uncertainty even more critical, especially given recent hydropower growth in the developing world (Zarfl et al. 2015).

We can also assess the decision implications of different expert-elicited prior probabilities. For simplicity, we set the prior probability of alteration (at least one alteration exceeding a threshold) to 0.5 to avoid initially favoring hydropower or ecological interests. In our example, one may argue that $P(A) = 0.75$ is more appropriate since we apply the hypothesis test to two separate AFDC quantiles assumed to be independent. With $P(A) = 0.75$, we obtain an even more compelling recommendation to change to run-of-river hydropower operations ($HP_{ROR}$), as the hydropower loss probability is just 2.7% compared to the ecological loss probability of 69.3%. Conversely, a prior probability of alteration of 10% would cause the decision tree to recommend maintaining $HP_{max}$ instead since the hydropower loss probability (7.7%) is much lower than its ecological counterpart (43.7%). This sensitivity demonstrates the pronounced influence that expert knowledge could have on recommendations arising from our Bayesian decision-tree approach.

Next, one might wonder if pre- and postdam comparisons can be made before dam operations irreversibly harm riverine ecosystems, as several years of pre- and postdam data are needed to compute the exact $U$ statistic of the MWW test (Fay and Proschan 2010) and at least 8 years are necessary for the normal approximation to hold (Mann and Whitney 1947). However, in many cases, excessive hydrologic alteration may signal a reversible decline in ecosystem function, as postdam ecological equilibria can take some time to become established. Perkin et al. (2017) observed fewer native opportunistic species and more nonnative generalist species downstream of a reservoir approximately a decade after its impoundment compared to the first few postdam years. Taylor et al. (2014) detected fewer changes in predam fish assemblages during a 6-year postimpoundment period than during the ensuing 7 years.



J. Water Resour. Plann. Manage., 2020, 146(5): 04020017

In addition, the decision tree can identify sites where excessive flow alteration motivates more intensive ecological monitoring. It can also be inverted to examine the restoration of predam flow conditions. Such screening-level decision tree applications should be conducted only when flow alteration is believed to be the limiting factor constraining riverine ecosystems.

Our statistical decision tree can also be applied in other water resource settings and with other flow alteration indicators and two-sample tests. For instance, applications using different flow indicators could examine excessive changes to daily flow hydrographs, subdaily flows from peaking plants, and durations of high- and low-flow pulses. Even habitat indicators, such as habitat suitability duration curves (e.g., Ceola et al. 2018) could be applied. It can also be applied in basins undergoing concurrent climate or land-use changes by (1) employing hydrologic models calibrated to predam data to estimate what postdam inflows would have been under predam watershed conditions and then (2) comparing this modeled time series to the postdam record just below a dam. Other nonparametric two-sample tests that examine ecologically critical changes in variability could also be assessed (Siegel and Tukey 1960; Marozzi 2013; Kroll et al. 2015). More detailed appraisals of two-sample tests with small samples featuring highly nonnormal flows downstream of hydropower dams are also warranted, as are comparisons of nonparametric tests with Welch's t-test accounting for unequal sample variances. The importance of type II errors also encourages additional work on specifying alternative hypotheses, including ones less reliant upon distributional assumptions given the wide range of dam impacts on flow regimes (McManamay et al. 2012). Finally, this work offers a possible path forward for incorporating the uncertainty surrounding exceedances of long-term alteration thresholds into other multistakeholder water-resources tradeoffs and motivates extensions of the decision tree that consider more than two stakeholders.

## Data Availability Statement

The streamflow data used in our hypothetical case study, along with code for the reservoir simulation model, Mann-Whitney test (including power analysis), and Bayesian decision tree, are available in a Github repository at https://github.com/jshecht/EnvFlows_DecisionTree. Additional code produced during the study is available from the corresponding author by request.

## Acknowledgments

This work builds on a conference paper (see Hecht et al. 2015) and dissertation (Hecht 2017). The Hydro Research Foundation provided a research award to the first author 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 (NSF OIA #0966093) and Vermont Experimental Program to Stimulate Competitive Research (EPSCoR) program (NSF OIA #1556770) also provided in-kind support. Ryan McManamay was supported by the US DOE, Office of Energy Efficiency and Renewable Energy, Water Power Technologies Office. This manuscript has been authored by UT-Battelle, LLC under Contract No. DE-AC05-00OR22725 with the US 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 nonexclusive, 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).

## Supplemental Data

Appendixes S1–S4, Table S1, and Figs. S1–S5 are available online in the ASCE Library (www.ascelibrary.org).

## References

Acreman, M., et al. 2009. "Environmental flows from dams: The water framework directive." *Proc. Inst. Civ. Eng. Eng. Sust.* 162 (1): 13–22. https://doi.org/10.1680/ensu.2009.162.1.13.

Archfield, S. A., P. A. Steeves, J. D. Guthrie, and K. G. Ries III. 2013. "Towards a publicly available, map-based regional software tool to estimate unregulated daily streamflow at ungauged rivers." *Geosci. Model Dev.* 6 (1): 101–115. https://doi.org/10.5194/gmd-6-101-2013.

Bark, R. H. 2013. "Understanding the sources of uncertainty to reduce the risks of undesirable outcomes in large-scale freshwater ecosystem restoration projects: An example from the Murray-Darling Basin, Australia." *Environ. Sci. Policy* 33 (Nov): 97–108. https://doi.org/10.1016/j.envsci.2013.04.010.

Bellera, C. A., M. Julien, and J. A. Hanley. 2010. "Normal approximation to the distributions of the Wilcoxon statistics: Accurate to what N? Graphical insights." *J. Stat. Educ.* 18 (2): 1–17. https://doi.org/10.1080/10691898.2010.11889486.

Carlisle, D. M., T. E. Grantham, K. Eng, and D. M. Wolock. 2017. "Biological relevance of streamflow metrics: Regional and national perspectives." *Freshwater Sci.* 36 (4): 927–940. https://doi.org/10.1086/694913.

Carlisle, D. M., D. M. Wolock, and M. R. Meador. 2011. "Alteration of streamflow magnitudes and potential ecological consequences: A multiregional assessment." *Front. Ecol. Environ.* 9 (5): 264–270. https://doi.org/10.1890/100053.

Ceola, S., A. Pugliese, M. Ventura, G. Galeati, A. Montanari, and A. Castellarin. 2018. "Hydro-power production and fish habitat suitability: Assessing impact and effectiveness of ecological flows at regional scale." *Adv. Water Resour.* 116 (Jun): 29–39. https://doi.org/10.1016/j.advwatres.2018.04.002.

Douglas, E. M., R. M. Vogel, and C. N. Kroll. 2000. "Trends in floods and low flows in the United States: Impact of spatial correlation." *J. Hydrol.* 240 (1–2): 90–105. https://doi.org/10.1016/S0022-1694(00)00336-X.

Downes, B. J., L. A. Barmuta, P. G. Fairweather, D. P. Faith, M. J. Keough, P. S. Lake, B. D. Mapstone, and G. P. Quinn. 2002. *Monitoring ecological impacts: Concepts and practice in flowing waters*. Cambridge, UK: Cambridge University Press.

EPSA (Electricity Power Supply Association). 2017. "Electricity primer: The basics of power and competitive markets." Accessed October 8, 2018. https://www.epsa.org/industry/primer/?fa=wholesaleMarket.

Eriyagama, N., V. Smakhtin, and K. Jinapala. 2016. "The Sri Lanka environmental flow calculator: A science-based tool to support sustainable national water management." *Water Policy* 18 (2): 480–492. https://doi.org/10.2166/wp.2015.158.

Fay, M. P., and M. A. Proschan. 2010. "Wilcoxon-Mann-Whitney or t-test? On assumptions for hypothesis tests and multiple interpretations of decision rules." *Stat. Surv.* 4: 1–39. https://doi.org/10.1214/09-SS051.

FERC (Federal Energy Regulatory Commission). 2016. "Hydropower licensing." Accessed June 11, 2019. https://www.ferc.gov/resources/guides/hydropower/hydro-guide.asp.

Field, S. A., A. J. Tyre, N. Jonzén, J. R. Rhodes, and H. P. Possingham. 2004. "Minimizing the cost of environmental management decisions by optimizing statistical thresholds." *Ecol. Lett.* 7 (8): 669–675. https://doi.org/10.1111/j.1461-0248.2004.00625.x.

FitzHugh, T. W. 2014. “EFCAM: A method for assessing alteration of environmental flow components.” *River Res. Appl.* 30 (7): 825–844. https://doi.org/10.1002/rra.2681.

Gao, Y., R. M. Vogel, C. N. Kroll, N. L. Poff, and J. D. Olden. 2009. “Development of representative indicators of hydrologic alteration.” *J. Hydrol.* 374 (1–2): 136–147. https://doi.org/10.1016/j.jhydrol.2009.06.009.

Gelman, A. 2008. “Objections to Bayesian statistics.” *Bayesian Anal.* 3 (3): 445–449. https://doi.org/10.1214/08-BA318.

Gillespie, B. R., S. Desmet, P. Kay, M. R. Tillotson, and L. E. Brown. 2015. “A critical analysis of regulated river ecosystem responses to managed environmental flows from reservoirs.” *Freshwater Biol.* 60 (2): 410–425. https://doi.org/10.1111/fwb.12506.

Growns, I. 2004. “Monitoring environmental flow regimes: Potential problems with management decision errors and their consequences.” *Ecol. Manage. Restor.* 5 (3): 216–218. https://doi.org/10.1111/j.1442-8903.2004.209-3.x.

Hecht, J. S. 2017. “Making multi-stakeholder water resources decisions with limited streamflow information.” Ph.D. dissertation, Dept. of Civil and Environmental Engineering, Tufts Univ.

Hecht, J. S., R. M. Vogel, and R. A. McManamay. 2015. “Testing hypotheses of hydropower-induced hydrologic alteration beyond ecological thresholds.” Accessed October 8, 2018. http://s36.a2zinc.net/clients/pennwell/HVI2015/Custom/Handout/Speaker53379_Session11881_1.pdf.

Hering, D., et al. 2010. “The European water framework directive at the age of 10: A critical review of the achievements with recommendations for the future.” *Sci. Total Environ.* 408 (19): 4007–4019. https://doi.org/10.1016/j.scitotenv.2010.05.031.

Hobbs, B. F., P. T. Chao, and B. N. Venkatesh. 1997. “Using decision analysis to include climate change in water resources decision making.” *Clim. Change* 37 (1): 177–202. https://doi.org/10.1023/A:1005376622183.

Hochberg, Y., and A. C. Tamhane. 1987. *Multiple comparison procedures*, 5. New York: Wiley.

Jager, H. I. 2014. “Thinking outside the channel: Timing pulse flows to benefit salmon via indirect pathways.” *Ecol. Modell.* 273 (Feb): 117–127. https://doi.org/10.1016/j.ecolmodel.2013.11.007.

Jager, H. I., and B. T. Smith. 2008. “Sustainable reservoir operation: Can we generate hydropower and preserve ecosystem values?” *River Res. Appl.* 24 (3): 340–352. https://doi.org/10.1002/rra.1069.

Jain, S. K. 2015. “Assessment of environmental flow requirements for hydropower projects in India.” *Curr. Sci.* 108 (10): 1815–1825.

Kendy, E., C. Apse, K. Blann, M. P. Smith, and A. Richardson. 2012. “Practical guide to environmental flows for policy and planning.” Accessed October 8, 2018. https://www.conservationgateway.org.

Kennard, M. J., S. J. Mackay, B. J. Pusey, J. D. Olden, and N. Marsh. 2010. “Quantifying uncertainty in estimation of hydrologic metrics for ecohydrological studies.” *River Res. Appl.* 26 (2): 137–156. https://doi.org/10.1002/rra.1249.

Kern, J. D., G. W. Characklis, M. W. Doyle, S. Blumsack, and R. B. Whisnant. 2012. “Influence of deregulated electricity markets on hydropower generation and downstream flow regime.” *J. Water Resour. Plann. Manage.* 138 (4): 342–355. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000183.

Kopf, R. K., C. M. Finlayson, P. Humphries, N. C. Sims, and S. Hladyz. 2015. “Anthropocene baselines: Assessing change and managing biodiversity in human-dominated aquatic ecosystems.” *BioScience* 65 (8): 798–811. https://doi.org/10.1093/biosci/biv092.

Kroll, C. N., K. E. Croteau, and R. M. Vogel. 2015. “Hypothesis tests for hydrologic alteration.” *J. Hydrol.* 530 (Nov): 117–126. https://doi.org/10.1016/j.jhydrol.2015.09.057.

Mann, H. B., and D. R. Whitney. 1947. “On a test of whether one of two random variables is stochastically larger than the others.” *Ann. Math. Stat.* 18 (1): 50–60. https://doi.org/10.1214/aoms/1177730491.

Mapstone, B. D. 1995. “Scalable decision rules for environmental impact studies: Effect size, type I, and type II errors.” *Ecol. Appl.* 5 (2): 401–410. https://doi.org/10.2307/1942031.

Marozzi, M. 2013. “Nonparametric simultaneous tests for location and scale testing: A comparison of several methods.” *Commun. Stat. Simul. Comput.* 42 (6): 1298–1317. https://doi.org/10.1080/03610918.2012.665546.

McManamay, R. A., D. J. Orth, and C. A. Dolloff. 2012. “Revisiting the homogenization of dammed rivers in the southeastern US.” *J. Hydrol.* 424–425 (Mar): 217–237. https://doi.org/10.1016/j.jhydrol.2012.01.003.

Mims, M. C., and J. D. Olden. 2013. “Fish assemblages respond to altered flow regimes via ecological filtering of life history strategies.” *Freshwater Biol.* 58 (1): 50–62. https://doi.org/10.1111/fwb.12037.

Mudge, J. F., L. F. Baker, C. B. Edge, and J. E. Houlahan. 2012a. “Setting an optimal $\alpha$ that minimizes errors in null hypothesis significance tests.” *PLoS One* 7 (2): e32734. https://doi.org/10.1371/journal.pone.0032734.

Mudge, J. F., J. J. Barrett, K. R. Munkittrick, and J. E. Houlahan. 2012b. “Negative consequences of using $\alpha = 0.05$ for environmental monitoring decisions: A case study from a decade of Canada’s environmental effects monitoring program.” *Environ. Sci. Tech.* 46 (17): 9249–9255. https://doi.org/10.1021/es301320n.

Nikghalb, S., A. Shokoohi, V. P. Singh, and R. Yu. 2016. “Ecological regime versus minimum environmental flow: Comparison of results for a river in a semi-Mediterranean region.” *Water Resour. Manage.* 30 (13): 4969–4984. https://doi.org/10.1007/s11269-016-1488-2.

Pastor, A. V., F. Ludwig, H. Biemans, H. Hoff, and P. Kabat. 2014. “Accounting for environmental flow requirements in global water assessments.” *Hydrol. Earth Syst. Sci.* 18 (12): 5041–5059. https://doi.org/10.5194/hess-18-5041-2014.

Perkin, J. S., N. E. Knorp, T. C. Boersig, A. E. Gebhard, L. A. Hix, and T. C. Johnson. 2017. “Life history theory predicts long-term fish assemblage response to stream impoundment.” *Can. J. Fish. Aquat. Sci.* 74 (2): 228–239. https://doi.org/10.1139/cjfas-2015-0593.

Peterman, R. M. 1990. “Statistical power can improve fisheries research and management.” *Can. J. Fish. Aquat. Sci.* 47 (1): 2–15. https://doi.org/10.1139/f90-001.

Poff, N. L., J. D. Olden, D. M. Merritt, and D. M. Pepin. 2007. “Homogenization of regional river dynamics by dams and global biodiversity implications.” *Proc. Nat. Acad. Sci.* 104 (14): 5732–5737. https://doi.org/10.1073/pnas.0609812104.

Poff, N. L., and J. K. H. Zimmerman. 2010. “Ecological responses to altered flow regimes: A literature review to inform the science and management of environmental flows.” *Freshwater Biol.* 55 (1): 194–205. https://doi.org/10.1111/j.1365-2427.2009.02272.x.

R Core Team. 2019. “R: A language and environment for statistical computing.” Accessed June 11, 2019. https://www.R-project.org.

Railsback, S. F., B. C. Harvey, S. J. Kupferberg, M. M. Lang, S. McBain, and H. H. Welsh Jr. 2016. “Modeling potential river management conflicts between frogs and salmonids.” *Can. J. Fish. Aquat. Sci.* 73 (5): 773–784. https://doi.org/10.1139/cjfas-2015-0267.

Reitsch, A. G. 1976. “Selecting alpha by minimizing expected regret: An example.” *Educ. Psychol. Meas.* 36 (3): 675–678. https://doi.org/10.1177/001316447603600313.

Richter, B. D., M. M. Davis, C. Apse, and C. Konrad. 2012. “A presumptive standard for environmental flow protection.” *River Res. Appl.* 28 (8): 1312–1321. https://doi.org/10.1002/rra.1511.

Rosner, A., R. M. Vogel, and P. H. Kirshen. 2014. “A risk-based approach to flood management decisions in a nonstationary world.” *Water Resour. Res.* 50 (3): 1928–1942. https://doi.org/10.1002/2013WR014561.

Shieh, G., S.-L. Jan, and R. H. Randles. 2007. “Power and sample-size determinations for the Wilcoxon signed-rank test.” *J. Stat. Comput. Simul.* 77 (8): 717–724. https://doi.org/10.1080/10629360600635245.

Siegel, S., and J. W. Tukey. 1960. “A nonparametric sum of ranks procedure for relative spread in unpaired samples.” *J. Am. Stat. Assoc.* 55 (291): 429–445. https://doi.org/10.1080/01621459.1960.10482073.

Smakhtin, V., C. Revenga, and P. Döll. 2004. “A pilot global assessment of environmental water requirements and scarcity.” *Water Int.* 29 (3): 307–317. https://doi.org/10.1080/02508060408691785.

Suen, J.-P., and J. W. Eheart. 2006. “Reservoir management to balance ecosystem and human needs: Incorporating the paradigm of the ecological flow regime.” *Water Resour. Res.* 42 (3): W03417. https://doi.org/10.1029/2005WR004314.

Taylor, J. M., T. S. Seilheimer, and W. L. Fisher. 2014. "Downstream fish assemblage response to river impoundment varies with degree of hydrologic alteration." *Hydrobiologia* 728 (1): 23–39. https://doi.org/10.1007/s10750-013-1797-x.

USACE. 2012. "John H. Kerr dam and reservoir master plan." Accessed January 27, 2019. https://www.saw.usace.army.mil/Portals/59/docs/recreation/Master%20Plan/Kerr%20Master%20Plan_FINAL%20JAN%202013.pdf.

USGS. 2019. "National water information system data available on the world wide web (USGS water data for the nation)." Accessed September 10, 2019. http://waterdata.usgs.gov/nwis/uv?02080500.

Vogel, R. M., and N. M. Fennessey. 1994. "Flow-duration curves I: New interpretation and confidence intervals." *J. Water Resour. Plann. Manage.* 120 (4): 485–504. https://doi.org/10.1061/(ASCE)0733-9496(1994)120:4(485).

Vogel, R. M., and N. M. Fennessey. 1995. "Flow duration curves II: A review of applications in water resources planning." *Water Resour. Bull.* 31 (6): 1029–1039. https://doi.org/10.1111/j.1752-1688.1995.tb03419.x.

Vogel, R. M., J. Sieber, S. A. Archfield, M. P. Smith, C. D. Apse, and A. Huber-Lee. 2007. "Relations among storage, yield and instream flow." *Water Resour. Res.* 43 (5): W05403. https://doi.org/10.1029/2006WR005226.

Wald, A. 1939. "Contributions to the theory of statistical estimation and testing hypotheses." *Ann. Math. Stat.* 10 (4): 299–326. https://doi.org/10.1214/aoms/1177732144.

Webb, J. A., S. C. De Little, K. A. Miller, M. J. Stewardson, I. D. Rutherfurd, A. K. Sharpe, L. Patulny, and N. L. Poff. 2015. "A general approach to predicting ecological responses to environmental flows: Making best use of the literature, expert knowledge, and monitoring data." *River Res. Appl.* 31 (4): 505–514. https://doi.org/10.1002/rra.2832.

Williams, J. G. 2018. "Building hydrologic foundations for applications of ELOHA: How long a record should you have?" *River Res. Appl.* 34 (1): 93–98. https://doi.org/10.1002/rra.3143.

Zarfl, C., A. Lumsdon, J. Berlekamp, L. Tydecks, and K. Tockner. 2015. "A global boom in hydropower dam construction." *Aquat. Sci.* 77 (1): 161–170. https://doi.org/10.1007/s00027-014-0377-0.