

Analysis of Water Balance in the Henry's Fork Watershed

Rob Van Kirk, Henry's Fork Foundation
April 2, 2020

Summary

To inform stakeholders and help conservation organizations develop effective strategies for improving streamflow for fisheries, I analyzed surface-water data to investigate water balance (inflow minus outflow) in the Henry's Fork watershed. I conducted the analysis using multi-model inference with Akaike's Information Criterion (AIC), which allows simultaneous testing of multiple working hypotheses and incorporation in parameter estimates of uncertainty across competing models. I used hydrologic data from US Geological Survey (USG; streamflow), US Bureau of Reclamation (USBR; reservoir volume), and Water District 01 (WD01; diversions).

Water leaves the river through diversion for irrigation. Some of that water returns to the river primarily through shallow groundwater, in what are called river "reach gains." The difference between diversion and reach gain is "net diversion." There are three pathways for water to leave the watershed: 1) outflow in the river, 2) consumptive use by crops, and 3) groundwater that travels out of the watershed via the Eastern Snake Plain Aquifer (ESPA). The difference between surface-water inflow and outflow is the net withdrawal of water from the watershed.

The following research questions and answers are presented in great detail in this document.

1. Has surface-water diversion in the Henry's Fork watershed changed since 1978?
Yes. Annual surface-water diversion has decreased by 215,016 ac-ft since 1978.
2. How is reach gain (return flow from irrigation diversion through shallow aquifers) in the Henry's Fork watershed related to diversion, and have reach gains changed since 1978?
Reach gain was strongly correlated with diversion, and annual reach gain has decreased by 210,569 ac-ft since 1978.
3. Has net diversion of surface water in the Henry's Fork watershed changed since 1978?
No. Diversion and reach gain have both decreased by the same amount since 1978.
4. Has surface-water inflow minus outflow in the Henry's Fork changed since 1978?
No.
5. Does surface-water inflow minus outflow measure the same quantity as net diversion?
Yes. It measures net withdrawal of water from the Henry's Fork basin.

Reduction in both diversion and reach gain has been due primarily to replacement of flood irrigation with more efficient sprinkler irrigation. Even though both quantities have decreased by the same amount, loss of return flows from flood irrigation has increased reliance on draft of Island Park Reservoir to meet irrigation demand and decreased input of cool groundwater to the lower Henry's Fork during summer. Furthermore, increased irrigation efficiency is well known to increase consumptive use of water by crops. Thus, these results suggest that irrigation efficiency improvements must be carefully considered to avoid unintended negative consequences to fisheries and that properly timed managed aquifer recharge holds promise for enhancing streamflow and water temperature for fish in the lower Henry's Fork.

Introduction

Irrigated agriculture in the upper Snake River basin is a major contributor to the economy of the state of Idaho and accounts for the vast majority of water withdrawn from the Snake River, its tributaries, and local and regional aquifers. Disagreement on management, administration, and policy among water users and other stakeholders has a long history in the basin, but the number and scope of legal and administrative actions governing the water resources of the Upper Snake River basin, including the Eastern Snake Plain Aquifer (ESPA), has increased in the past 40 years. Some argue that the Swan Falls Settlement Agreement of 1984 is the most important of these, setting in motion the Snake River Basin Adjudication, among other actions. Establishment of conjunctive management rules (in reality, conjunctive *administration* rules) in the mid-1990s was also important, in turn resulting in numerous water calls by senior surface users against junior groundwater users. The outcome of one of those water calls was the 2015 settlement agreement between groundwater and surface water users; another settlement involving cities on the ESPA followed. Numerous planning documents and policies have been developed by the Idaho Water Resource Board in recent years, some resulting in substantial funding for new water projects such as the state's managed aquifer recharge program.

All of these actions have occurred against a backdrop of variability in water supply over time scales ranging from years to decades and changes in irrigation practices. Scientific analysis and modeling of these hydrologic changes has informed legal and policy actions, but the science itself has been a source of disagreement, and many stakeholders do not receive unbiased information. To provide information to stakeholders, as well as help conservation organizations develop effective strategies and partnerships for improving streamflow at times and places most important for fisheries, I have analyzed surface-water data to investigate temporal trends in water balance in the Henry's Fork watershed.

Research Questions

1. Has surface-water diversion in the Henry's Fork watershed changed since 1978?
2. How is reach gain in the Henry's Fork watershed related to diversion, and has reach gain changed since 1978?
3. Has net diversion (diversion minus gain) of surface water in the Henry's Fork watershed changed since 1978?
4. Has mass balance (inflow minus outflow) of surface water in the Henry's Fork watershed changed since 1978)?
5. Does surface-water inflow minus outflow measure the same quantity as net diversion?

Methods

Statistical Modeling using Akaike's Information Criterion

Rather than using traditional hypothesis testing, I used model selection and multi-model inference with Akaike's Information Criterion (AIC) for analysis of annual time series data. The theory of this type of analysis dates back to the 1970s but has become widely used only in the past 15 years, due in part to availability of sufficient computational power and software. The basic method is to propose a set of candidate models, rank them according to a particular information criterion (in this case AIC; there are several others out there), and then use a

measure of relative evidence for the models in the candidate set to calculate a final model that is a weighted average of all models in the set (Burnham and Anderson 2002; Anderson 2008; Claeskens and Hjort 2008). I and my scientific colleagues have used this method in recent publications in groundwater hydrology (Boggs et al. 2014b), fish ecology (Kuzniar et al. 2017), and limnology (McLaren et al. 2019).

The AIC is a relatively easily understood information criterion that has firm mathematical basis in theory of both statistical likelihoods and information. The basic AIC formula is

$$AIC = -2 \log(\mathcal{L}) + 2p, \quad (1)$$

where $\mathcal{L}$ is the statistical likelihood of a fitted model (essentially the product of the probability density of all data points used to fit the model), $\log$ is the natural logarithm, and p is the number of parameters fitted in the model, including all structural parameters such as means, slopes, and intercepts, and any and all parameters describing the probability structure of the model such as variances, covariances, and autocorrelation coefficients. Lower AIC scores indicate “better” models in the sense that the data provides more evidence for that particular model among all in the candidate set.

The $2p$ term “penalizes” models for the number of parameters included. The likelihood term in the AIC formula can be made as small as desired by simply adding more parameters to the model, analogous to the way in which R^2 in linear models can be made as close to 1 as desired by adding more parameters to the model. In teaching statistics, I always assigned students an exercise to generate a random set of 10 observations from any distribution they wanted and then fit a 9th order polynomial to that random data. Of course the polynomial fit perfectly every time, because it fit 10 parameters to 10 observations. However, the model was of no use in predicting the behavior of the population of data from which the sample was drawn; a different sample would produce a completely different set of coefficients. Overfitting results in large standard errors around parameter estimates because some parameters, especially if they are correlated with others, contribute very little to model likelihood (“fit”) but use up degrees of freedom. Under-fitting may miss important structural aspects of the data and result in biased parameter estimates, even if those estimates are highly precise. A simple example of the latter occurs when the data show a strong linear trend but are described with a model that includes only the intercept (the mean). This mean is very precisely estimated, but it is biased well away from the actual nonzero slope. The AIC weighs bias against precision and tends to favor models with fewer parameters than might be selected based solely on statistical significance of parameters estimates, especially in data sets with relatively small sample sizes—on the order of 10 to 100. When dealing with small sample sizes, a modification of AIC known as AICc (AIC with small-sample correction) is used. The AICc includes an additional term that increases the overfitting “penalty” when the number of fitted parameters becomes large relative to the sample size.

The absolute AICc scores are not important in the model selection process, only the differences in AICc among models. The “best” model out of the candidate set is the one with lowest AICc score, and then all other models are ranked in ascending order of $\Delta AICc$ with respect to this “best” model. These $\Delta AICc$ values can be converted into model weights w_i via the formula

$$w_i = \frac{\exp\left(-\frac{1}{2}\exp(\Delta AICc_i)\right)}{\sum_j \exp\left(-\frac{1}{2}\exp(\Delta AICc_j)\right)}, \quad (2)$$

where the sum in the denominator is taken over all models in the candidate set. This normalization produces a set of weights that sum to 1. The weight of a given model thus decreases exponentially with $\Delta AICc$. The weights give relative evidence for or against the respective models, given the data and the other models in the candidate set. Weighted averages of model parameters, fitted values, and covariances yield an evidence-based final model that reflects the relative evidence for all models in the candidate set and parameter estimates that have optimal balance between bias and standard error. Model weights can also be used to identify particular model components that are more strongly or less strongly supported by the data. For example, in a set of models, some of which contain a linear trend term, the sum of weights over all models containing that term is relative evidence for existence of a linear trend in the population from which the data sample was drawn.

Model averaging using AICc offers numerous advantages over traditional statistical hypothesis testing. First, it allows simultaneous evaluation of a number of scientific hypotheses, each represent by a particular model or subset of models in the candidate set (e.g., all models with a linear trend term contain the hypothesis that the data show a linear trend). This is in contrast to statistical hypothesis testing, where only one null hypothesis can be tested at a time by comparing a given model with that to a nested model with one or more parameters removed. Except in completely balanced experimental designs, the results of the null hypothesis tests depend on the order in which null hypotheses are tested i.e., the order in which parameters are added or removed. The AICc does not depend on the ordering of model parameters, so the effect of potential confounding variables can be assessed in all combinations with variables of interest, without first testing for significance of the confounding variable and then electing to include or not include it before testing significance of the variables of interest. Second, the AICc can be used to compare models that are not nested within each other, a requirement for null hypothesis testing. A model with linear trend can be compared against a model with exponential trend, for example. Third, the AICc can compare models with different variance and distributional structures, for example comparing models with lognormal variance against analogous models with normal variance or time series models with different orders of autocorrelation. This cannot be done with null hypothesis testing. In this application of AICc, adherence to distributional assumptions can be evaluated based on evidence from the data concurrent with the structural models.

One drawback of the AICc method is that results depend on the particular models in the candidate set. Obviously, a model not in the set can't be evaluated, and this fact can be exploited to bias the results toward certain outcomes. The set of candidate models should be chosen based on reasonable and parsimonious descriptions of the data that are grounded in knowledge of the system being studied and on the underlying physics or biology. Simulation modeling suggests that once a given set of parameters and model structures is determined, all possible combinations of the parameters and structures should be used in order to obtain the most appropriate set of model weights and averaged model (Doherty et al. 2012). The candidate set should always

include a “null” model, one that tests the hypothesis that none of the potential predictor variables has any effect on the response. In most cases, the null model simply describes the data with a single structural parameter, namely the mean. The null model may appear with different variance/covariance models.

After ranking the models by $\Delta AICc$, the additional step of removing redundant or “pretending” models must be taken to correctly calculate model weights. Redundant models occur when a parameter with no predictive power occurs in a particular model that otherwise has reasonable model weight. In this case, the addition of the poor predictor is equivalent to adding a parameter whose value is 0. The addition of that parameter to a particular model does not increase the model likelihood but increases AIC by 2 (or AICc by something slightly larger than two), the value of the $2p$ term when p is increased by one. Thus, the “pretending” model may still look very good in relation to other models, but the pretending model is actually just redundant with the first one and should be removed so that the remaining models receive appropriate weights.

Specific Candidate Models Used in this Analysis

All of the data analyzed here occur in annual time series, which runs from 1978 to 2019 (sample size = 42), coinciding with the set of irrigation years over which the modern Water District 01 (WD01) accounting model has been used. Given that the data occur in time series, all models were fit in the framework of autoregressive (AR) models with exogenous predictors.

Structural models tested

- Null model: data described by a single mean (one structural parameter).
- Piecewise constant: data described by two means, one for each of two distinct time periods (two structural parameters describing the means plus a third to determine the time period breakpoint).
- Linear trend (two structural parameters).
- Piecewise trend: data described by linear trend over one time period and constant mean over the other (three structural parameters plus a fourth to determine the time period breakpoint).
- Quadratic (three structural parameters).

Potential covariate

- Natural flow (water supply) was used as a potential covariate because diversion is generally greater in years of greater water supply. Incorporation of water supply as a covariate removes the confounding effect of short-term variability in water supply on actual long-term trends (one additional parameter).

Autocorrelation models

- First-order autocorrelation (AR1, one additional parameter).

Distributional models

- Normal distribution (one additional parameter).
- Lognormal distribution (one additional parameter).

Unless there was a clear reason not to include every possible model combination (e.g., model diagnostics clearly showed normally distributed residuals), I used the all-possible-combinations approach. Given five structural models and two choices for each of the other components, this gave a maximum of 40 possible models. In most cases, there was no need to test models with both normal and lognormal distributions, so generally 20 models were fit. After removing redundant models, most AICc results were based on fewer than 15 models, and among those models, generally fewer than 10 accounted for 99.9% of total model weight. With a sample size of 42 and the relatively parsimonious and simple model structures used in the candidate sets, the data supported the number of models and parameters fit with little chance of spurious results.

Data Compilation

Total surface-water diversion for each of irrigation years 1978-2019 was calculated from daily WD01 diversion data, as reported in the accounting database. This excludes diversions in headwater areas such as Teton Valley and Island Park. Total natural flow (basin inflow) is defined as the sum of natural flow in Henry's Fork at Ashton, Fall River at Chester, and Teton River at St. Anthony, each calculated using the formula

$$\text{natural flow} = \text{regulated flow} + \text{diversions} - \text{return flow} + \Delta \text{storage} + \text{reservoir evap.} \quad (3)$$

For these river reaches, I set the return flow term to zero because there is very little in the way of irrigation return flow either through surface returns or groundwater pathways upstream of Ashton, Chester, and the Crosscut Canal, respectively. There are two differences between my natural flow calculation and that of WD01. First, I count direct precipitation on reservoir surfaces in the evaporation term; if precipitation exceeds evaporation, this term is negative. Second, when applied to river reaches downstream of Ashton, Chester, and the Crosscut Canal, I assume that reach gain is equivalent to irrigation return flow and thus subtract it from natural flow. Regardless of source, reach gains are natural flow from an accounting standpoint but are not "natural" flow from the standpoint of separating basin supply from water recycled in the lower watershed through the irrigation system.

I define total lower-watershed reach gain using the standard WD01 formula

$$\text{reach gain} = \text{outflow} - \text{inflow} + \text{diversions} + \Delta \text{storage} + \text{reservoir evap}, \quad (4)$$

where the storage and reservoir evaporation terms are zero in the lower watershed, outflow is streamflow in Henry's Fork at Rexburg and inflow is sum of regulated flow in Henry's Fork at Ashton, Fall River at Chester, and Teton River at St. Anthony. To avoid double-counting water diverted into the Crosscut Canal that is diverted again in the Teton River, I have subtracted Crosscut injection from total diversion.

Again, assuming that lower-watershed reach gain is equivalent to return flow, equations (3) and (4), applied to the whole watershed, yield

$$\begin{aligned} & \text{basin inflow} - \text{basin outflow} \\ & = \text{diversion} - \text{reach gain} + \Delta \text{storage} + \text{reservoir evap.} \end{aligned} \quad (5)$$

This equation is valid on any given day. If reservoir storage is the same at the beginning/end of each irrigation year, then at the annual time scale, $\Delta storage = 0$, and the difference between inflow and outflow is equal to the difference between diversion and return flow, less the net difference between direct precipitation on reservoir surfaces and reservoir evaporation. Small differences between these annual quantities are expected because of year-to-year differences in reservoir storage at the beginning/end of each irrigation year. Analysis of inflow and outflow by themselves was done with water years, whereas all of the other analysis was done with irrigation years.

Results

Because AICc tables are large and dense in information, I have included only one in this document, to provide an example of how they are used in multi-model inference. Otherwise, I provide a summary of information from the AICc analysis and the averaged models. I have also provided some graphical examples and accompanying statistical explanations of how candidate models are averaged by the AICc weights and how autocorrelation and inflow as a covariate affect model fit and interpretation. These examples are intended to provide insight into both the multi-model inference approach and to the hydrology itself.

1. Has surface-water diversion in the Henry's Fork watershed changed since 1978?

Yes. Annual surface-water diversion has decreased by 215,016 ac-ft since 1978.

Models that included terms quantifying decrease in diversion accounted for essentially 100% of the model weight (Table 1). Evidence was strongest for a stepwise reduction in diversion beginning in 2001; 95.7% of the model weight was accounted for by piecewise constant and piecewise trend models. Evidence was strongest for a model consisting of two means—one for 1978-2000 and another for 2001-2019. Six versions of this model, each with different distributional and covariate parameterizations, accounted for 75.5% of model weight. Annual diversion averaged 1,084,551 ac-ft over the earlier time period and 869,535 ac-ft over the recent time period. Correlation between diversion and basin inflow was 0.46, and the effect of inflow appeared in models accounting for 50.2% of total weight. Temporal autocorrelation was moderate, accounting for 83.4% of model weight and averaging 0.39 over all models. Models with lognormal distribution of residuals accounted for 92.6% of model weight, indicating right skewness of raw residuals. Note the importance of including covariates and different distributional parameterizations in the models. The “naïve” model—one with linear trend, no covariates, no autocorrelation, and normal distribution of residuals—does not even appear in Table 1 because its weight was essentially zero.

Table 1. AICc table for analysis of Henry's Fork surface-water diversion, after removal of redundant models. Models with weight less than 0.002 are not shown.

Structural model	Inflow included	AR term	Log transform	p	$\log(\mathcal{L})$	AICc	ΔAICc	weight	Cum. weight
Piecewise constant	NO	YES	YES	5	52.32	-92.97	0.00	0.432	0.432
Piecewise constant	YES	YES	YES	6	53.08	-91.75	1.21	0.236	0.668
Piecewise trend	YES	NO	YES	6	52.25	-90.10	2.86	0.103	0.772
Piecewise trend	YES	YES	YES	7	53.49	-89.69	3.28	0.084	0.856
Piecewise constant	NO	YES	NO	5	49.95	-88.24	4.72	0.041	0.896
Piecewise constant	YES	NO	YES	5	49.32	-86.97	6.00	0.022	0.918
Piecewise constant	YES	YES	NO	6	50.47	-86.53	6.43	0.017	0.935
Linear trend	YES	YES	YES	5	49.03	-86.40	6.56	0.016	0.952
Linear trend	YES	NO	YES	4	47.54	-86.00	6.96	0.013	0.965
Piecewise trend	NO	NO	YES	5	48.74	-85.81	7.16	0.012	0.977
Piecewise constant	NO	NO	YES	4	46.82	-84.56	8.41	0.007	0.983
Linear trend	NO	YES	NO	4	46.76	-84.43	8.53	0.006	0.989
Piecewise trend	YES	NO	NO	6	48.60	-82.80	10.16	0.003	0.992
Linear trend	YES	YES	NO	5	47.13	-82.60	10.37	0.002	0.995

The averaged model showed continuous decline in diversion since 1978, with a large decrease in 2001 (Figure 2). The nature of the decrease reflects a combination of change in irrigation practices throughout the 1980s and 1990s and lower water supply since 2000.

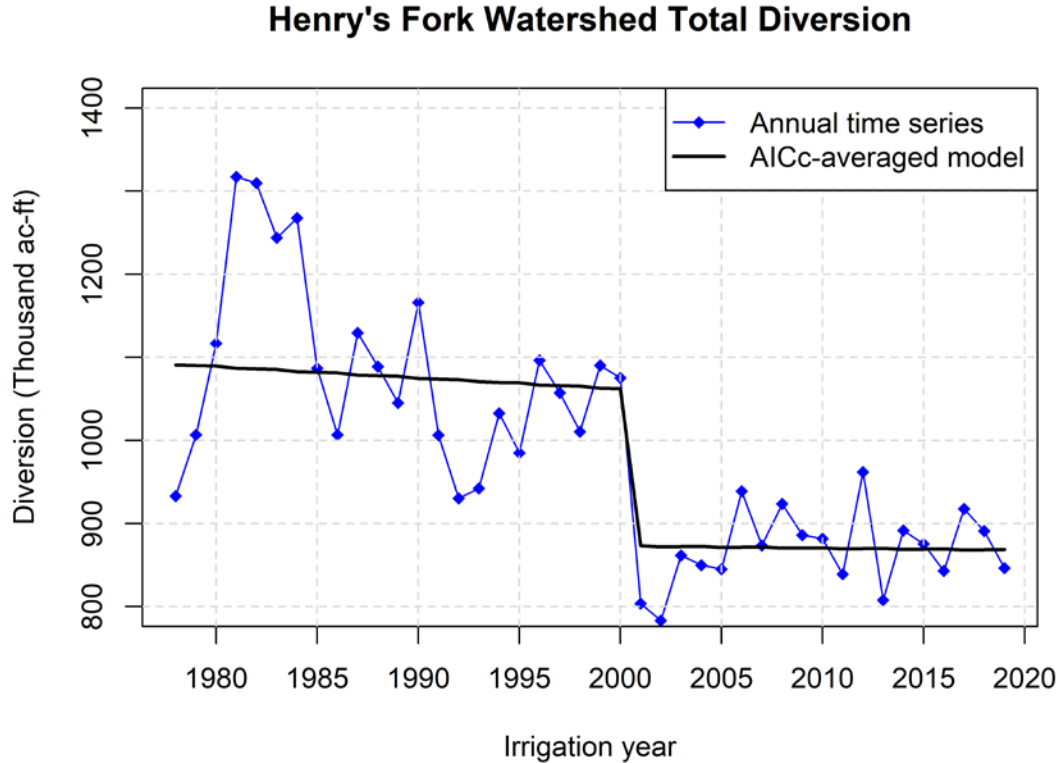


Figure 2. Trend in surface-water diversion in the Henry's Fork watershed, 1978-2019.

2. How is reach gain in the Henry's Fork watershed related to diversion, and have reach gains changed since 1978?

Reach gain was strongly correlated with diversion, and annual reach gain has decreased by 210,569 ac-ft since 1978.

Correlation between reach gain and diversion was 0.83 (Figure 3); correspondingly, evidence for decrease in reach gains since 1978 was strong. Models with some sort of decreasing trend accounted for over 75% of the model weight and all incorporated appropriate logarithmic data transformation. This resulted in “linear” trends in the transformed data appearing as exponential trends in the original data. Not surprisingly, the top model was the same as that for diversion—a piecewise constant model with one mean for the 1978-2000 period and another for the 2001-2019 period (Figure 4). Reach gain averaged 247,677 ac-ft/year from 1978 to 2000 and 34,108 ac-ft since then. Differences in results between diversion and reach gain all reflect proportionally greater influence of groundwater on reach gain (Table 2).

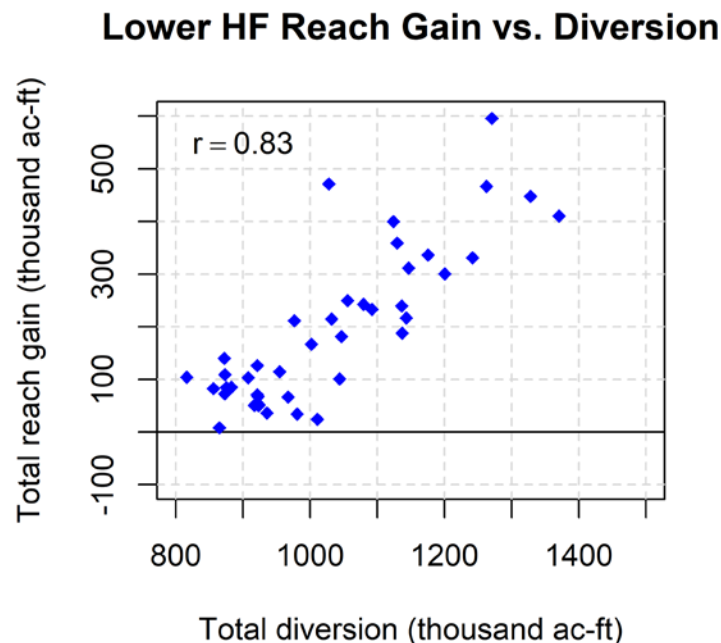


Figure 3. Relationship between total watershed diversion and lower-watershed reach gain.

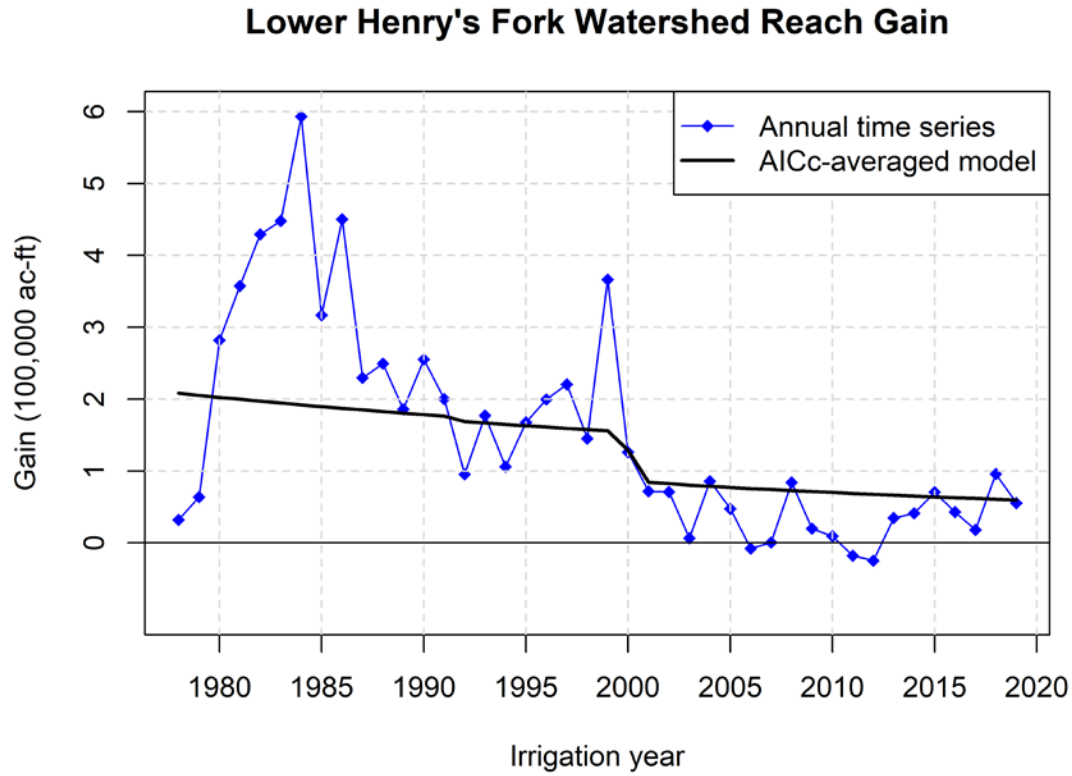


Figure 4. Trend in surface-water diversion in the Henry’s Fork watershed, 1978-2019.

Table 2. Comparison of AICc analysis results for annual diversion and annual reach gain in the Henry’s Fork watershed, 1978-2019.

Model characteristic	Diversion	Reach gain
Weight of models with step change around the year 2000	95.7%	40.1%
Weight of models with continuous trend	3.7%	31.4%
Weight of models with no temporal trend (null models)	0%	24.3%
Weight of models including basin inflow as a covariate	50.2%	42.6%
Model-averaged autocorrelation (1 st -order AR term)	0.39	0.57
Mean decrease between 1978-2000 and 2001-2019 (ac-ft)	215,016	210,569

On the face of it, the AICc-averaged model for trend in reach gain does not look to “fit” the data well (Figure 4). The linear trend prior to 2001 looks as if it should be steeper, and the 2001-2019 mean looks as if it should be higher. The reason for apparent poor fit to the raw data is because of high autocorrelation and moderate evidence (24.3% of model weight) for null models. The null models pull the averaged model toward that of a single mean for the whole time series. The two null models, which differed from each other only by presence of inflow as a covariate, have the form of the basic AR1 time series model:

$$y_t = \mu + \phi_1(y_{t-1} - \mu) + \varepsilon, \quad (6)$$

where y_t is the state of the system at time t , μ is the mean state of the system, $0 < \phi_1 < 1$ is the first-order autocorrelation (autoregression) coefficient, and ε is independent, random error. The structural part of this model is simply a constant—there is no systematic temporal change in the state of the system. At any given time, the difference between the current state and the mean is just a fraction ϕ_1 of the difference between the state and the mean at the last time step, plus some random noise. Averaged autocorrelation in the two null models was $\phi_1 = 0.78$. Exponential decay at this rate has a half-life of 2.8 years, so the effect of the current state persists several years into the future. Moderate model weight for the basic AR1 time series model provides some evidence that the apparent long-term trend in reach gains could be due to short-term trends caused by strong persistence of hydrologic conditions from one year to the next. Random large-magnitude changes “reset” the starting point of these short-term trends, as occurred in the early 1980s and 1999 (Figure 5). Beyond the last observation in the time series, the difference between the system state and the mean decays exponentially with rate ϕ_1 to 0, meaning that in absence of any prior information, the predicted system state in the future is just the constant, μ .

All of the AICc-averaged models illustrated as fits to the time series in this document include only the structural model components, not the additional AR1 terms. This is done for the reason explained above: in absence of any current information about the system (i.e., well into the future), the autoregressive component of the model decays exponentially, leaving only the structural part of the model as the future projection (the random error term has mean 0). Thus, although the full autoregressive model may fit the data quite well, the structural part alone may appear to fit relatively poorly (Figure 6). Inclusion of appropriate autoregressive terms in time-series models is critical to meeting distributional assumptions necessary for unbiased estimates of structural model parameters and their uncertainty. In general, failure to include autocorrelation overestimates the magnitude of structural parameters and underestimates standard errors because a substantial fraction of information about the system comes from the system itself rather than from external factors. From a statistical standpoint, redundant internal information in time series effectively reduces the number of independent data points (sample size), which in turn increases standard errors (recall that standard error is inversely proportional to the square root of sample size) and hence uncertainty about the effect of external factors on the system. The large effect of autocorrelation on multi-model inference about temporal trends in reach gain is a good illustration of why multi-model inference is a much better analysis tool than hypothesis testing and why simple trend analysis can lead to erroneous conclusions (Figure 6).

Lower Henry's Fork Watershed Reach Gain

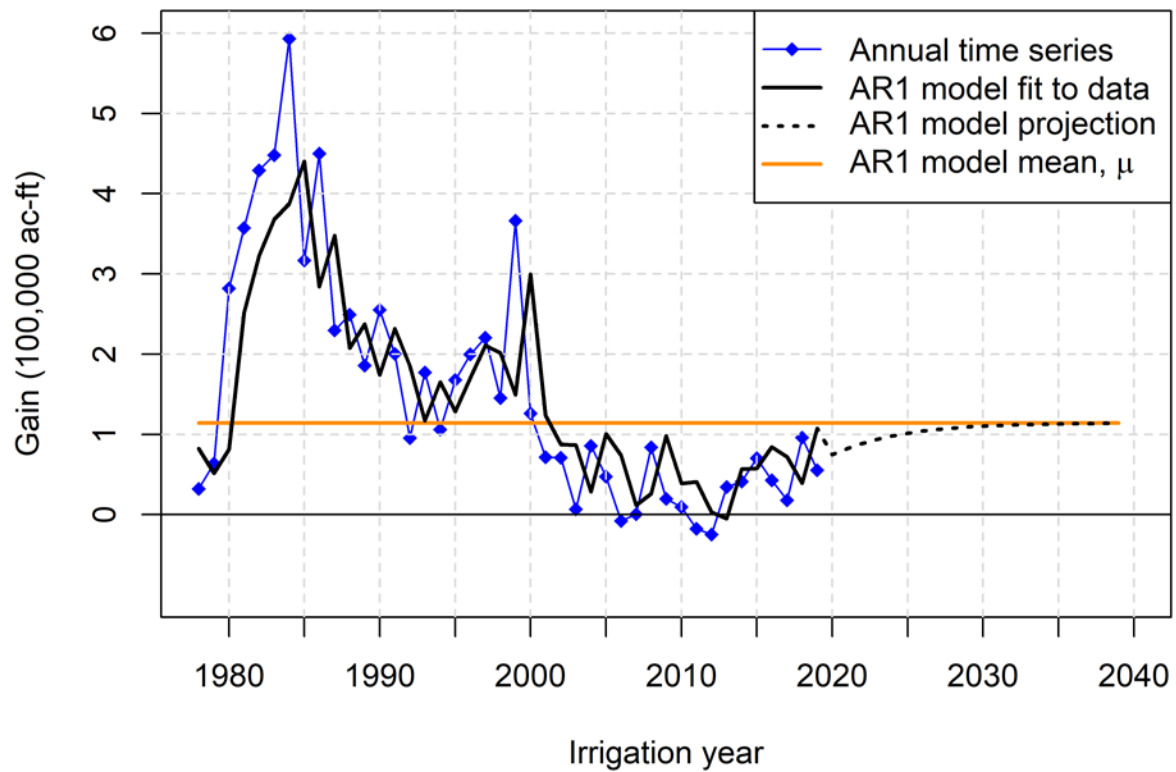


Figure 5. First-order autoregressive (AR1) model with constant mean, fit to the reach gain time series and projected 20 years beyond the last observation. The model shown here is a weighted mean of the two null models, which together accounted for around 24% of total model weight.

Lower Henry's Fork Watershed Reach Gain

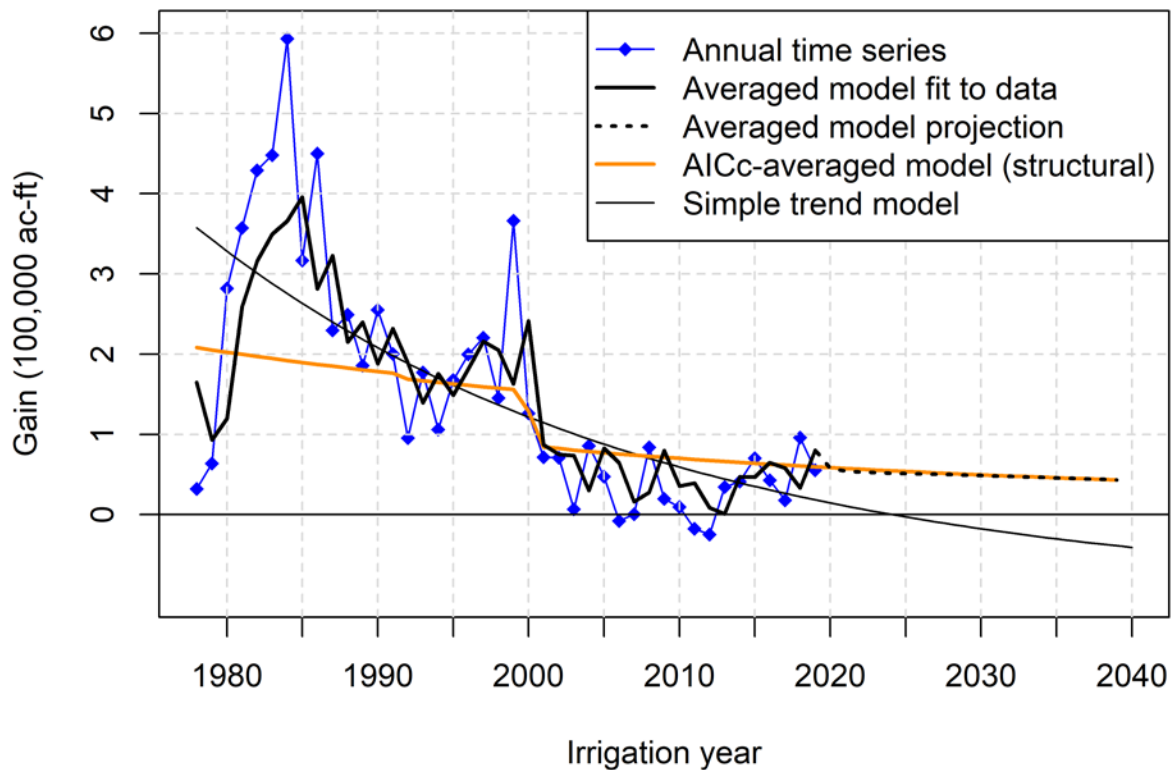


Figure 6. Averaged model, including autocorrelation, fit to the reach gain time series and projected 20 years beyond the last observation. The AICc-averaged model shown in Figure 4 is shown here in orange, including its projection 20 years beyond the last observation. Also shown is the simple trend model, which does not include either inflow as a covariate or autocorrelation. It appears to “fit the data” better but had a model weight of 0.03%.

3. Has net diversion of surface water in the Henry's Fork watershed changed since 1978?

No. Diversion and reach gain have both decreased by the same amount since 1978.

For diversion minus reach gain, i.e., “net diversion”, only two models had nonzero weight, and both of those were models of a single constant mean. One of these used a logarithmic transformation, and neither included autocorrelation or dependence on basin inflow. These latter observations indicate that year-to-year variability in net diversion is random and independent. Mean net diversion was 834,862 ac-ft per year, with an interannual coefficient of variation of 10% (Figure 7). From above, diversion decreased by 215,016 ac-ft per year between pre- and post-2001 time periods, and reach gain dropped by 210,569 between pre- and post-2000 time

periods. Statistically, the difference between these two quantities is zero. Roughly, then, diversion and reach gain have both decreased by about 213,000 ac-ft since 2000.

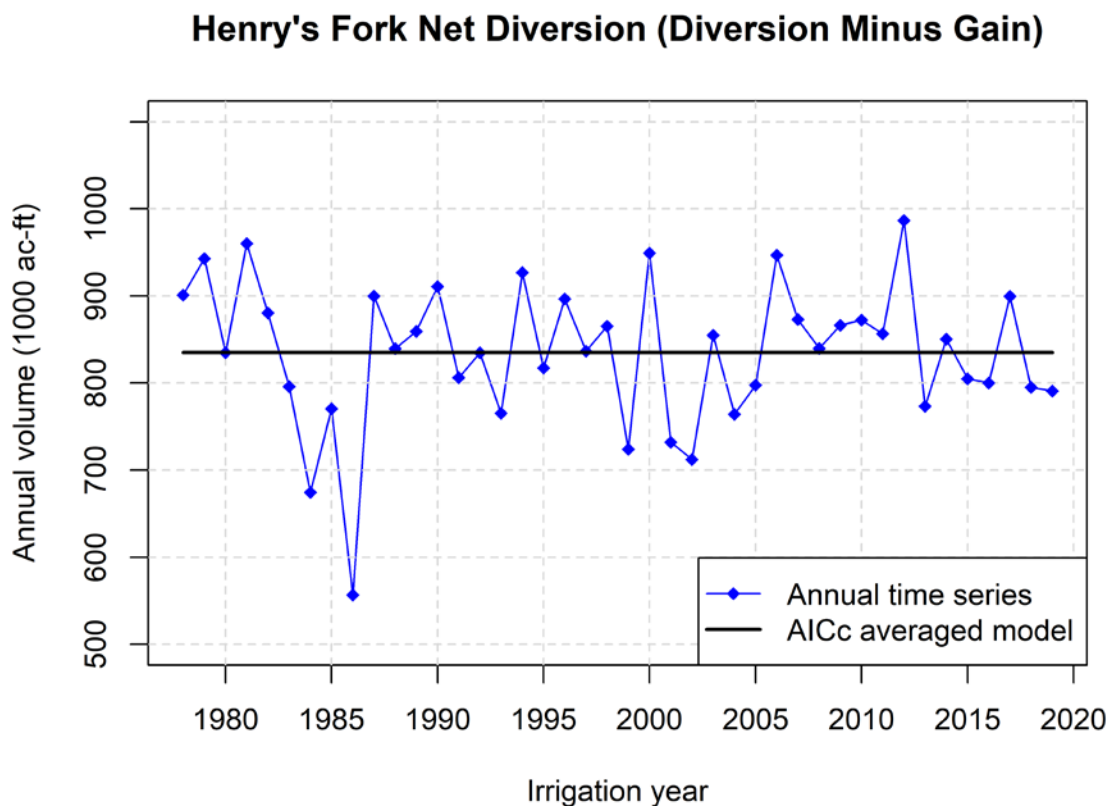


Figure 7. Henry's Fork net diversion, 1978-2019.

4. Has mass balance (inflow minus outflow) of surface water in the Henry's Fork watershed changed since 1978)?

No.

Basin inflow and outflow were highly correlated, with strongest correlation in logarithms (Figure 8). Neither inflow nor outflow on their own showed any strong temporal trend or pattern, although there was slight evidence for a stepwise decrease in both around the year 2000 (Figure 9). Inflow minus outflow showed no correlation with inflow, and there was no evidence for any change in inflow minus outflow since 1978 (Figure 10).

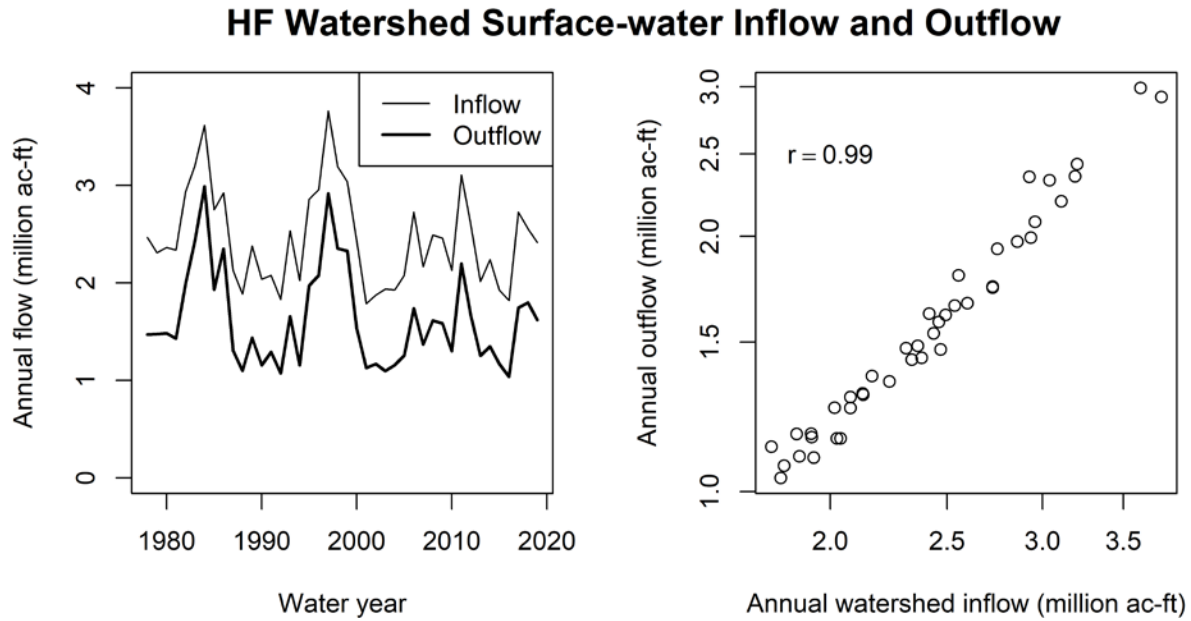


Figure 8. Relationship between Henry's Fork basin inflow and outflow, as a time series (left) and scatterplot (right). The right-hand panel has a log-log scale.

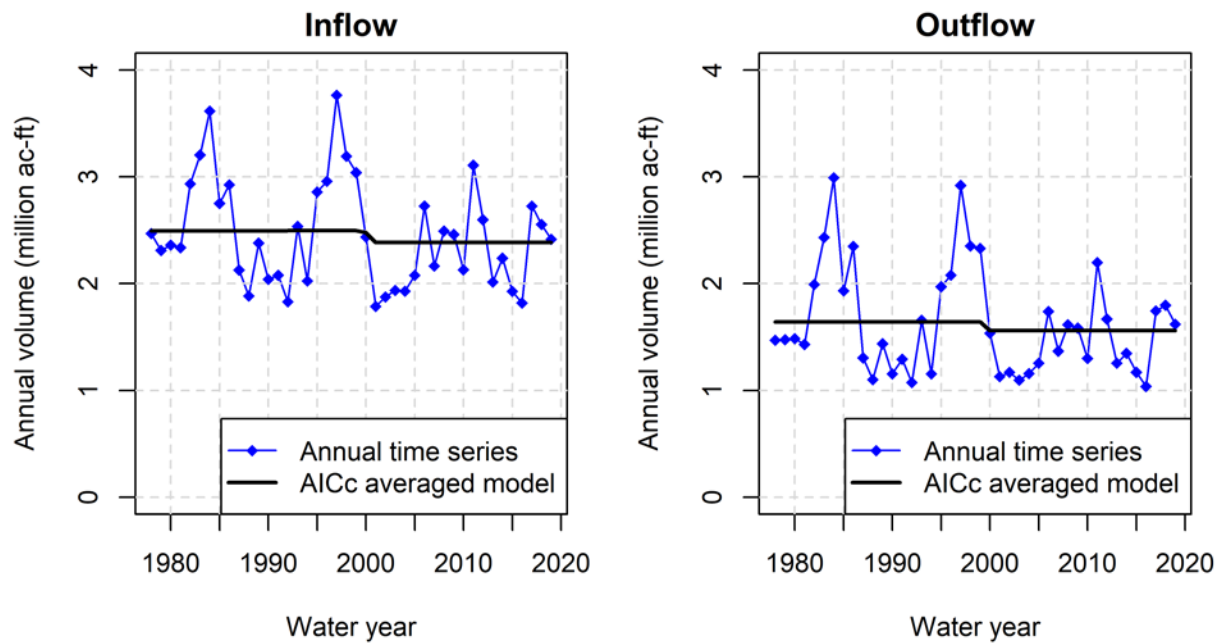


Figure 9. Henry's Fork basin inflow (left) and outflow (right).

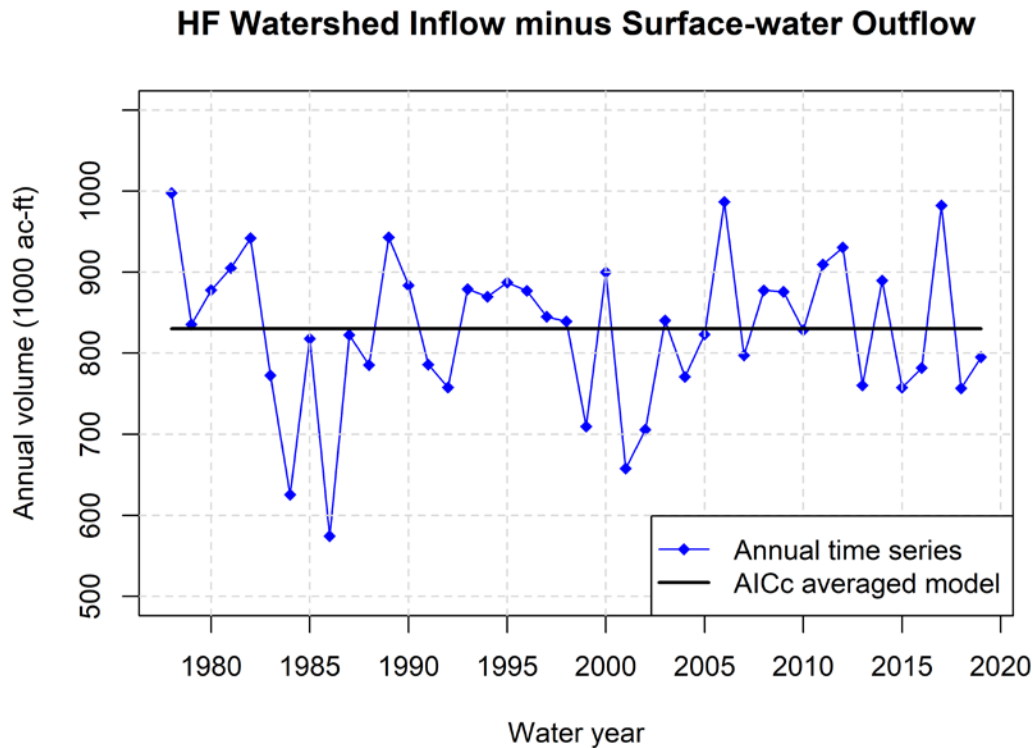


Figure 10. Henry's Fork watershed inflow minus outflow.

5. Does surface-water inflow minus outflow measure the same quantity as net diversion?

Yes. It measures net withdrawal of water from the Henry's Fork basin.

Net diversion and surface-water inflow minus outflow were highly correlated (Figure 11). Among several models that described the relationship between inflow minus outflow and net diversion (diversion minus reach gain), 100% of the model weight was accounted for by a linear relationship with zero intercept. The slope of the fitted line, with 95% confidence interval, was 0.993 ± 0.010 . Thus, the relationship between inflow minus outflow and diversion minus reach gain is statistically equivalent to a 1:1 linear relationship (Figure 12). Basin inflow minus outflow averaged 829,931 ac-ft/year, and net diversion averaged 834,862 ac-ft/year, a difference of -4,931 ac-ft per year. From equation (5), the difference should be equal to $\Delta storage + reservoir\ evap$. Over the period 1978-2019, my calculations give $\Delta storage + reservoir\ evap = -6,001$ ac-ft/year, a difference of only 1,070 ac-ft/year or 0.13% of net diversion. This difference includes effect of inflow from tributaries in the lower watershed that is not included in my calculation of basin inflow and to injection of water from the exchange wells. Within a water year, the difference between inflow minus outflow and net diversion is the combined $\Delta storage + reservoir\ evap$ term, which is positive during storage season and negative during irrigation season (Figure 13).

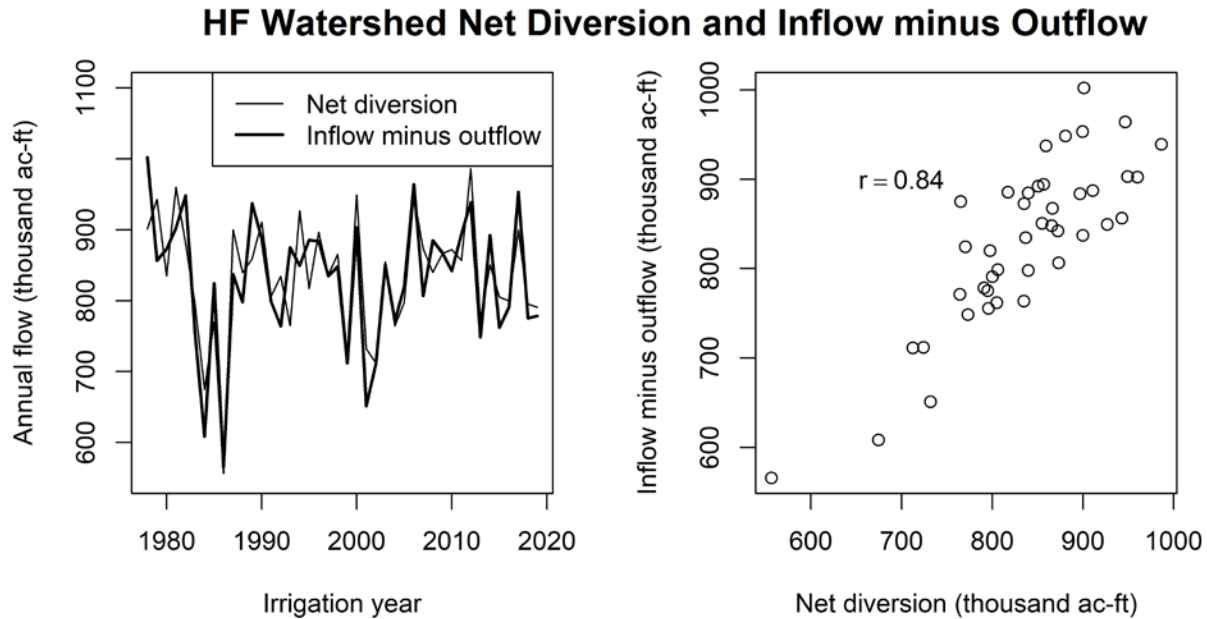


Figure 11. Relationship between Henry's Fork basin net diversion (diversion minus reach gain) and surface-water mass balance (inflow minus outflow), as a time series (left) and scatterplot (right). The right-hand panel has a log-log scale.

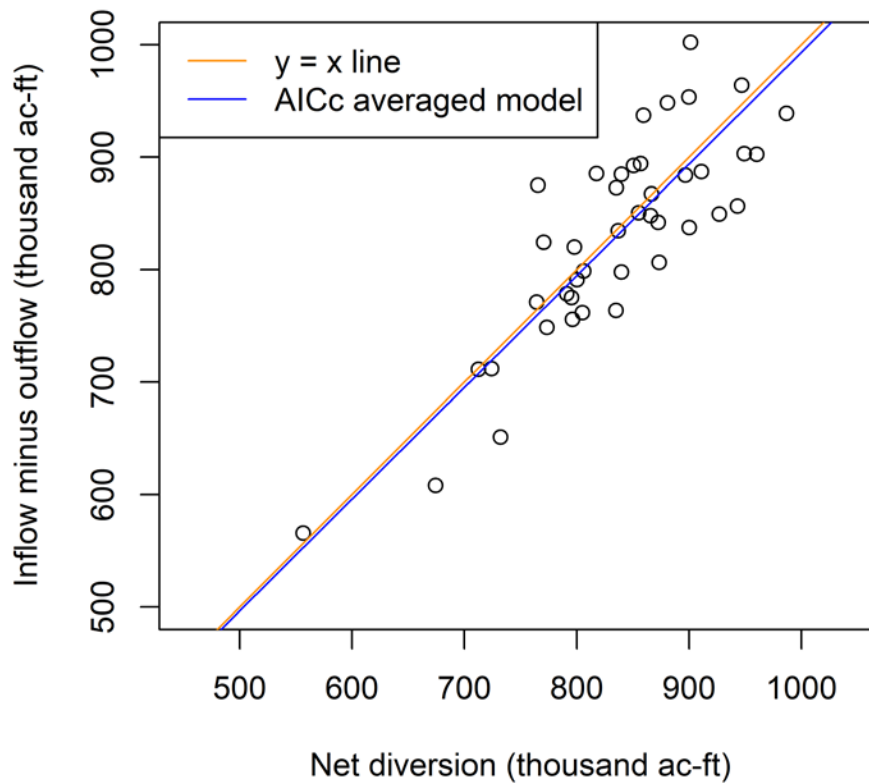


Figure 12. Statistical relationship between inflow minus outflow and net diversion in the Henry's Fork watershed, 1978-2019.

Henry's Fork Surface Water Balance

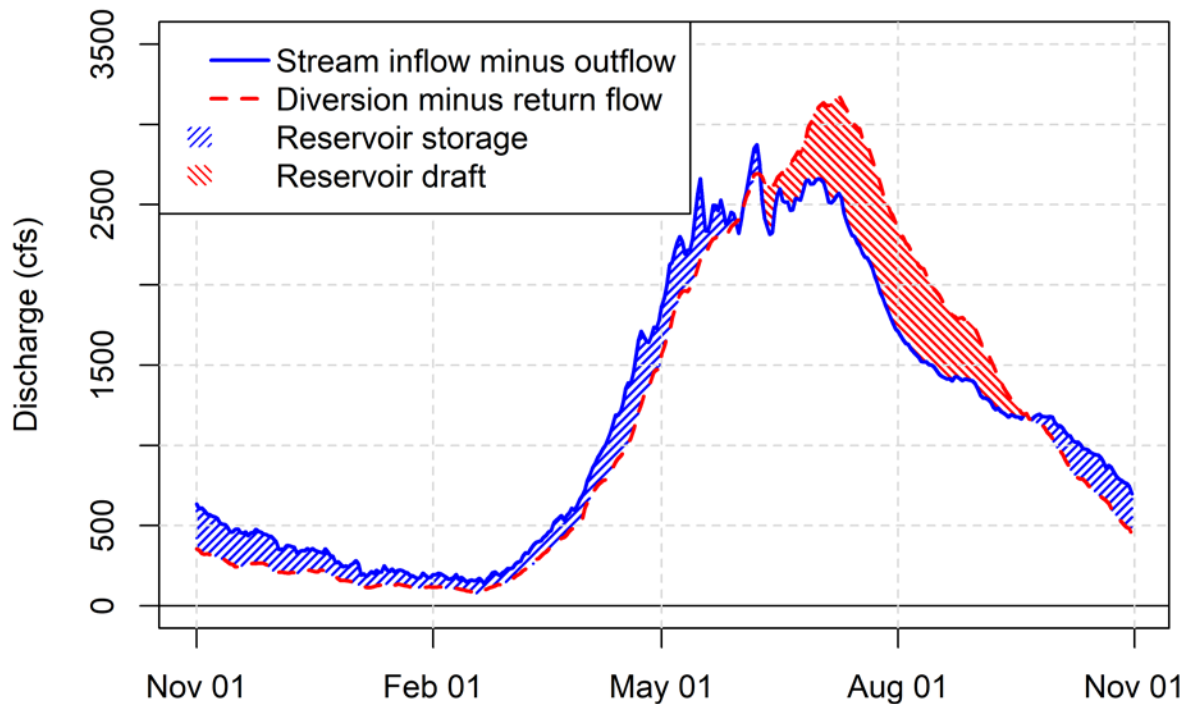


Figure 13. Surface water balance for Henry's Fork watershed. Reservoir storage and draft are net quantities that include direct precipitation to and evaporation from reservoir surfaces.

Because inflow minus outflow is equivalent to diversion minus return flow, we can call this single quantity "net basin withdrawal," which averages around 835,000 ac-ft/year. Net withdrawal is the sum of consumptive use within the basin and outflow from the basin as groundwater. Consumptive use includes crop ET, evaporation from irrigation canals and sprinklers, and all other ET that is not already reflected in basin inflow. Note that total watershed precipitation exceeds basin inflow by about 2.3 million ac-ft/year, and this ET is already accounted for because precipitation supporting it is not ever realized as surface inflow. Using data from 1979-2008, I estimated crop ET on surface-irrigated lands in the watershed at 312,000 ac-ft/year (USBR 2015). Additional consumptive use is probably on the order of 50,000 ac-ft/year, although it is not clear that this use would be reflected in the surface-water calculations done here. In any case, it is likely that somewhere around 500,000 ac-ft/year leave the Henry's Fork watershed as groundwater, subsequently contributing to a combination of consumptive use on and discharge from the ESPA.

Discussion

A natural question to ask is: “If net withdrawal from the surface-water system has not changed in 40 years, does it really matter that reach gains have declined?” The answer is yes. Decrease in reach gain since 2000 has been greater during the summer than winter (Figure 14). During July and August, when draft of Island Park Reservoir is needed to maintain streamflow in the lower watershed, reach gains have averaged 441 cfs less since 2001 than prior. Because reach gains are considered natural flow from an administrative standpoint, decreased reach gains result in earlier priority dates basin-wide and decreased availability of natural flow for water users, even if the same amount of physical water flows down the river. Furthermore, the delay between when water is diverted into the canal system and when it results in inflow to the river temporarily stores natural flow diverted early in the irrigation season for release later in the season, providing the mechanism by which decreased reach gains reduce natural flow availability for water users. When a decreased proportion of lower-watershed streamflow comes from reach gains, greater draft of the reservoir system is needed to maintain a given streamflow, which has negative consequences for both fisheries and water users. Reach gains that travel through groundwater pathways are cooler during summer, providing local refuge for trout. In 2019, Christina Morrisett, a Ph.D. student at Utah State University, used thermal infrared imagery and temperatures measurements to quantify temperature difference between groundwater inputs and ambient surface-water temperature (Figures 15 and 16), finding statistically and biologically significant reductions in water temperature along a gradient from the main river channel to the groundwater input source. Decreased reach gains decrease the thermal benefit of these groundwater inputs, even if total streamflow remains the same. This work has been recently published in a peer-reviewed journal (Van Kirk et al. 2020).

In the Henry’s Fork basin, diversion and reach gain have both decreased by the same amount since 1978, and the decrease has been due primarily to replacement of flood irrigation with sprinkler irrigation (Figure 17). Flood irrigation generally applies more water than is needed by crops, and the difference percolates into the ground beyond the root zone, ultimately recharging local and regional aquifers. Some of that groundwater returns to the river. Sprinkler irrigation is more efficient at applying only the amount of water needed by the plants, resulting in less percolation, less aquifer recharge, and less return flow to the river. As a result, net withdrawal from the basin has not changed since 1978, whether net withdrawal is measured by net diversion (diversion minus reach gain) or by basin surface-water inflow minus outflow. Because changes in local aquifer storage are small and response of reach gains to changes in irrigation practices and local aquifer storage occurs over short time scales (months to a few years), water balance in the Henry’s Fork is relatively simple. Net withdrawal from the Henry’s Fork basin is combination of in-basin consumptive use and outflow from the basin as groundwater, that is,

$$HF \text{ inflow} - HF \text{ outflow} = HF \text{ consumptive use} + \text{groundwater outflow}. \quad (7)$$

Data not analyzed here suggest that the majority of basin withdrawal is outflow as groundwater to the ESPA and that in-basin consumptive use has increased by a small fraction since the late 1970s. The small increase in consumptive use is due to increased irrigation efficiency. It is now fairly well established in the scientific community that increased irrigation efficiency actually increases rather than decreases the total amount of water consumed by crops (see Van Kirk et al. 2019 and references therein).

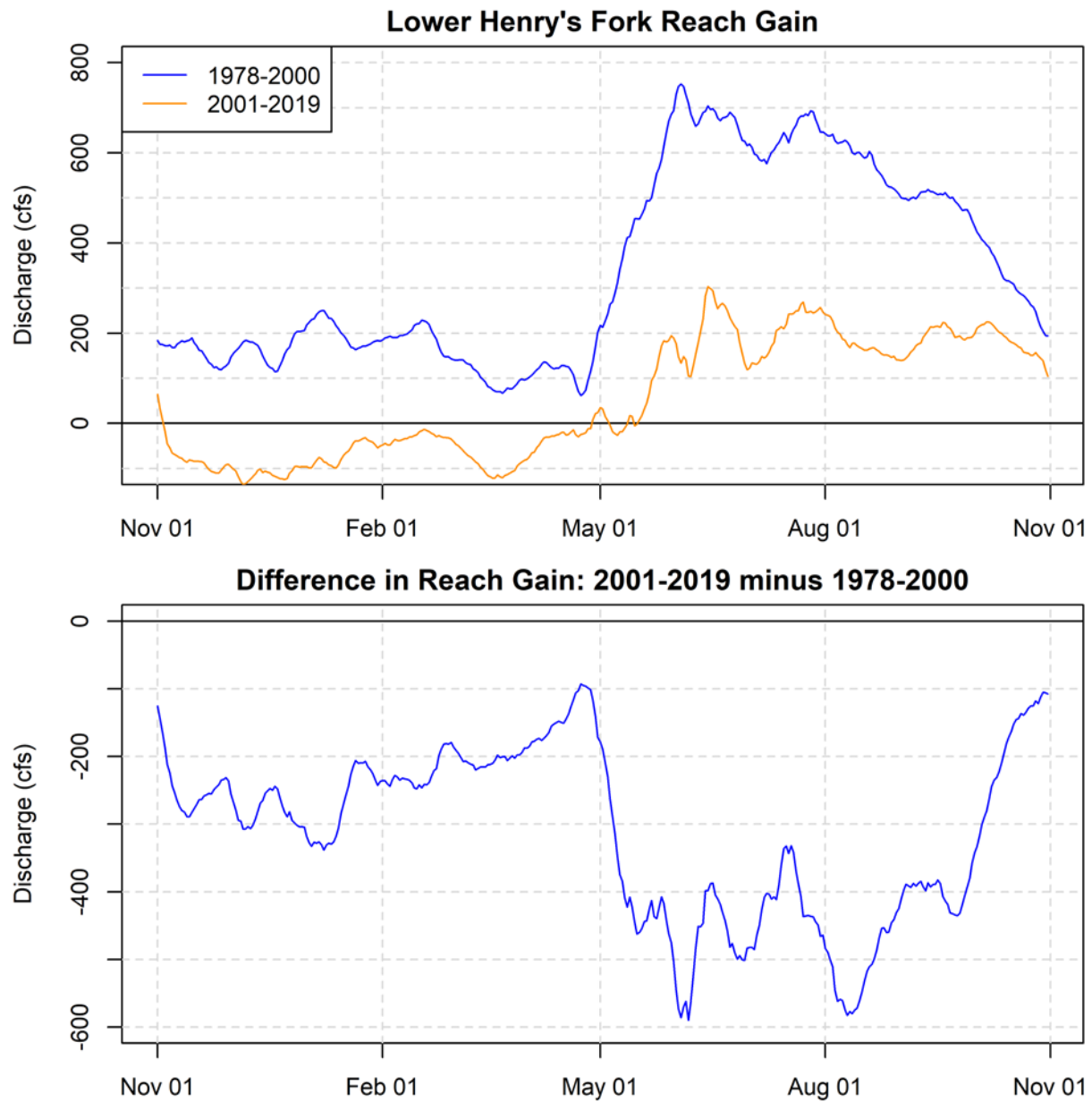


Figure 14. Comparison of mean reach gain hydrographs for the periods 1978-2000 and 2001-2019. Bottom panel is the difference between the 2001-2019 and 1978-2000 hydrographs shown in the top panel.

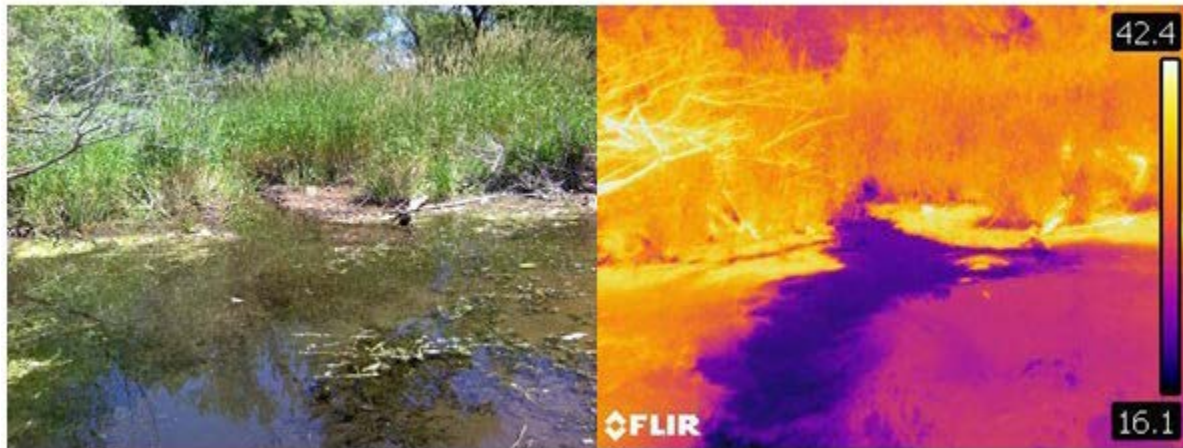


Figure 15. A side-by-side comparison of a visual image (left) and thermal infrared image (right) of a groundwater spring entering the lower Henry's Fork. Temperature is indicated in degrees Celsius.

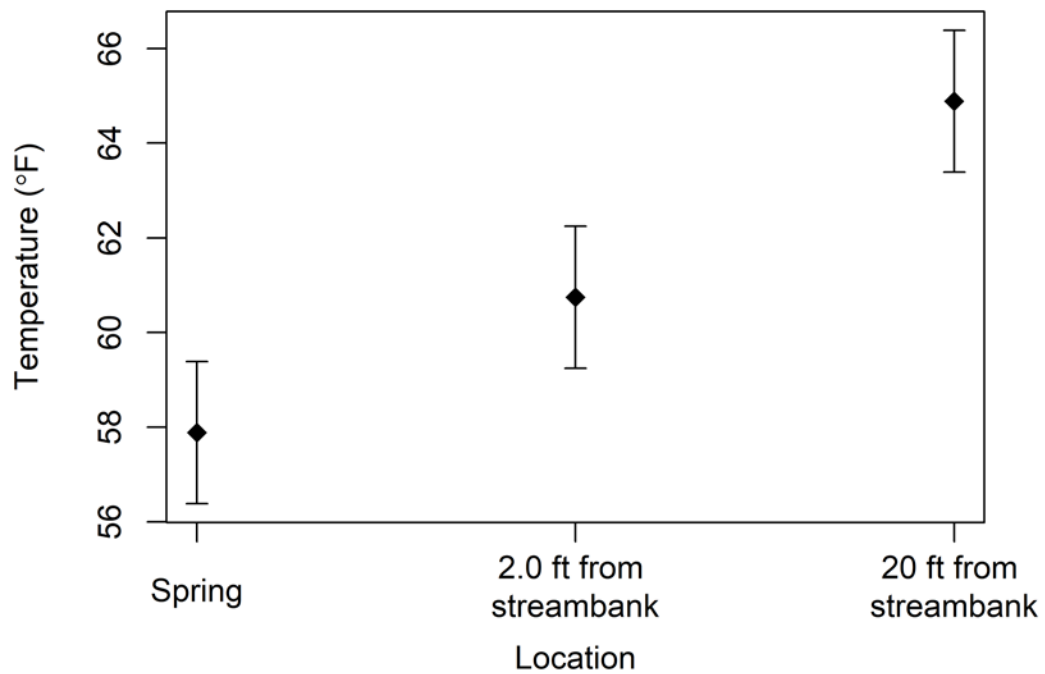


Figure 16. Mean water temperature, with 95% confidence intervals, at three locations measured at each of 20 distinct spring inflow points along the lower Henry's Fork.

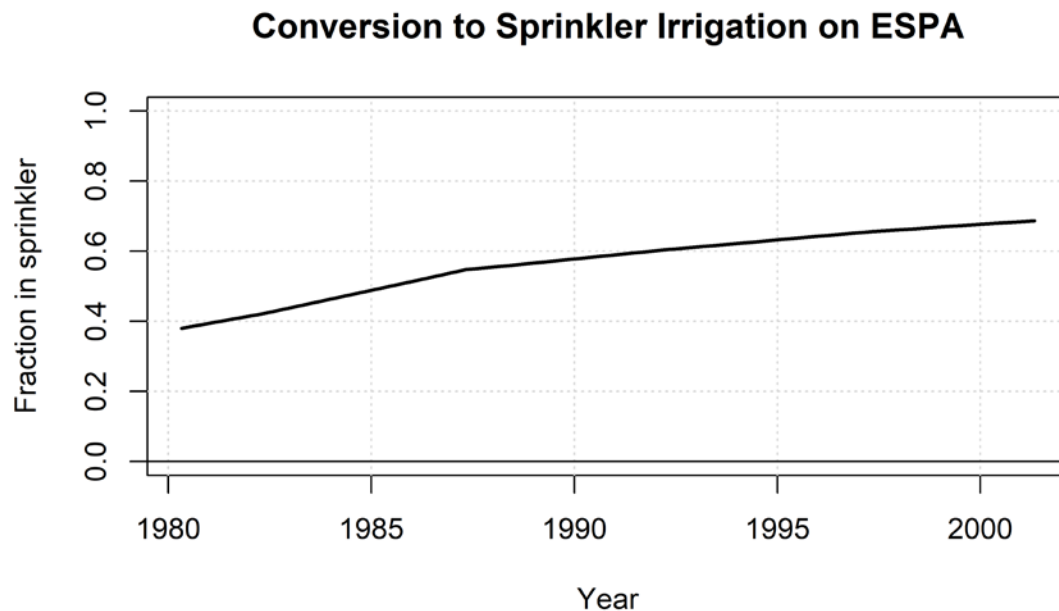


Figure 17. Conversion from gravity to sprinkler irrigation on the ESPA. Data from Contor (2004).

Implications for Conservation

This analysis shows that irrigation efficiency improvements must be carefully considered to avoid unintended negative consequences to fisheries. More efficient sprinkler irrigation can decrease diversion from the river during the middle of the summer, hence reducing the need for draft of Island Park Reservoir to meet irrigation demand. However, the loss of return flow from more efficient irrigation can reduce reach gains, thereby negating any benefit of reduced diversion, possibly requiring even greater draft of Island Park Reservoir, and reducing inputs of cool water to the lower Henry's Fork. Managed aquifer recharge is a mechanism for increasing groundwater returns, which can enhance streamflow and decrease water temperature during the middle of summer on the lower Henry's Fork. If water can be diverted for managed aquifer recharge during spring and fall, this diversion will not increase the need for draft of Island Park Reservoir. A combined strategy of increased mid-summer irrigation efficiency and increased off-season managed aquifer recharge has the potential to benefit fisheries throughout the watershed.

References

- Anderson, D.R. 2008. *Model Based Inference in the Life Sciences*. Springer, New York, NY.
- Boggs, K.G., R.W. Van Kirk, G.S. Johnson, and J.P. Fairley. 2014a. Forecasting aquifer discharge using a data-driven, statistical approach. *Groundwater* 52:853-863.
- Boggs, K.G., G.S. Johnson, R.W. Van Kirk, and J.P. Fairley. 2014b. Forecasting aquifer discharge using a numerical model and convolution. *Groundwater* 52:503-513.
- Boggs, K.G., R.W. Van Kirk, G.S. Johnson, J.P. Fairley, and P.S. Porter. 2010. Analytical solutions to the linearized Boussinesq equation for assessing the effects of recharge on aquifer discharge. *Journal of the American Water Resources Association* 46:1116-1132.

- Burnham, K.P. and D.R. Anderson. 2002. *Model Selection and Multimodel Inference*, 2nd ed. Springer, New York, NY.
- Claeskens, G. and N.L. Hjort. 2008. *Model Selection and Model Averaging*. Cambridge University Press, Cambridge.
- Contor, B. 2004. Delineation of sprinkler and gravity application systems. Eastern Snake Plain Aquifer Model Enhancement Project Scenario Document DDW-022. Technical Report 04-005, Idaho Water Resources Research Institute, University of Idaho, Moscow.
- Doherty, P.F., G.C. White, and K.P. Burnham. 2012. Comparison of model building and selection strategies. *Journal of Ornithology* 152 (Suppl 2):S317-S323.
- Kuzniar, Z.J., R.W. Van Kirk, and E.B. Snyder. 2017. Seasonal effects of macrophyte growth on rainbow trout habitat availability and selection in a low-gradient, groundwater-dominated river. *Ecology of Freshwater Fish* 26:653-665.
- McLaren, J.S., T.V. Royer, R.W. Van Kirk, and M.L. Muradian. 2019. Management and limnology drive water temperature patterns in a middle Rockies river-reservoir system. *Journal of the American Water Resource Association* 55:1323-1334.
- U.S. Bureau of Reclamation (USBR). 2015. Henry's Fork Basin Study Final Report. USBR Pacific Northwest Region, Boise, ID.
- Van Kirk, R.W., B.A. Contor, C.N. Morrisett, S.E. Null, and A.S. Loibman. 2020. Potential for managed aquifer recharge to enhance fish habitat in a regulated river. *Water* 12:673.
- Van Kirk, R., B. Hoffner, A. Verbeten, and S. Yates. 2019. New approaches to providing instream flow for fisheries in the American West: Embracing prior appropriation and the marketplace. Pp. 515-564 in D.C. Dauwalter, T.W. Birdsong, and G.P. Garrett, editors. *Multispecies and Watershed Approaches to Freshwater Fish Conservation*. American Fisheries Society Symposium 91, Bethesda, MD.