

Prepared in cooperation with the Bureau of Reclamation (Interagency Agreement R16PG00120)

Use of Boosted Regression Trees to Quantify Cumulative Instream Flow Resulting from Curtailment of Irrigation in the Sprague River Basin, Oregon

Scientific Investigations Report 2019-5130

Use of Boosted Regression Trees to Quantify Cumulative Instream Flow Resulting from Curtailment of Irrigation in the Sprague River Basin, Oregon

By Tamara M. Wood

Prepared in cooperation with the Bureau of Reclamation (Interagency Agreement R16PG00120)

Scientific Investigations Report 2019-5130

U.S. Department of the Interior
U.S. Geological Survey

U.S. Department of the Interior
DAVID BERNHARDT, Secretary

U.S. Geological Survey
James F. Reilly II, Director

U.S. Geological Survey, Reston, Virginia: 2020

For more information on the USGS—the Federal source for science about the Earth, its natural and living resources, natural hazards, and the environment—visit <https://www.usgs.gov> or call 1–888–ASK–USGS.

For an overview of USGS information products, including maps, imagery, and publications, visit <https://store.usgs.gov>.

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Although this information product, for the most part, is in the public domain, it also may contain copyrighted materials as noted in the text. Permission to reproduce copyrighted items must be secured from the copyright owner.

Suggested citation:

Wood, T.M., 2019, Use of boosted regression trees to quantify cumulative instream flow resulting from curtailment of irrigation in the Sprague River basin, Oregon: U.S. Geological Survey Scientific Investigations Report 2019-5130, 25 p., <https://doi.org/10.3133/sir20195130>.

ISSN 2328-0328 (online)

Contents

| | |
|--|----|
| Abstract..... | 1 |
| Introduction..... | 1 |
| Use Of Boosted Regression Trees To Model Streamflow..... | 5 |
| Surface-Water Budget..... | 5 |
| Boosted Regression Trees..... | 5 |
| Data Used To Develop Sprague River Discharge Boosted Regression Trees Model..... | 6 |
| Response Variable..... | 6 |
| Predictor Selection..... | 6 |
| Predictors Related To Interannual Variability In Water Table Elevation..... | 8 |
| Predictors Related To Current Water Year Weather And Hydrology..... | 9 |
| Building And Evaluating The Sprague River Discharge Boosted Regression Trees Model..... | 10 |
| Identification Of Model Start Date And Selection Of Training Data..... | 10 |
| Statistical Methods..... | 11 |
| Selection Of Model Parameters..... | 11 |
| Model Optimization And Testing With Holdout Data..... | 13 |
| Using The Boosted Regression Trees Model To Quantify Cumulative Instream Flow Resulting From Curtailment Of Irrigation..... | 17 |
| Conclusion..... | 22 |
| References Cited..... | 22 |

Figures

| | |
|---|----|
| 1. Map showing the upper Klamath and Sprague River Basins, Oregon..... | 2 |
| 2. Graph showing percent relative influence, averaged over all trees, of the final selection of 17 predictor variables in the boosted regression trees model..... | 8 |
| 3. Graph showing depth-to-water in Oregon Water Resources Department well No. 2145, interpolated with a locally weighted scatterplot smoothing to filter out intra-annual variability, Sprague River Basin, Oregon..... | 9 |
| 4. Graph showing relation between model complexity as defined by the number of trees and model performance as measured by root-mean-square error of the tenfold cross-validation testing prediction for four values of learning rate..... | 13 |
| 5. Graph showing relation between model complexity as defined by the number of trees and model performance as measured by root-mean-square error of the tenfold cross-validation testing prediction for four values of learning rate..... | 14 |
| 6. Graph showing mean error in 33 holdout years, as a function of the measured daily streamflow, averaged over model dates in each year..... | 15 |
| 7. Boxplot showing correlation coefficient in 33 holdout years, for years in which daily streamflow, averaged over model dates, was high..... | 15 |
| 8. Observed and predicted target variable in 4 holdout years, Sprague River at Chiloquin, Oregon..... | 16 |
| 9. Graph showing cumulative difference between predicted and measured streamflow in each of 33 holdout years and in 4 years when irrigation was regulated, Sprague River at Chiloquin, Oregon, starting on July..... | 17 |

Figures—Continued

| | |
|--|----|
| 10. Boxplot showing cumulative difference between predicted and measured streamflow in 33 holdout years, Sprague River at Chiloquin, Oregon, starting July 1 and ending September 30..... | 18 |
| 11. Graph showing thresholds for quantification of cumulative instream flow resulting from the curtailment of irrigation, as a function of days from the start of regulation action, Sprague River at Chiloquin, Oregon..... | 19 |
| 12. Graphs showing measured and predicted streamflow in 4 post-regulation years..... | 19 |
| 13. Graphs showing difference between boosted-regression-trees-predicted streamflow and streamflow observed at the streamgage in 4 post-regulation years, Sprague River at Chiloquin, Oregon | 20 |
| 14. Graph showing model estimate of the cumulative instream flow made available through the end of September by irrigation curtailment in post-regulation years, Sprague River at Chiloquin, Oregon | 21 |

Tables

| | |
|---|----|
| 1. Response variable and model predictors. | 7 |
| 2. Results of model training and testing on the entire 33-year training dataset, with a tenfold cross-validation..... | 12 |
| 3. Results of model testing on holdout data | 15 |
| 4. Cumulative instream flow from irrigation curtailment (RIF) at the Chiloquin gage on the Sprague River, calculated in 4 post-regulation years | 21 |

Conversion Factors

U.S. customary units to International System of Units

| Multiply | By | To obtain |
|--|-----------|--|
| Length | | |
| mile (mi) | 1.609 | kilometer (km) |
| Area | | |
| acre | 4,047 | square meter (m ²) |
| acre | 0.4047 | hectare (ha) |
| acre | 0.004047 | square kilometer (km ²) |
| square mile (mi ²) | 2.590 | square kilometer (km ²) |
| Volume | | |
| cubic foot (ft ³) | 0.02832 | cubic meter (m ³) |
| acre-foot (acre-ft) | 1,233 | cubic meter (m ³) |
| Flow rate | | |
| cubic foot per second (ft ³ /s) | 0.02832 | cubic meter per second (m ³ /s) |

International System of Units to U.S. customary units

| Multiply | By | To obtain |
|--|-----------|--|
| Length | | |
| meter (m) | 3.281 | foot (ft) |
| kilometer (km) | 0.6214 | mile (mi) |
| Area | | |
| square meter (m ²) | 0.0002471 | acre |
| hectare (ha) | 2.471 | acre |
| hectare (ha) | 0.003861 | square mile (mi ²) |
| square kilometer (km ²) | 0.3861 | square mile (mi ²) |
| Volume | | |
| cubic meter (m ³) | 35.31 | cubic foot (ft ³) |
| cubic meter (m ³) | 0.0008107 | acre-foot (acre-ft) |
| Flow rate | | |
| cubic meter per second (m ³ /s) | 35.31 | cubic foot per second (ft ³ /s) |

Abbreviations

| | |
|--------|---|
| ABT | aggregated boosted regression trees |
| BRT | boosted regression trees |
| CV | cross-validation approach |
| ENSO | El Niño-Southern Oscillation |
| ET | evapotranspiration |
| GCM | global-scale general circulation models |
| LOWESS | locally weighted scatterplot smoothing |
| MEA | mean absolute error |
| R | correlation coefficient |
| RIF | regulated instream flow |
| RMSE | root mean square error |
| SE | standard error |
| TNI | Trans-Niño Index |
| USGS | U.S. Geological Survey |

Use of Boosted Regression Trees to Quantify Cumulative Instream Flow Resulting from Curtailment of Irrigation in the Sprague River Basin, Oregon

By Tamara M. Wood

Abstract

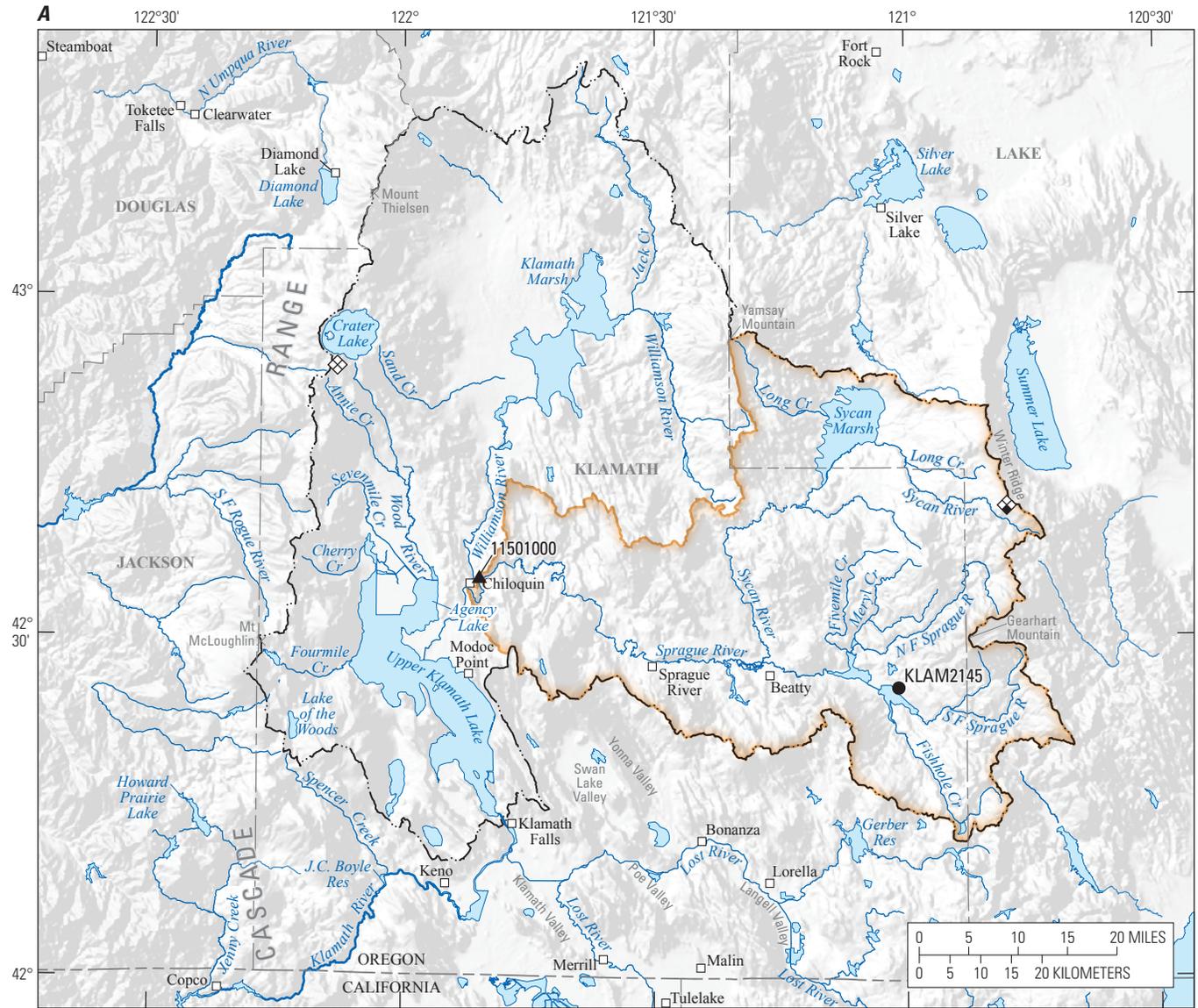
A boosted regression trees (BRT) approach was used to estimate the amount by which streamflow is increased when irrigation is regulated (curtailed) upstream of a streamgage on the Sprague River in southern-central Oregon. The BRT approach differs from most other approaches that require baseline conditions for comparison, where those baseline conditions are determined from past observations by searching for hydrologically similar years when irrigation was not regulated. Such baseline conditions are always imperfect estimates of the true baseline conditions. The BRT approach instead estimates unique baseline conditions for any year in which irrigation is regulated by calculating the baseline condition based on measurements of precipitation and weather observations that determine evapotranspiration, and other measurements that are proxies for the effects of climate and regional groundwater pumping on water-table elevation, using a model that has been trained in years of no regulation. The amount by which streamflow is increased by regulation is then calculated by subtracting the estimated baseline conditions from the measured streamflow. The approach is challenged by the fact that the streamflow increase may be a small fraction of the total streamflow; nonetheless, during 2 years in which regulation was started early and was implemented consistently through the season, the increased flow made up about one third of the flow past the streamgage during the regulation period. An advantage of this approach is that with rigorous model testing with holdout data, the threshold for detecting streamflow increase and intervals around the estimates of increase at a desired level of confidence can be quantified. The model relies on datasets that are readily available and updated continuously and therefore can be used operationally to inform resource management.

Introduction

Water managers in the upper Klamath Basin of southern-central Oregon (fig. 1) have long struggled with the challenges of managing water for competing demands in a semi-arid basin-and-range landscape. Water in the Klamath basin above Upper Klamath Lake is managed according to the doctrine of prior appropriations (Oregon Water Resources Department, 2018), which applies a priority date and amount to each water right. When insufficient water is available to satisfy all water rights, water users with senior priority dates make a “call” on users with junior water rights so that the junior users are shut off until the rights of the senior users making the call are satisfied. This includes instream rights (rights that keep water in the stream channel for the benefit of wildlife) which, in the upper Klamath Basin, are senior to consumptive-use rights (rights that remove water from the stream channel for off-channel use). For this reason, enforcement of senior water rights results in curtailment of diversions in the tributaries flowing into Upper Klamath Lake, including the Williamson and Sprague Rivers, and in more water being retained in these streams.

The final ruling of the Klamath Basin water adjudication in March, 2013 (State of Oregon, 2018), allowed regulation (curtailment of diversions) based on water rights throughout the basin. That was not the case prior to 2013, and regulation activities were small-scale and related mostly to complaints between neighbors during dry conditions. Water is critical to the region’s agriculture and is the basis for healthy ecosystems in streams flowing into the lake. In many years, water has been inadequate to meet both agricultural and environmental demands. Management of this resource requires quantifying water flows and the effect of management decisions on flows with the greatest possible accuracy. For this reason, in years in which water diversions of junior water rights are curtailed, it is of great interest to the water managers, water users, and stakeholders in the upper Klamath Basin to quantify the increased volume of water flowing into Upper Klamath Lake as a result of the regulation of water diversions on streams flowing into upper Klamath Lake.

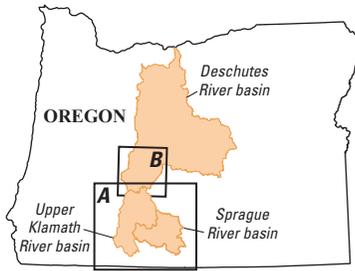
2 Using Boosted Regression Trees to Quantify Instream Flow Resulting from Curtailment of Irrigation



Base modified from U.S. Geological Survey and other digital sources. Coordinate system: Universal Transverse Mercator, Zone 10N, North American Datum of 1927.

- EXPLANATION**
- Upper Klamath River basin
 - Sprague River basin
 - ▲ U.S. Geological Survey streamgage
 - Oregon Water Resources Department well
 - ◆ NOAA meteorological station, Crater Lake NPS
 - ◆ NRCS Summer Rim SNOTEL

[NOAA, National Oceanic and Atmospheric Administration
 NRCS, Natural Resources Conservation Service
 NPS, National Park Service
 SNOTEL, Snow Telemetry]



- EXPLANATION**
- Deschutes River basin

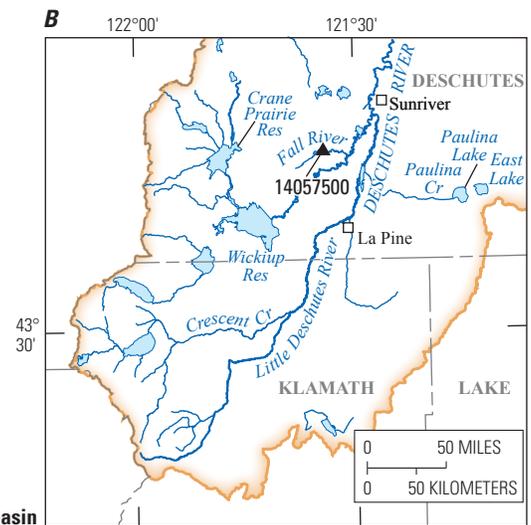


Figure 1. Upper Klamath and Sprague River Basins, Oregon.

The streamflow resulting from regulation of water diversions (referred to as “regulation instream flow” [RIF] hereinafter) is difficult to quantify. Generally, water-right diversions are not metered at the point of diversion, and metering all diversions would still only provide an upper bound on the volume of water ultimately reaching the lake as a result of the regulation of diversions because not all diverted water is lost to the stream system through evapotranspiration (consumptive use). Some water will return to the stream as surface or groundwater flow. The inventory of issued water rights is also not an accurate way of estimating the RIF, because the diversion rate listed in a water right is an upper limit on the amount of water which may be diverted; the true amount varies with landowner water-management decisions and can vary with environmental conditions and crop demands, according to in-season weather, climate cycles, and antecedent conditions controlled by current and prior years’ precipitation, temperature, and snowpack. Confounding the issue is the fact that the “signal” of the RIF may be small relative to streamflow observed at a given downstream streamgauge site, because in this basin total upstream irrigated acres are typically a fraction of the total acres above a given streamgauge.

An important aspect of the problem is the fact that the RIF is calculated by comparison to an unknown baseline condition—the streamflow that would have occurred if the environmental conditions were the same but irrigation was not curtailed. There are other approaches to calculating the RIF, but each must solve the problem of this unknown baseline condition. For example, Hess and Stonewall (2014) estimated the RIF at a streamgauge relative to a baseline condition that was obtained by searching the historical record for hydrologically similar years, and then merging those similar years into a single composite index year. A fundamentally different approach is to calculate the RIF from an energy balance by calculating the difference in evapotranspiration (ET) over irrigated areas between a regulated year and a baseline year (Senay and others, 2016). Both approaches have limitations in that no year in the historical record of either streamflow or Landsat imagery will have exactly the same hydroclimate and antecedent conditions as the target year. The index year method seeks to mitigate this problem by creating a composite baseline based on several years and additionally can be modified to use only the baseflow portion of the hydrograph in order to remove the effects of precipitation events that complicate the identification of appropriate baseline hydrographs (Hess and Stonewall, 2014). The remote sensing method can be modified to account for differences in precipitation between the target and baseline years (Reitz and others, 2017).

In this study, a fundamentally different approach is taken to estimate RIF. We estimate the baseline condition directly from measured hydroclimatic and hydrologic data, and we calculate the effects of irrigation regulation based on differences between observed streamflow at a streamgaged site and our estimated baseline. Furthermore, the baseline is estimated using machine learning; that is, by learning the baseline response from the statistical relationship between observable environmental conditions and historical flow observations rather than calculating the baseline response based on an understanding of underlying physical processes. The baseline condition that we estimate is the streamflow at a streamgauge located on the Sprague River at Chiloquin, Oregon. The RIF is then calculated as the difference of streamgauge measurements from the estimated baseline streamflow. This method resembles more closely the index year method than the remote sensing method in that it is not spatially explicit, instead estimating RIF over the watershed delineated upstream of a streamgauge. A potential advantage of this method over the index year method is that it does not rely on finding a hydroclimatically similar year or set of years in the historical record to provide a baseline; rather, a statistical model is used to estimate a baseline that is unique to each target year.

The BRT model is created by training and testing on a long period of measured streamflow prior to 2013, the year that irrigation curtailment was first imposed on the basin upstream of Upper Klamath Lake. The premise of this approach is that surface water diversions for irrigation in response to a given set of measured hydroclimatic conditions, even if the response is complicated and nonlinear, can be “learned” by the model if it is trained on a long period of record covering a wide range in conditions prior to 2013. The assumption also is made that the response of these diversions to hydroclimatology was consistent (statistically stationary) through the training period.

Beginning in 2013, the baseline streamflow, unmodified by regulation, is simulated with the trained and tested model, and the RIF is calculated as the difference between the simulated baseline and the measured streamflow. An additional required assumption is that the difference is largely attributable to regulation activities and not to other, unmeasured causes that could include, for example, timber removal activities, juniper expansion, changes in the beaver population, and road building.

The problem of estimating the unobserved baseline has some similarity to the problem of forecasting streamflows, particularly at short lead times, to which machine learning has been applied successfully (Rasouli and others, 2012;

4 Using Boosted Regression Trees to Quantify Instream Flow Resulting from Curtailment of Irrigation

Erdal and Karakurt, 2013; Humphrey and others, 2016; Lima and others, 2016; Yaseen and others, 2016), but there is an important distinction between this problem and short-lead forecasting. Streamflow forecasts often include the most recent streamflow observations as predictor variables; in this case, that is not possible because the model is used to estimate a baseline streamflow that is not observed. This also means that the model cannot be updated regularly with new information, which has been shown to improve forecasts to the extent that new observations contain novel event information not included in the training data (Lima and others, 2016). For these reasons, a more apt comparison might be to statistical downscaling that relates large-scale climate information obtained from global-scale general circulation models (GCMs) directly to localized streamflow without the intermediate step of running a deterministic hydrologic model. Statistical downscaling can be used to examine historical trends in streamflow over long time periods based on global re-analysis data (Cannon and Whitfield, 2002), to predict climate change impacts using GCM simulations of future climate at monthly (Ghosh and Mujumdar, 2008; Sachindra and others, 2013; Umut and Gul, 2015; Nasser and others, 2017) or daily (Tisseuil and others, 2010) time scales, or to forecast water supply (Erdal and Karakurt, 2013). The conclusion that nonlinear statistical methods perform better than linear methods for this problem is the norm, when the two methods are directly compared (Cannon and Whitfield, 2002; Tisseuil and others, 2010; Rasouli and others, 2012; Joshi and others, 2013; Sachindra and others, 2013). The modeling approach described hereinafter differs from a downscaling problem in that it uses high-resolution, gridded surface meteorological data (Abatzoglou, 2013) instead of re-analysis data, and the gridded data is augmented with local observations to provide the model site-specific information regarding climate cycles and current year weather and hydrology.

It is not the purpose of this study to compare the performance of a selection of machine learning methods. In this study, the boosted regression trees (BRT) technique is used. A direct comparison of BRT to artificial neural networks and Bayesian networks for the prediction of nitrate in groundwater of the Central Valley in California (Nolan and others, 2015) and the meta-modeling of surface water depletion expressed as the source of water-to-wells in the Lake Michigan Basin (Fienen and others, 2016) showed that

BRT performed comparably to, or slightly better than, the other two methods for these hydrologic problems at a regional scale. Snelder and others (2009) used BRT to predict flow regimes in ungaged catchments in France based on watershed characteristics and found BRT to have better performance than linear discriminant analysis or classification and regression trees. For the problem of statistical downscaling, which is closely related to our problem of estimating RIF, aggregated boosted regression trees (ABT, a variation on BRT in which multiple BRT models built on subsets of the data are averaged together) slightly outperformed artificial neural networks and Bayesian networks, though the authors note that all three models performed comparably (Tisseuil and others, 2010). BRT has been used successfully to relate hydroclimate and basin topography to multi-decadal streamflow trends at the large spatial scales of the continental and sub-continental United States (Rice and others, 2016) and to forecast streamflow at monthly time scales (Erdal and Karakurt, 2013). Based on these previous studies, we determined that BRT is an appropriate choice of machine learning algorithm to use for estimating the unknown baseline streamflow.

From the water managers' perspective, a model for calculating RIF is most useful if it is "operational," meaning that it can make calculations at time scales of 2 weeks or less to inform decisions. This requires temporal resolution of a few days or less and also requires that model calculations be based on environmental measurements that are updated frequently. For this reason, the model was developed based on daily measurements and uses datasets that are available with a delay of a few days or less. Second, because the RIF accumulates over sub-seasonal time scales based on operational decisions that occur intermittently at intervals from days to weeks, the temporal resolution required is greater than that required for a medium-to-long range forecast with the purpose of providing water availability information in advance of an irrigation season (Risley and others, 2005). Third, because there will never be observations against which to compare and evaluate the model estimation of the baseline in the years of curtailment, it becomes particularly crucial to understand the expected accuracy of the estimation based on the performance of the model in estimating observations made in years prior to the implementation of curtailment. For this reason, an important component of this study is the use of an ensemble of tests on holdout data¹ to quantify the uncertainty in the cumulative RIF through an irrigation season.

¹ As used herein, holdout data are a 1-year subset of the full 33-year dataset that provides a final estimate of the BRT model's performance after it has been trained and tested with the rest of the data. Holdout data are not used in the training of the BRT model.

Use Of Boosted Regression Trees To Model Streamflow

Surface-Water Budget

A representation of the surface water budget for the drainage basin above the Chiloquin streamgage is:

$$SW_{out} = PRECIP - ET - GW_r + GW_d \quad (1)$$

where,

SW_{out} is the streamflow measured at the streamgage,

$PRECIP$ and ET are the precipitation and evapotranspiration, respectively, integrated over the entire basin,

GW_r is water at the surface infiltrating to the water table, and

GW_d is the streamflow gained from groundwater discharge.

All of the terms on the right-hand side of the equation are variable over the landscape and through time.

ET incorporates evapotranspiration from all landscape types in the basin—forests, grasslands, wetlands, and irrigated and non-irrigated agriculture. Process-based calculation of evapotranspiration requires spatially dense data describing land cover and soil characteristics—soil moisture, temperature profile—as well as the atmospheric weather data that drives the evapotranspiration, including solar radiation, wind, air temperature, and humidity. Weather observations include current observations but also some integration of observations over previous weeks to months that determine antecedent conditions, primarily soil moisture. Basin-scale variability in the atmospheric drivers must also be incorporated for accurate calculations of $PRECIP$ and ET , particularly in a basin like the Sprague River basin that varies in elevation.

The two remaining terms, groundwater recharge and discharge, are not continuous over the basin; recharge is largely limited to higher elevations, and discharge occurs at discrete locations. Over annual time scales groundwater recharge and discharge can probably be considered approximately in balance with each other (Gannett, 2007), but over longer time scales the small imbalances in these terms lead to variability in water-table elevation and consequent variability in baseflow to the stream. This interannual variability in baseflow is correlated to climate cycles or long-term changes in storage owing to groundwater pumping and is highly relevant to the current problem because baseflow dominates the streamflow during the summer and early autumn months—exactly the season when accurate estimates of RIF are desired.

Machine learning was tested in this study because, if appropriate predictor variables representing forcing functions can be selected, the algorithms of the model can be relied on to find the complex nonlinear relationships that define the response (SW_{out} in the mass balance) to those forcing functions

without the intermediate step of explicitly calculating $PRECIP$, ET , GW_r , and GW_d . An inherent assumption is that the human behavioral response to hydroclimatic forcing (in terms of crop type and acreage and corresponding unregulated irrigation practices) is consistent through the training period and can be learned by the model.

Boosted Regression Trees

The theory and use of boosted regression trees are provided eloquently in classic texts and papers on the subject (Friedman, 2001; Hastie, 2001; Friedman, 2002; De'ath, 2007), and is not repeated here. It is useful, however, to briefly contrast the basic mathematical form of the BRT model of streamflow with the mass balance above:

$$SW_{out} = \sum_{m=1}^M \beta_m b(\mathbf{p}; \gamma_m) \quad (2)$$

In BRT formulation, the streamflow at the streamgage is represented as a summation over M basis functions b (the regression trees), each of which is a function of the vector of predictor variables \mathbf{p} . The set of potential predictor variables comprises short-term weather observations, soil characteristics, irrigated acres, and long-term climate observations. Variables derived from weather observations that represent a 1–6-month memory of weather and hydrologic conditions in the system are also included in \mathbf{p} .

The model is built sequentially by fitting a new tree at each iteration m to the residuals remaining after the last iteration, $m - 1$. The parameters β and γ at each iteration are calculated to minimize a loss function. The improvement made by the new tree at each iteration is focused on a small region of the predictor space, so that the overall improvement in the representation of the response variable at each iteration is small, but the process of adding together trees created over many iterations (“boosting”) creates a good solution. The total number of trees M is generally hundreds to thousands; the process of optimizing M is part of the model-building process. Minimization of a generalized loss function can be made more efficient by applying the method of steepest descent (“gradient boosting”; Friedman, 2001), and performance (in terms of error, computation time, and reduction of over-fitting) can be improved by selecting only a fraction of the data to fit a tree at any given iteration (“stochastic gradient boosting”; Friedman, 2002). Thus, the model uses “stochastic, gradient-boosted regression trees,” which is abbreviated to “boosted regression trees” (BRT).

BRT model calculations were done using version 2.1.3 of the *gbm* package (Greenwell and others, 2018) in the R computing environment (R Foundation for Statistical Computing, 2018). Parameter selection and holdout testing was done using task-specific wrapper scripts for the *gbm* package written by the author, also in the R computing environment.

Data Used To Develop Sprague River Discharge Boosted Regression Trees Model

Response Variable

The model response variable is derived from daily mean flow in the Sprague River near Chiloquin, the site of U.S. Geological Survey (USGS) streamgage 11501000 (U.S. Geological Survey, 2017). The Sprague River joins with the Williamson River downstream of Chiloquin to form the largest tributary to Upper Klamath Lake (fig. 1). The Sprague River drains 1.002 million acres to the east of Upper Klamath Lake, originating in volcanic uplands that border the eastern side of the upper Klamath Basin. The hydrograph at this streamgage is typically characterized by high flows driven by rainfall runoff during the winter followed by snowmelt-driven peak flows in the spring which, following a spring recession period, give way to base flow conditions in the summer and autumn. Irrigated acres in the Sprague River basin upstream of the Chiloquin streamgage make up approximately 6.7 percent of the 1.02 million acres in the basin.

The record of flow at this streamgage goes back to 1921, but the training and testing period is constrained by the atmospheric predictor data to start in 1980. The calendar-day-mean flow was calculated over 33 years (1980–2012) to get an average seasonal cycle. This long-term average cycle was subtracted from the daily mean flow to create the model response variable.

The dates covered by the model in each year were based on streamflow at the streamgage. The model start date in each year is variable and is defined as the first day following the day that the flow was 500 cubic feet per second (ft³/s) for the last time in the water year², or March 2 if springtime flows never reach 500 ft³/s. The end date of the model in each year is September 30th, the last date in the water year. The number of model dates in each year varies from 72 in 1995 (a year with high spring flows in which the model starts on July 22) to 214 in 1992 (a year in which water year flows never were as high as 500 ft³/s and in which the model starts on March 2).

²The 12-month period from October 1 of any given year to September 30 of the following year.

Predictor Selection

The relative influence of each predictor in an individual regression tree can be quantified as the sum of the squared improvements at all splits determined by that predictor (Breiman and others, 1984). The relative influence of the predictor in the model as a whole is then the average over each tree, and for a cross-validated model, the average over each tree in each fold.

Selection for the predictor variables with the most important explanatory power is inherent to the BRT algorithms—variables with little explanatory power have small values of relative influence and very little effect on the outcome. Nonetheless, having a long list of predictor variables adds unnecessary computation time and can make interpretation of the model dependencies more difficult. Therefore, before beginning the process of rigorous parameter selection, a process of model simplification was executed to limit the number of predictors. A reasonable value of learning rate (0.01) and number and complexity of trees (2,000 trees with an interaction depth of 5) was assumed, and a list of 35 potential predictors was reduced to 20 in an iterative process by dropping the variable with the lowest relative influence score at each iteration and running the model again; this process was repeated until only 5 predictor variables were left. In this way it was determined that the relative improvement gained with each additional predictor began to drop off more quickly after the number of predictors reached 17, and an additional 3 predictors were initially retained to provide some flexibility in the model to re-order predictors during the parameter selection process. After the learning rate and number of trees were optimized, the three variables with the smallest relative influence were dropped from the model. The final list of 17 variables with their relative influence in the model (averaged over 10 folds of cross validation) is provided in table 1 and in figure 2.

Table 1. Response variable and model predictors.

[The length of record used in all cases is 33 years (1980–2012). Relative influence is based on the chosen best, simplest model during model training and testing, with final parameters as in table 2. USGS, U.S. Geological Survey; —, not applicable]

| Source | Data type | Data interval and statistic | Derived variables | Relative influence | Short name |
|---|------------------------|---|--|--------------------|--------------------|
| Response variable | | | | | |
| USGS Sprague River gage | | | | | |
| Sprague River near Chiloquin, Oregon, streamgage No. 11501000 | Streamflow | Daily mean | 33-yr calendar day mean subtracted | — | — |
| Predictor variables | | | | | |
| Oregon Water Resources Department | | | | | |
| Upper Sprague River, Oregon, well KLAM2145 | Depth to water | Instantaneous value at variable intervals, approximately 4 times per year | Interpolated to daily values using locally weighted scatterplot smoothing (LOWESS) | 19.24 | wl2145_dtw |
| Fall River, Oregon, streamgage No. 14057500 | Streamflow | Daily mean | Cumulative flow since beginning of current WY | 8.60 | fr_cum_WY |
| | | Annual by water year | Cumulative flow over entire water year, lagged by 1 water year | 3.51 | fr_WY_lag1yr |
| University of Idaho GridMET | | | | | |
| | Precipitation | Daily accumulation | Lagged by 5 days | 6.96 | precip_lag5days |
| | | Daily accumulation | Change over last 5 days | 5.28 | precip_dif5days |
| | | Daily accumulation | Accumulation from beginning of current water year | 3.36 | precip_cum_WY |
| | Air temperature | Daily maximum | — | 3.54 | airt_dailymax |
| | | Daily maximum | Mean over last 180 days | 13.31 | airt_mean180d |
| | Relative humidity | Daily minimum | — | 2.18 | relhum_dailymin |
| | | Daily minimum | Mean over last 30 days | 7.86 | relhum_mean30d |
| | Specific humidity | Daily mean | Mean over previous 180 days | 7.09 | sphum_mean180d |
| | | Daily mean | Mean over previous 30 days | 2.17 | sphum_mean30d |
| Natural Resources Conservation Service SNOTEL | | | | | |
| Summer Rim, Oregon, site number 20G02S | Precipitation | Daily accumulation | Accumulation from beginning of current WY | 3.76 | sr_precip_cum_WY |
| National Oceanic and Atmospheric Administration, National Centers for Environmental Information | | | | | |
| Crater Lake National Park Headquarters weather station, site GHCND:USC00351946 | Precipitation | Daily accumulation | Lagged by 3 days | 2.19 | cl_precip_lag3days |
| | Snowfall/precipitation | Daily accumulation | Ratio of snowfall accumulation from beginning of current WY to precipitation accumulation from beginning of current WY | 2.36 | cl_sn2pr_cum_WY |
| | | Daily accumulation | Ratio of snowfall accumulation over entire WY to precipitation accumulation over entire WY, lagged by 2 years | 2.43 | sl_sn2pr_WY_lag2yr |
| — | Serial day | — | Days since target variable was 500 cubic feet per for the last time in the water year | 6.17 | ser_day |

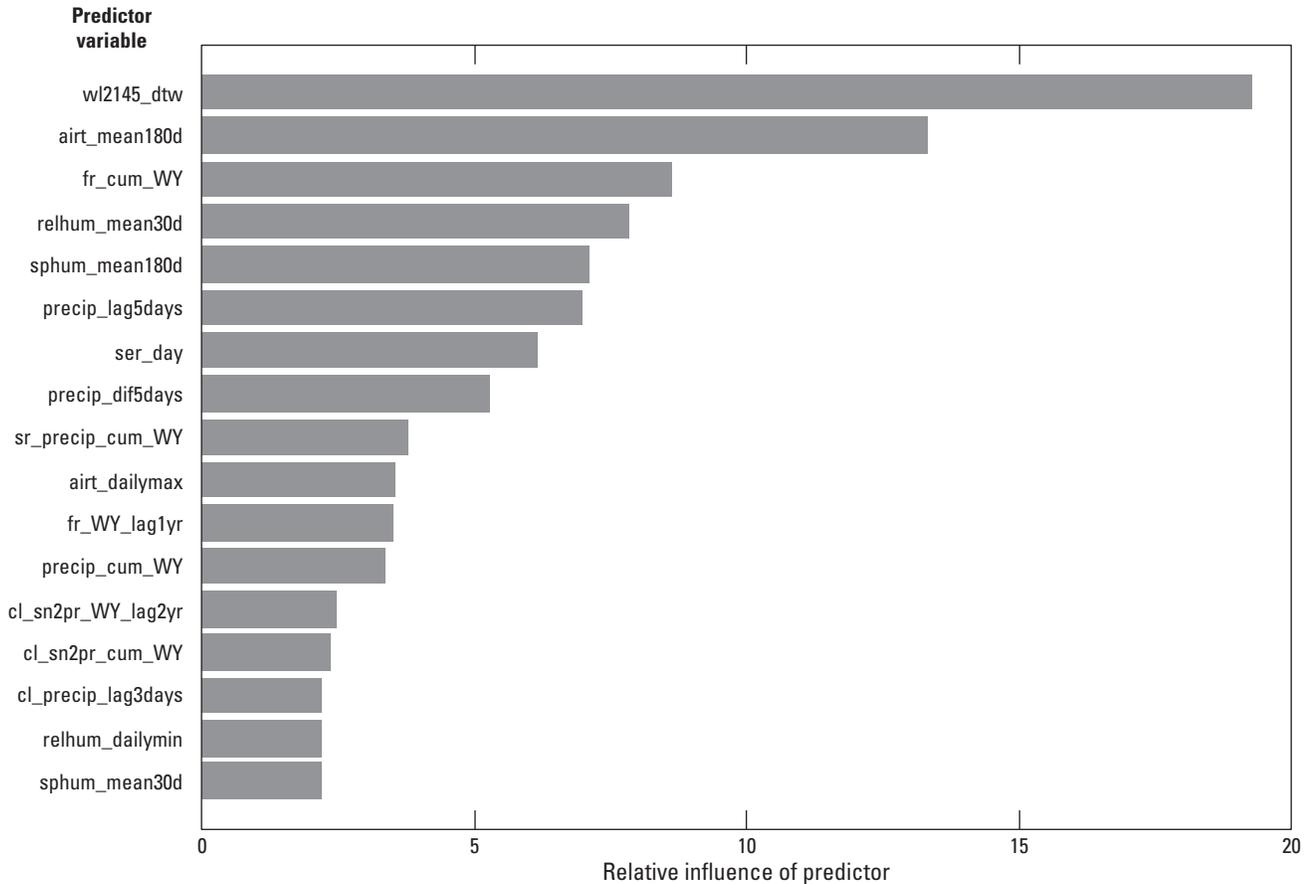


Figure 2. Percent relative influence, averaged over all trees, of the final selection of 17 predictor variables in the boosted regression trees model. [Variable names are defined in table 1. The relative influence of each predictor in an individual regression tree was quantified as the sum of the squared improvements at all splits determined by that predictor.]

Predictors Related To Interannual Variability In Water Table Elevation

The upper Klamath Basin is situated between the Pacific Northwest, where the strong correlation between winter precipitation and the El Niño-Southern Oscillation (ENSO) results in anomalously low streamflow associated with the El Niño phase of the oscillation, and the desert Southwest, where anomalously high streamflow is associated with the El Niño phase of the oscillation (Redmond and Koch, 1991). In this relatively narrow mid-latitude band, the Southern Oscillation Index (SOI) is not a useful climate index as it is in other parts of western North America. Kennedy and others (2009) demonstrated that, while the SOI is not particularly useful, the equatorial Pacific Ocean sea surface temperature gradient, expressed as the Trans-Niño Index (TNI), is an appropriate climate index to incorporate into streamflow forecast models with 1–4 month lead times. Thus, there is an established

teleconnection between climate, particularly as indexed by equatorial Pacific Conditions, and surface-water flow in the Upper Klamath Basin. However, because the purpose of this study is not to forecast streamflow but rather to “hindcast” streamflow, there is no lead time required. This suggests that leading indicators, particularly global scale interannual climate indices such as the Pacific Decadal Oscillation (Beebee and Manga, 2004; McCabe and Dettinger, 2002) and the Trans-Niño index (Kennedy and others, 2009), might not be as useful, in terms of explanatory power, as measurements of the localized response to these indicators.

The discharge measured at the Chiloquin streamgage during winter and spring months is dominated by surface and sub-surface runoff originating at the higher elevations of the watershed, but the time period of interest for this study is the summer and early autumn dry season. Runoff in response to precipitation events occurs during this season as well, but the discharge is dominated by groundwater-influenced baseflow.

Groundwater recharge in the upper Klamath Basin integrates the effects of climate over decadal-scale drought cycles; the effect on groundwater storage is in turn reflected in the groundwater discharge to streams as the water table and head gradients fluctuate in response to climate cycles (Gannett, 2007) and to long-term trends in regional groundwater pumping as the number of permitted groundwater-irrigated acres increased from 14,400 in 1980 to 25,000 in 2007, after which it did not increase. The response of the water table in the upper Sprague basin to these cycles is well-described by the depth-to-water in well KLAM2145 near Bly, Oregon (Oregon Water Resources Department, 2018a) that has been monitored at approximately quarterly intervals by the Oregon Water Resources Department (OWRD). The water level in this well responds to climate-driven variations in recharge, is correlated with precipitation at Crater Lake National Park Headquarters, and also demonstrates a long-term trend in response to multi-decadal increases in regional groundwater pumping. The water level in this well has been identified as a potentially useful predictor variable for seasonal forecasts of streamflow in the upper Klamath Basin (Risley and others, 2005; Kennedy and others, 2009). After removing sporadic drawdowns that were obviously caused by pumping in the well itself, the irregular data were interpolated to a daily interval using a Lowess smooth (Cleveland and Devlin, 1988), with the smoothing parameter set to utilize a window of about 3.5 years, which removed most variability associated with intra-annual fluctuations while retaining the climate and regional groundwater signal inherent in the inter-annual fluctuations (fig. 3).

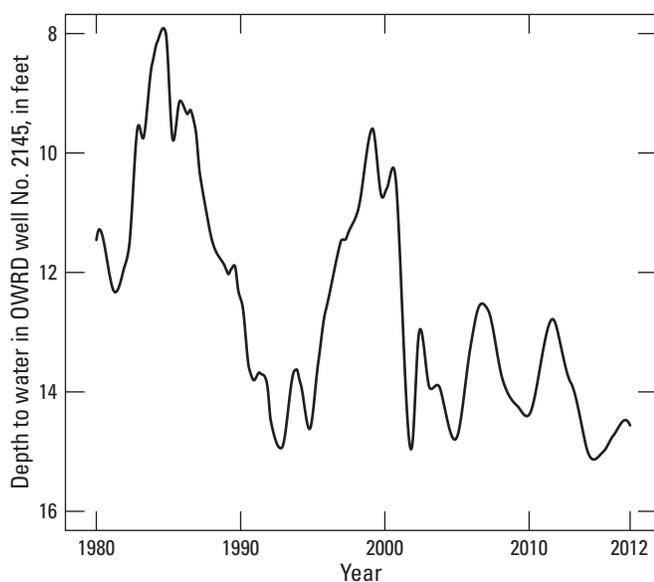


Figure 3. Depth-to-water in Oregon Water Resources Department well No. 2145, interpolated with a locally weighted scatterplot smoothing to filter out intra-annual variability, Sprague River Basin, Oregon. [OWRD, Oregon Water Resources Department]

Fall River is located in the Deschutes Basin to the north of the upper Klamath Basin, in a similar geologic and geographic setting to groundwater dominated streams in the upper Klamath Basin, but unaffected by regulation of irrigation. Streamflow in Fall River consists entirely of discharge from springs and the measurements at the OWRD streamgauge near La Pine, Oregon, reflect regional climate, as integrated over a few years (Risley and others, 2005). The streamflow in Fall River also was shown to be a useful predictor for seasonal forecasts of streamflow in the upper Klamath Basin (Risley and others, 2005; Kennedy and others, 2009). The daily mean discharge in Fall River is correlated with the depth-to-water in well 2145, which was found to have a greater relative influence in the BRT model, and therefore daily mean discharge was not retained in the model simplification process. A derived variable was created from the Fall River daily mean discharge by summing the flow over the entire water year, and lagging by 1, 2, and 3 years. The total water year flow with a lag of 2 and 3 years was removed from the model during model simplification, and the total water year flow with a lag of 1 year was retained.

The depth-to-water in well 2145 and Fall River discharge are correlated to the long-term cumulative departure from average precipitation measured at the Crater Lake National Park (CLNP) headquarters (Gannett, 2007). Conditions at CLNP clearly contain information about regional climate cycles, and a derived variable based on the precipitation at CLNP was calculated as the ratio of precipitation as snow and ice to the precipitation as rain. This variable is essentially an annual temperature index and was found to have a greater relative influence than precipitation alone. This derived variable was based on the total accumulation of both types of precipitation in the water year, lagged by 1, 2, and 3, years. The water year ratios with a lag of 1 and 3 years were removed from the model during model simplification, and the water year ratio with a lag of 2 years was retained.

Predictors Related To Current Water Year Weather And Hydrology

Current meteorological data to be used as predictor variables in a model for streamflow at a single streamgauge and at sub-weekly time scales is most useful at high spatial and temporal resolution. Because meteorological observation networks are rarely dense enough to provide the required spatial resolution in unpopulated areas, it is common practice to use some form of model-based “reanalysis” data that incorporates observations in a systematic and robust way and disaggregates the data into uniformly spaced spatial and temporal data. For short lead-time streamflow forecasting,

global re-forecasts that incorporate local observations, such as the re-forecast generated by the National Oceanic and Atmospheric Administration (NOAA; Kalnay and others, 1996), have proven very useful in many downscaling and streamflow forecasting studies, but the spatial resolution is coarse for watershed studies—about 210 kilometer at the equator. The NOAA publishes North American Land Data Assimilation System (NLDAS) re-analysis data with a resolution of 1/8 degree and temporal resolution from hourly to monthly to drive land surface hydrologic models (Xia and others, 2012). The University of Idaho gridMet data (Abatzoglou, 2013) melds temporal attributes from NLDAS with the physiographically based interpolation of observations from the Parameter-elevation Regressions on Independent Slopes Model (PRISM; Daly and others, 2008). The resulting product is a spatially and temporally complete, high-resolution (1/24 degree, or about 4-kilometer) gridded dataset of surface meteorological variables. The time series of daily gridMET data were downloaded as a weighted average over the entire Sprague River basin from the USGS geo-data portal (U.S. Geological Survey, 2018).

GridMET variables include precipitation (daily accumulation), air temperature (daily minimum and maximum), relative humidity (daily minimum and maximum), daily specific humidity, wind speed and direction, and incoming shortwave solar radiation. The entire suite of variables was tested for predictive power in the model, but after model simplification, solar radiation and wind speed were dropped. Precipitation was retained, lagged by 5 days, and the change in precipitation over the last 5 days also was retained. The 5-day lag in precipitation maximized the linear correlation of gridMET precipitation with streamflow at the Chiloquin streamgauge. Daily maximum air temperature and daily minimum relative humidity were retained. Additionally, several derived variables that were expected to capture intermediate-term memory in the system (1–6 months) related to hydrology and evapotranspiration, particularly the integration of conditions that sets antecedent soil moisture, were found to be useful: the current water year accumulation in precipitation, the mean of the maximum air temperature over the previous 180 days, the mean of the minimum relative humidity over the previous 30 days, and the mean over the previous 30 and 180 days of the specific humidity (table 1).

Previous studies have established the predictive value of the Natural Resource Conservation Service Snow Telemetry (SNOTEL) sites to prediction of streamflow in the upper Klamath Basin, particularly in the Sprague River, which receives seasonal snowmelt from higher elevations to the east (Risley and others, 2005; Kennedy and others, 2009). Cumulative precipitation in the current water year and snow water equivalent (SWE) observations from SNOTEL

sites at Taylor Butte and Summer Rim were tested for their explanatory value. The observations at Summer Rim were retained for the model. SWE observations at Summer Rim were lagged by 10, 30, and 60 days and, after the variable selection process, only the cumulative current water year precipitation was retained, indicating that the role of snow melt during the last part of the hydrograph recession that is included in the model's 500 ft³/s threshold is better captured by the retained GridMET variables, particularly air temperature.

Crater Lake National Park, while not in the Sprague River basin, is at the northern boundary of the upper Klamath Basin, and weather data from CLNP is important both for the information it contains regarding climate cycles, which has already been discussed and for the measurements of current, local storm systems coming from the west. Data from the NOAA meteorological station at Crater Lake National Park Headquarters is available from NOAA National Centers for Environmental Information (National Oceanic and Atmospheric Administration, 2017). Published data include daily minimum and maximum temperature, daily accumulation of precipitation as rain, daily accumulation of precipitation as snow, and snow depth. As with SWE at Summer Rim, because the model focuses on the low-flow season, the accumulation of snow at CLNP was not a useful predictor, and daily temperature was highly correlated with temperature in the gridMET dataset, so those data were dropped during the model simplification process. The daily accumulation of precipitation as rain at a 3-day lag was retained; the 3-day lag in precipitation maximized the linear correlation of CLNP precipitation with streamflow at the Chiloquin streamgauge. A derived variable based on the precipitation at CLNP was calculated as the ratio of cumulative precipitation in the current water year as snow and ice to the cumulative precipitation in the current water year as water. This variable is a form of temperature index and was found to have a greater relative influence than temperature or precipitation alone.

Building And Evaluating The Sprague River Discharge Boosted Regression Trees Model

Identification Of Model Start Date And Selection Of Training Data

The daily data from water years 1980 to 2012 comprise the potential training dataset. A more accurate model was obtained by concentrating the training on the lower streamflows dominated by base flow that are of the most interest when calculating the RIF. The model start date in each year, therefore, was based on streamflow rather than a fixed calendar date. The daily contribution of groundwater discharge

(base flow) to the measured streamflow was estimated using a streamflow partitioning method implemented in the USGS program PART (Rutledge, 1998). A streamflow of 500 ft³/s was chosen as a good indicator of the transition between runoff-dominated and baseflow-dominated streamflow. At this streamflow, coming at the end of the spring recession of the hydrograph, base flow made up at least 50 percent of the streamflow in each year in the training dataset, and in half of the years in the training dataset, base flow made up more than 99 percent of the streamflow.

For each modeled water year, the start date is the latest of March 2 or the first day following the day that the flow was 500 ft³/s for the last time in the water year. The end date is the last day in the water year, September 30. Because the length of the hydrograph between start and end dates varied each year, the number of data points available for training varied in each year as well. The total number of daily observations included in the 33 years of training data is 3,859. The daily data between the first model date (determined in the same way as for the training data) and the end of the water year from water years 2013 to 2016 comprise the regulated dataset that was used to quantify the RIF.

Statistical Methods

Model performance was evaluated with root-mean-square error (RMSE), mean absolute error (MAE), mean error (ME), and the correlation coefficient (R). In the case of cross-validation training and testing, the estimate of the error statistic was the mean of the statistic across the 10 folds of a cross-validation, and the standard error (SE) of the estimate was the standard deviation across the 10 folds of the cross-validation, divided by $\sqrt{10}$. In the case of testing with holdout data, the estimate of the error statistic for an individual holdout test was the mean of the statistic across the 10 folds of a cross-validation, and the standard error (SE) of the estimate was the standard deviation across the 10 folds of the cross-validation, divided by $\sqrt{10}$. The estimate of the error statistic for the entire ensemble of 33 holdout tests was the mean of the statistic across 10 folds of each individual holdout test, and the SE of the estimate was the standard deviation of the error statistic across 10 folds of each individual holdout test, divided by $\sqrt{330}$.

Normality in the distribution of holdout errors was tested with the Shapiro Wilk test, with p values greater than 0.05 indicating a normal distribution, with 95 percent confidence. A confidence threshold for the cumulative model errors through time based on the ensemble of holdout tests was determined with a Z-score at the desired level of confidence. First it was established that the cumulative errors were normally distributed with zero mean. Then, the detection limit for the cumulative difference between the model baseline and streamgauge measurements in regulation years that exceeded

model error was determined by $z_{\alpha}\sigma$ where z_{α} is the Z-score for a desired level of confidence α (1.64 for 95 and 1.28 for 90 percent), and σ is the standard deviation of the cumulative errors. A one-tailed test was used because the cumulative difference between baseline and measurements in regulation years should only be negative. A confidence interval for the cumulative model errors through time also was determined with a Z-score at the desired level of confidence, but using a two-tailed test, as cumulative error can be positive or negative.

All statistical test calculations were done using the basic stats package in the R computing environment (R Foundation for Statistical Computing, 2018).

Selection Of Model Parameters

Selection of model parameters is a balance between the goals of (1) minimizing the error between the simulated values of the response variable and the measurements of that variable, and (2) avoiding over-training the model such that it learns to simulate the noise superimposed on the real information contained in the response variable, which degrades the model's predictive power. In the BRT model, three techniques—controlling the learning rate, constraining the number and complexity of regression trees, and subsampling the data—are effective means of limiting model over-training. The careful choice of the parameters that define these techniques is therefore an important part of the development of an efficient model with good predictive performance. Parameter selection generally followed the procedure outlined by Elith and others (2008).

The process of parameter choice started with assuming reasonable values for the rate of sub-sampling and tree complexity. Values of the sub-sampling rate between 0.2 and 0.8, meaning that a randomly chosen 20–80 percent of the data is used to create each tree, effectively reduce absolute error in the response variable while also reducing computation time (Friedman, 2002). The sub-sampling rate in this study was set to 0.2, resulting in a random 20 percent of the daily data being chosen to create each tree. More typical values of the sub-sampling rate are between 0.4 and 0.6 (De'ath, 2007); a low value was used in this case because the daily values in the training dataset could support a higher sub-sampling rate without much loss of information. The results of training the model on the entire training dataset with a sub-sampling rate of 0.2 were compared to the results of training the model on a training dataset created by subsampling at regular intervals of 5 and 10 days, with a sub-sampling rate of 0.5. The performance of the model using a sub-sampling rate of 0.2 and no a priori sub-sampling was better, without evidence of increased overfitting. The interaction depth, or tree depth, was set to 3, allowing up to second-order reactions among the predictors. While the maximum theoretical depth is limited to the number of predictors, the best tradeoff between bias and variance is generally achieved at much smaller trees when the number of predictors is high (Friedman, 2001). The

12 Using Boosted Regression Trees to Quantify Instream Flow Resulting from Curtailment of Irrigation

minimum number of observations required to create a new split was set to 10, which was not more constraining on tree complexity than the interaction depth of 3. With these choices of parameters held constant, model training and testing was repeated across a range in values of learning rate and number of regression trees. The learning rate varied from 0.001 to 0.1, and the number of regression trees varied from 200 to 3,000.

A tenfold cross-validation (CV) approach was used to train and test the BRT model for optimization of the learning rate and number of trees. The entire training dataset was separated into 10 subsets by randomly sampling 3 years of the training dataset without replacement and assigning all of the daily values from those years to the same subset, 9 times, and assigning all of the daily values from the remaining 6 years to the tenth subset. Creating subsets of the training data in this way proved to be a more demanding test of the model than combining all years of data first and randomly sampling the daily values into 10 subsets. The model was then trained on each of the 10 possible combinations of 9 subsets of data and tested by using the learned model weights and coefficients to predict the values of the response variable in the subset of data that was left out of the training.

The estimate of the testing RMSE is shown for four values of the learning rate in figure 4. Higher values of the learning rate are omitted for clarity; higher values resulted in larger and more variable RMSE across the range in the

number of trees. The combination of a learning rate of 0.01 and 1,200 trees resulted in the smallest testing RMSE. The RMSE of the training data (not shown) was lower for the same combination of parameters (table 2) and decreased with increasing numbers of trees across the entire range tested; this is expected because a model can be trained to fit noise in the data as well as real information, but this over-training coincides with degraded performance in testing (De'ath, 2007). Once the "best" model with the lowest testing RMSE was identified, the 1 SE rule (Hastie and others, 2001) was applied to pick the "best simpler" model, which was defined as the lowest learning rate first, and then the minimum number of trees, that had an estimated RMSE within 1 SE of the best testing model (fig. 4). Applying the rule resulted in a final optimized learning rate of 0.005 and an optimized number of trees of 1,000 (table 2). The mean error (ME) and correlation coefficient R of the best and best simpler models are provided in table 2 as additional measures of performance.

Once final values of learning rate and number of trees were determined, additional tenfold CV tests were performed to evaluate model performance by increasing values of the sub-sampling rate (up to 0.7). Increasing the sub-sampling rate resulted in slightly degraded performance relative to the best model identified in the previous step, so this parameter remained at 0.2 for subsequent testing.

Table 2. Results of model training and testing on the entire 33-year training dataset, with a tenfold cross-validation.

[Model parameters held constant include bag fraction (0.2), minimum number of observations at terminal nodes (10). The best model had the minimum root mean square error, and the best simpler model was the final model chosen with a RMSE within 1 standard error of the best model. Units of RMSE and mean absolute error are cubic foot per second. Coefficient of determination R has no units. The proportion of observations is approximate, as the assignment to training and testing is by year rather than individual data points, and each year has a different number of data points. **Abbreviations:** MAE, mean absolute error; ME, mean error; R, correlation coefficient R; RMSE, root mean square error]

| | Best model in 10-fold cross-validation validation to choose learning rate and number of trees | | Best simpler model within 1 standard error of best in tenfold cross-validation to choose learning rate and number of trees | | Final model (best simpler model within 1 standard error of best in tenfold hold-out testing to choose interaction depth) | | |
|------------------------------|---|------------|--|------------|--|------------|-------------|
| | Training | Testing | Training | Testing | Training | Testing | |
| Days of training or testing | 3,473 | 386 | 3,473 | 386 | 3,473 | 386 | |
| Years of training or testing | 30 | 3 | 30 | 3 | 30 | 3 | |
| Sub-sampling rate | | 0.2 | | 0.2 | | 0.2 | |
| Learning rate | | 0.01 | | 0.005 | | 0.005 | |
| Number of trees | | 1200 | | 1000 | | 1000 | |
| Interaction depth | | 3 | | 3 | | 5 | |
| RMSE | +/- 1 SE | 24.02±0.12 | 42.80±2.62 | 31.41±0.18 | 45.04±2.75 | 26.65±0.13 | 41.87±2.65 |
| MAE | +/- 1 SE | 18.71±0.08 | 33.67±2.21 | 24.38±0.14 | 35.55±2.26 | 20.68±0.12 | 33.04±2.38 |
| ME | +/- 1 SE | -0.03±0.02 | 0.44±4.78 | -0.06±0.02 | 0.74±4.74 | -0.01±0.01 | -1.24 ±3.43 |
| R | +/- 1 SE | 0.94±0.00 | 0.76±0.06 | 0.90±0.00 | 0.73±0.06 | 0.93±0.00 | 0.78 ±0.05 |

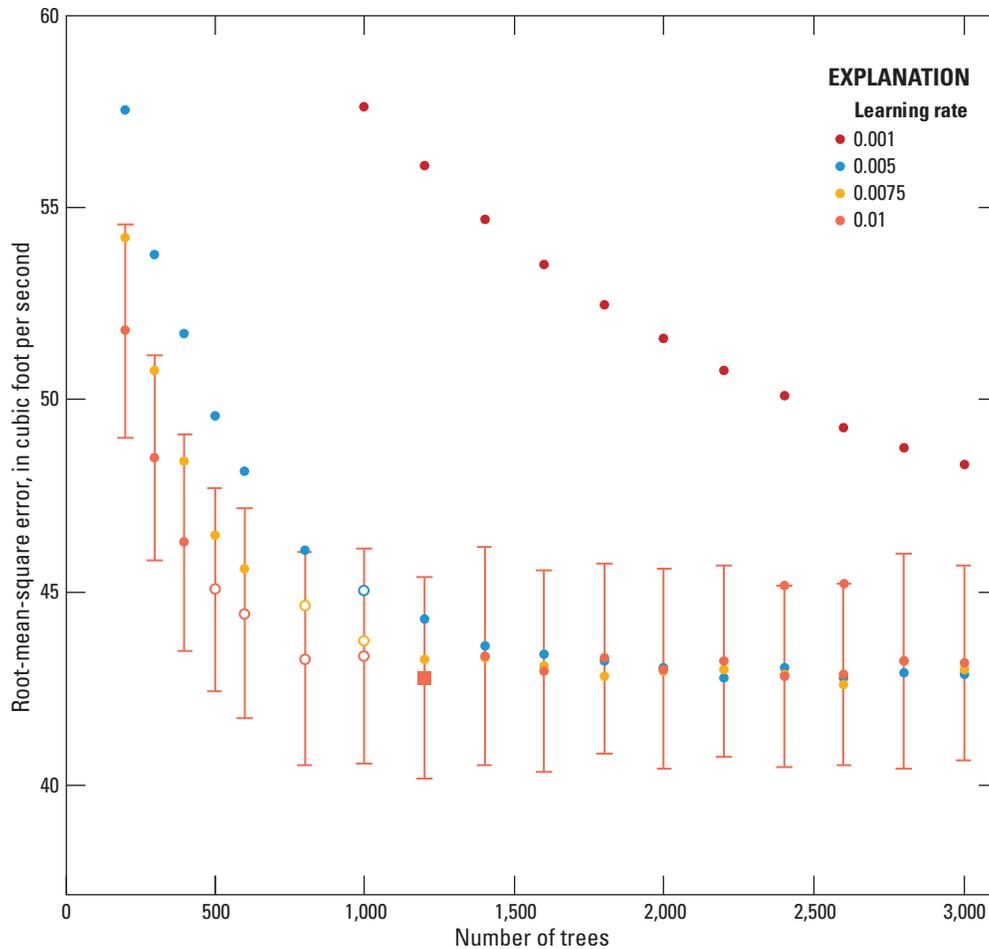


Figure 4. Relation between model complexity as defined by the number of trees and model performance as measured by root-mean-square error of the tenfold cross-validation testing prediction for four values of learning rate. [Bars indicating 1 standard error are provided for a learning rate of 0.01 only. The best model is indicated by the filled square; best simpler models by open circles.]

Model Optimization And Testing With Holdout Data

Once the learning rate and number of trees were finalized, the model was further optimized and tested by withholding a subset of the data from training and examining model performance in predicting the holdout data. An individual holdout test was implemented by removing all data corresponding to a single year from the training dataset, as this was the best way to evaluate model errors during curtailment years. An ensemble of 33 individual holdout tests was created by withholding each of the years in the entire training dataset in turn. Again, a tenfold cross validation procedure was used in each holdout test. The procedure was the same as that used for optimizing parameters, except that a single year of data was removed from the training dataset before it was partitioned into 10 mutually exclusive subsets. The model was trained on each of the 10 possible combinations of 9 subsets of data and then used to predict the response variable in the

holdout year, resulting in 10 predictions and 10 estimates of error statistics for each holdout year.

The ensemble of holdout tests was used to optimize the interaction depth of the trees in the model. Each of the 33 holdout tests was repeated while varying the interaction depth between 2 and 14, and the ensemble mean of RMSE (330 estimates) was used to identify the optimal interaction depth (fig. 5). Minimum ensemble RMSE was associated with an interaction depth of 12, but interaction depths as low as 5 resulted in an ensemble RMSE within 1 SE of the RMSE at an interaction depth of 12.

The tenfold CV training and testing was repeated with an interaction depth of 5; the results showed that increasing the interaction depth from 3 to 5 yielded an improvement in testing RMSE and MAE compared to the testing results of the best and best simpler (1 SE) models obtained previously. The testing RMSE of the best, best simpler, and final models was 42.80, 45.04, and 41.87 ft³/s, respectively, and the MAE of the three models was 33.67, 35.55, and 33.04 ft³/s, respectively (table 2). The final, “optimal” set of model parameters

14 Using Boosted Regression Trees to Quantify Instream Flow Resulting from Curtailment of Irrigation

included, therefore, a learning rate of 0.005, 1,000 trees, and a maximum interaction depth of 5, and this set of parameters was used in all subsequent model training, testing, and prediction.

Having identified the optimal set of model parameters, the performance of the model on holdout data could be compared to performance on testing data in a cross-validation. The overall mean of the performance metrics was comparable to the metrics obtained during CV testing (a RMSE of 41.03 ft³/s in table 3, in comparison to a RMSE of 41.87 ft³/s during CV testing of the final model with the same parameters in table 2, for example), indicating that the performance of the model in predicting holdout data was not degraded compared to testing with data that was included in the CV training and testing. The RMSE, MAE, and ME of the ensemble of 33 holdout tests were normally distributed as determined by a Shapiro-Wilk test ($p > 0.05$ in all cases). The 33 years were ranked according to the average flow during model dates (starting after flows reach 500 ft³/s in the spring to the end

of the WY), and the data were split into two groups. The RMSE, MAE, and ME of these two groups were significantly different as determined by a Kruskal-Wallis test (p less than 0.05 in all cases), meaning the magnitude of absolute errors measured by RMSE and MAE was higher, overall, in high flow years than in low flow years, which is not unexpected if the errors are proportional to the magnitude of the response variable. The ME was more likely positive in low flow years and negative in high flow years, although standard deviation of ME across all the holdout years is large, and the relation with flow is most apparent at the extremes—less than 200 or more than 300 ft³/s—of the range in flows, and ME of both signs is scattered across the middle of the range (fig. 6). The correlation coefficient was not significantly different between the high-flow and low-flow groups, although the highest R values were obtained in high-flow years and the lowest in low-flow years (fig. 7). Both R and ME were weakly but

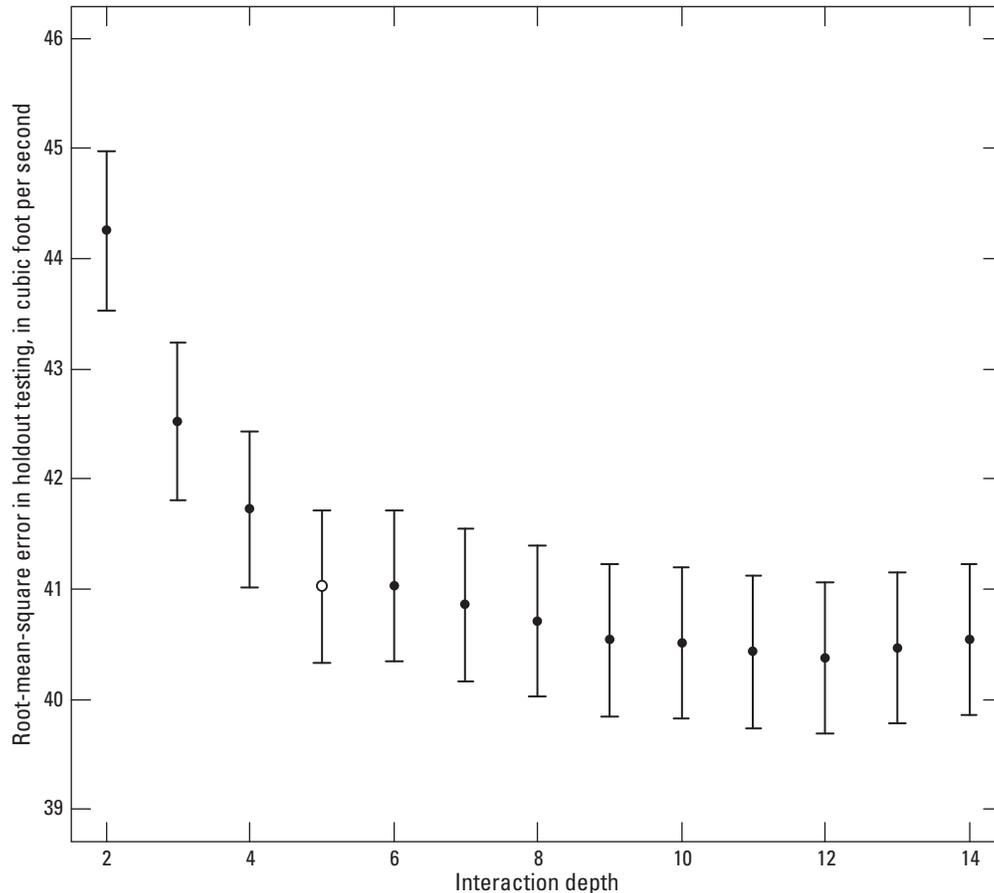


Figure 5. Relation between model complexity as defined by the maximum interaction depth of trees and model performance as measured by root-mean-square error of the holdout testing predictions. [Bars indicate 1 standard error. The optimal model is indicated by the open circle.]

Table 3. Results of model testing on holdout data.

[Units of root mean square error and mean absolute error are cubic foot per second. Correlation coefficient R has no units. Flow rank is based on the daily streamflow, averaged over model dates in each year. The rank is out of 33 years: 1980–2012, inclusive. Abbreviations: MAE, mean absolute error; ME, mean error; RMSE, root mean square error; SE, standard error]

| | Year | Flow rank | Number of observations | RMSE ± SE | R ± SE | ME ± SE | MAE ± SE |
|---------------------------------|------|-----------|------------------------|--------------|-------------|---------------|--------------|
| Aggregated means | | | | | | | |
| 33 years | – | – | 3859 | 41.03 ± 0.67 | 0.72 ± 0.01 | -0.40 ± 1.44 | 33.57 ± 0.58 |
| Years with flow rank 17-33 | – | – | 2152 | 36.60 ± 0.91 | 0.69 ± 0.01 | 9.34 ± 1.48 | 29.66 ± 0.75 |
| Years with flow rank 1-16 | – | – | 1707 | 45.73 ± 0.84 | 0.76 ± 0.01 | -10.76 ± 2.24 | 37.72 ± 0.75 |
| Individual years: best metrics | | | | | | | |
| Smallest positive ME | 2002 | 26 | 127 | 27.14 ± 0.45 | 0.69 ± 0.01 | 1.07 ± 1.79 | 22.39 ± 0.33 |
| Smallest negative ME | 2007 | 25 | 130 | 27.46 ± 0.35 | 0.73 ± 0.01 | -0.50 ± 0.61 | 21.82 ± 0.27 |
| Smallest RMSE | 2005 | 23 | 110 | 17.72 ± 0.81 | 0.83 ± 0.01 | 1.61 ± 1.57 | 14.40 ± 0.81 |
| Highest R | 1986 | 6 | 108 | 25.38 ± 0.62 | 0.95 ± 0.00 | 5.35 ± 1.35 | 18.72 ± 0.68 |
| Individual years: worst metrics | | | | | | | |
| Largest positive ME | 1998 | 10 | 82 | 57.93 ± 1.79 | 0.78 ± 0.01 | 49.49 ± 2.09 | 50.58 ± 1.82 |
| Largest negative ME | 2006 | 9 | 103 | 57.72 ± 1.13 | 0.35 ± 0.01 | -51.30 ± 1.32 | 51.37 ± 1.31 |
| Largest RMSE | 1987 | 5 | 142 | 65.87 ± 1.17 | 0.68 ± 0.01 | -51.80 ± 1.59 | 54.48 ± 1.34 |
| Lowest R | 2003 | 31 | 111 | 50.34 ± 0.46 | 0.25 ± 0.01 | 24.25 ± 0.85 | 46.46 ± 0.51 |

significantly correlated with flow (Spearman $\rho=0.36$ and -0.52 , respectively; $p<0.05$ in both cases).

The results of 8 of the 33 individual holdout tests with the best and worst performance metrics are reported in table 3. The response variable and model predictions for 4 of these years are shown in figure 8. This selection of individual holdout tests demonstrates that performance metrics vary greatly among individual years and that correlation with flow is weak. Three of the four best metrics characterized years from the lower half of the flow distribution, and three of the

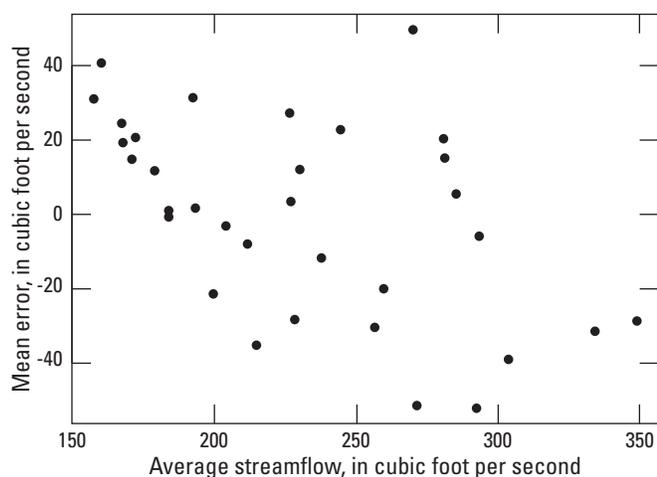


Figure 6. Mean error in 33 holdout years, as a function of the measured daily streamflow, averaged over model dates in each year.

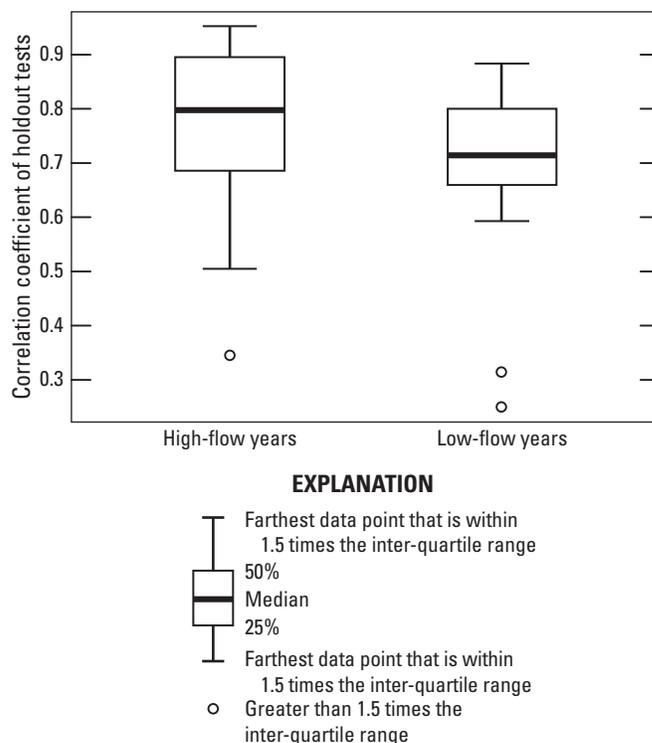


Figure 7. Correlation coefficient in 33 holdout years, for years in which daily streamflow, averaged over model dates, was high (rank 1-16) and low (rank 17-33).

four worst metrics characterized years from the upper half of the flow distribution. The highest R (0.95) was found for a year of high flow (1986; rank 6), and the lowest R (0.25) was found for a year of low flow (2003; rank 31). However, the third lowest R of 0.35 was found for a high flow year, 2006 (rank 9). Years with similar flow ranks can have quite different performance metrics, as demonstrated by the years shown in figure 8. The year characterized by highest R (0.95) was 1986 (flow rank 6), and the year of largest RMSE (65.9 ft³/s) was 1987 (flow rank 5), both high-flow years. The year with smallest RMSE (17.72 ft³/s) was 2005 (rank 23), and the year with the lowest R (0.25) was 2003 (rank 31), both low-flow years.

When model performance is measured in terms of overall bias in each year as measured by ME, the best metrics were found for low-flow years; the smallest positive and

negative ME were found for 2002 and 2007, ranked 26 and 25, respectively. Two years of similarly high flow were characterized by the largest positive ME (49.49 ft³/s in 1998, rank 10) and the largest negative ME (-51.30 ft³/s in 2006, rank 9). The bias results of the holdout tests have important implications for using the BRT model to quantify instream flow, because they indicate that, while the overall bias of the model in any individual year that was not used in training can be positive or negative, and as high as about 50 ft³/s, when the ensemble of years of holdout tests is considered, the errors are normally distributed, and the overall bias is approximately zero. The 33 holdout tests therefore provide an estimate of the distribution of model errors in predicting the baseline streamflow in years of irrigation curtailment.

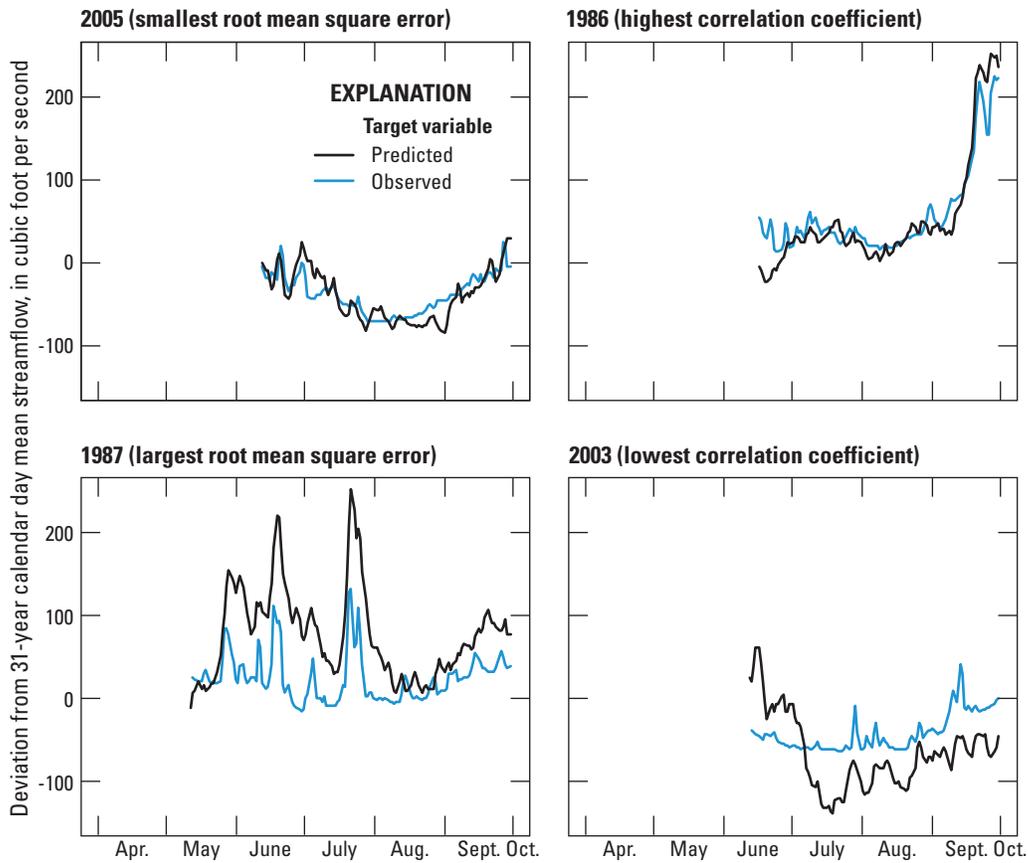


Figure 8. Observed and predicted target variable in 4 holdout years, Sprague River at Chiloquin, Oregon. [Years characterized by smallest root mean square error (2005), highest correlation coefficient R (1986), largest room mean square error (1987), and lowest correlation coefficient R (2003).]

Using The Boosted Regression Trees Model To Quantify Cumulative Instream Flow Resulting From Curtailment Of Irrigation

Water in the basin is allocated in terms of volume rather than streamflow, and it is convenient to quantify the RIF in terms of a cumulative volume, as determined by subtracting the BRT-predicted baseline from the streamgage measurements and accumulating those deviations through time. The results of the holdout tests were used as the basis for determining the “detection limit” of the RIF, as follows. As discussed above, the errors of the prediction with respect to the streamgage were normally distributed across 33 pre-regulation years, as was the accumulation of those errors over a fixed period of time in each year. This was demonstrated by cumulatively summing the difference between the daily prediction and the daily average flow from the streamgage, starting from July 1 (fig. 9). In some years, the calculation

begins at a later date because the model did not start until some days later, when the transition to 500 ft³/s occurred. During years before irrigation regulation, the cumulative difference in streamflow between the BRT prediction and the observations would be zero in each year if the model were perfect. Instead, error in the model predictions results in a non-zero cumulative difference in streamflow, and the envelope delimited by the highest and lowest values increases with time away from the starting date of the calculation. These non-zero values of cumulative error provide the distribution against which the cumulative differences in regulation years; that is, the BRT estimate of RIF, can be compared. Also shown on figure 9 is the cumulative difference between the BRT-predicted baseline streamflow and the streamgage in 4 regulation years. In a regulation year, the cumulative difference is the sum of the accumulated error and any instream flow resulting from irrigation curtailment. When the cumulative difference is negative, the streamgage measures more flow than the predicted baseline. Two regulation years appear to be well within the envelope of expected cumulative error after 4 months, and two are outside or at the limit of expected cumulative error after 4 months.

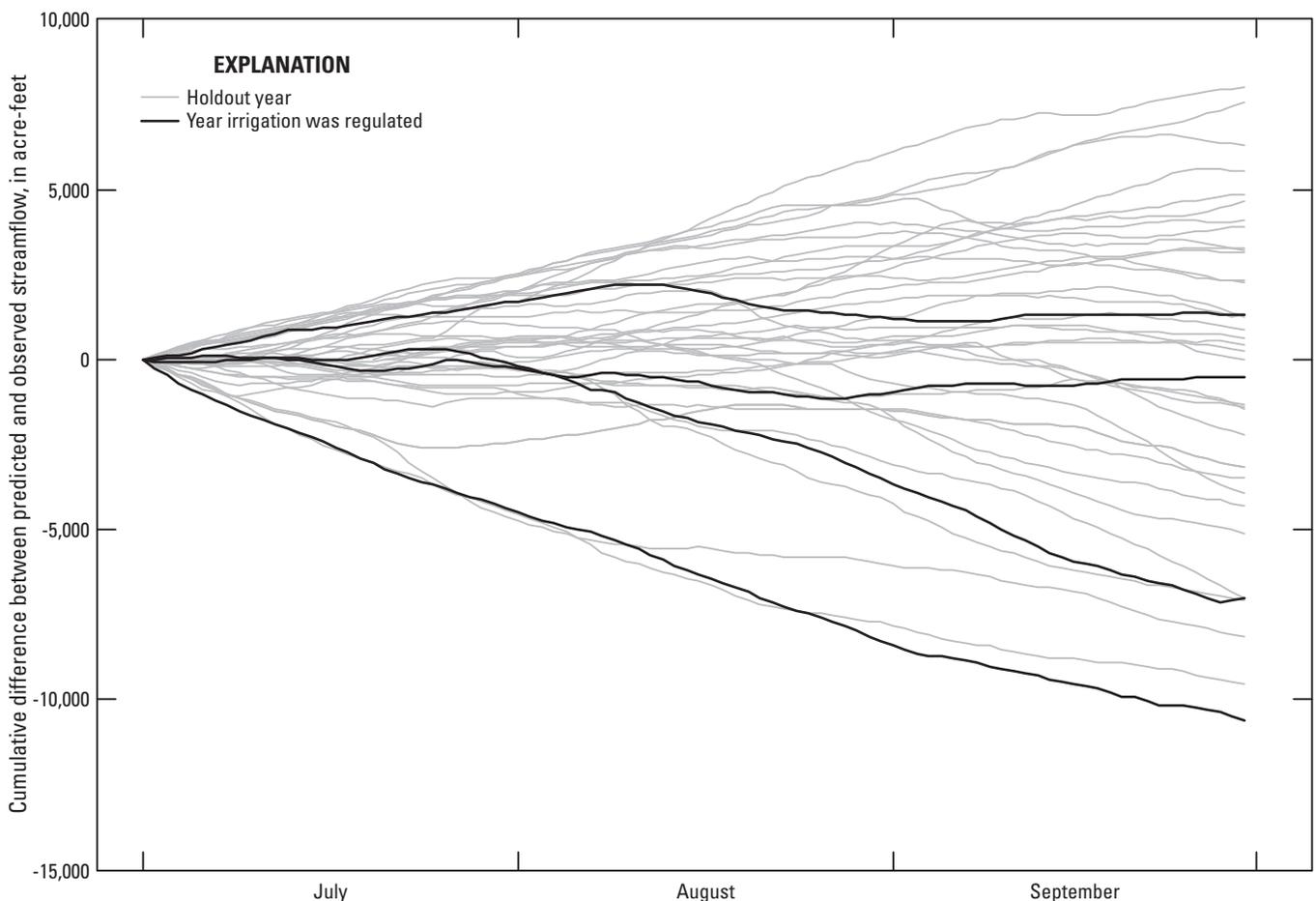


Figure 9. Cumulative difference between predicted and measured streamflow in each of 33 holdout years and in 4 years when irrigation was regulated, Sprague River at Chiloquin, Oregon, starting on July 1.

The cumulative error is normally distributed with a mean of zero and a standard deviation that increases through time, for all but a few days between July 1 and September 30. This distribution provides a measure of the cumulative model error that can be expected in any given year and applied to the regulation years, and it provides a measure of the cumulative difference in streamflow that could be attributed to model error in the prediction of a baseline hydrograph, rather than a real difference attributable to irrigation curtailment. To illustrate how this would work in practice, the distribution of cumulative errors, starting on July 1 and ending on September 30, is shown in figure 10, with thresholds for 90 and 95

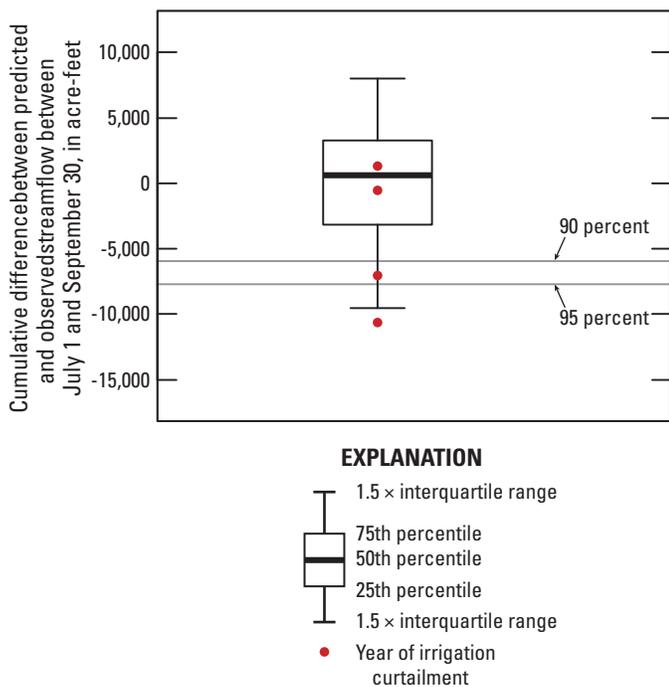


Figure 10. Boxplot showing cumulative difference between predicted and measured streamflow in 33 holdout years, Sprague River at Chiloquin, Oregon, starting July 1 and ending September 30. [The accumulation of the difference between the predicted baseline and streamgage measurements for 4 post-regulation years are superimposed on the boxplot. Thresholds indicate the levels at which surplus water from curtailment of irrigation can be identified and quantified, with 95 and 90 percent confidence.]

percent confidence for detection of the RIF shown. Based on this assessment, 1 regulation year is well outside of the 95 percent confidence threshold, and one is outside of the 90 percent confidence threshold. In 2 regulation years, the cumulative difference between the predicted reference and the gaged streamflow was not distinguishable from model error.

Because the date on which irrigation starts to be limited by holders of senior water rights varies from year to year, the accumulation of flow differences also should begin on a date that is determined by a unique timetable in each year. Increasing thresholds of model error can be calculated for days beyond the start of regulatory action. The calculation of the thresholds, which is based on the 33 holdout years, varies by a small amount depending on when the calculation starts. In figure 11, several curves, starting on June 1 and 15, July 1 and 15, and August 1 and 15, have been averaged and smoothed with a running average, to produce threshold curves along an arbitrary timeline starting with regulation action. In figure 11, more water in the stream from irrigation curtailment is shown as a positive value.

Not only does regulation take place on a different timeline each year, but also the amount of acreage affected varies through time, and can both increase and decrease multiple times within a year. The most important features of this timeline for the 4 regulation years have been reconstructed, using information from OWRD (Oregon Water Resources Department, 2018b), and are shown with streamgaged streamflow and the BRT-predicted baseline in figure 12. Regulation affects all acreage with rights dated on or after the date of regulation; thus, the farther back in time regulated water rights are dated, the more acreage is affected. For example, regulation starting at 1864 affects more acreage than regulation starting at 1905, because it includes all acreage with rights dated between 1864 and 1905. Regulation starting at time immemorial affects all acreage irrigated with surface water (and since 2014 some acreage irrigated with groundwater) in the basin. The measured streamflow in 4 regulation years shows some deviations from the predicted baseline that appear coordinated with regulation activities, most obviously in 2013, when the observations diverge from and are higher than the predicted baseline hydrograph soon after the date on which regulation action began (fig. 13). 2013 was the only year in which regulation of all surface-water rights in the basin was imposed early and maintained through the rest of the water year. Other times when divergence between the measured streamflow and the predicted reference streamflow seem coordinated with regulation activities include mid-July of 2015 and early August of 2016 (fig. 13).

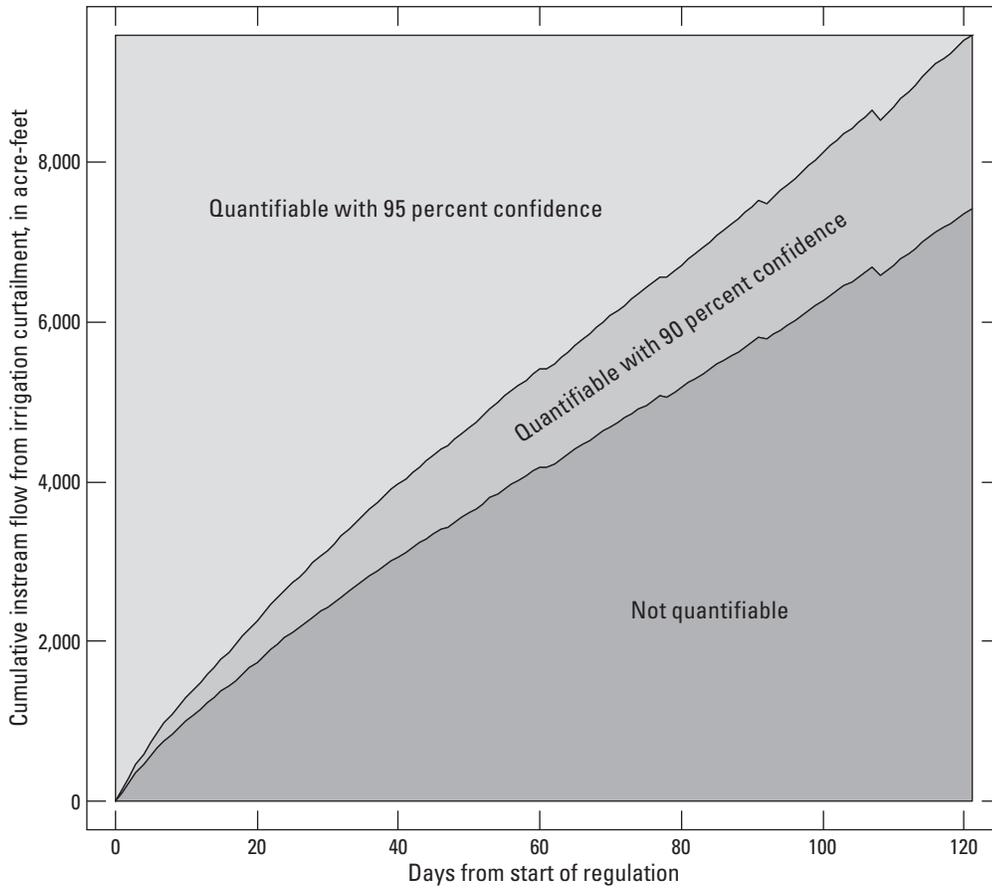
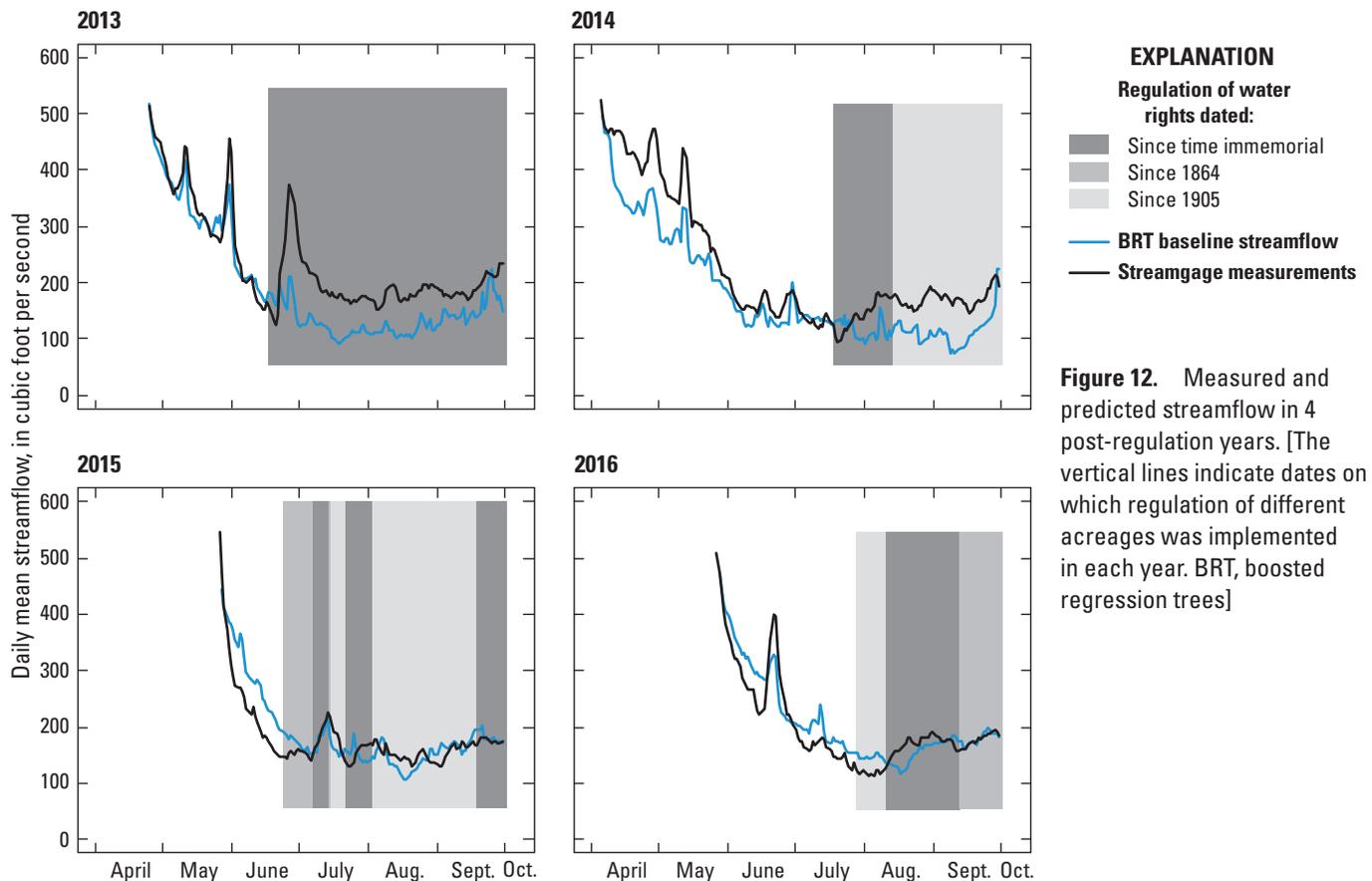


Figure 11. Thresholds for quantification of cumulative instream flow resulting from the curtailment of irrigation, as a function of days from the start of regulation action, Sprague River at Chiloquin, Oregon. [More water in the stream is shown as a positive quantity.]



EXPLANATION
Regulation of water rights dated:

- Since time immemorial
- Since 1864
- Since 1905

— BRT baseline streamflow
 — Streamgauge measurements

Figure 12. Measured and predicted streamflow in 4 post-regulation years. [The vertical lines indicate dates on which regulation of different acreages was implemented in each year. BRT, boosted regression trees]

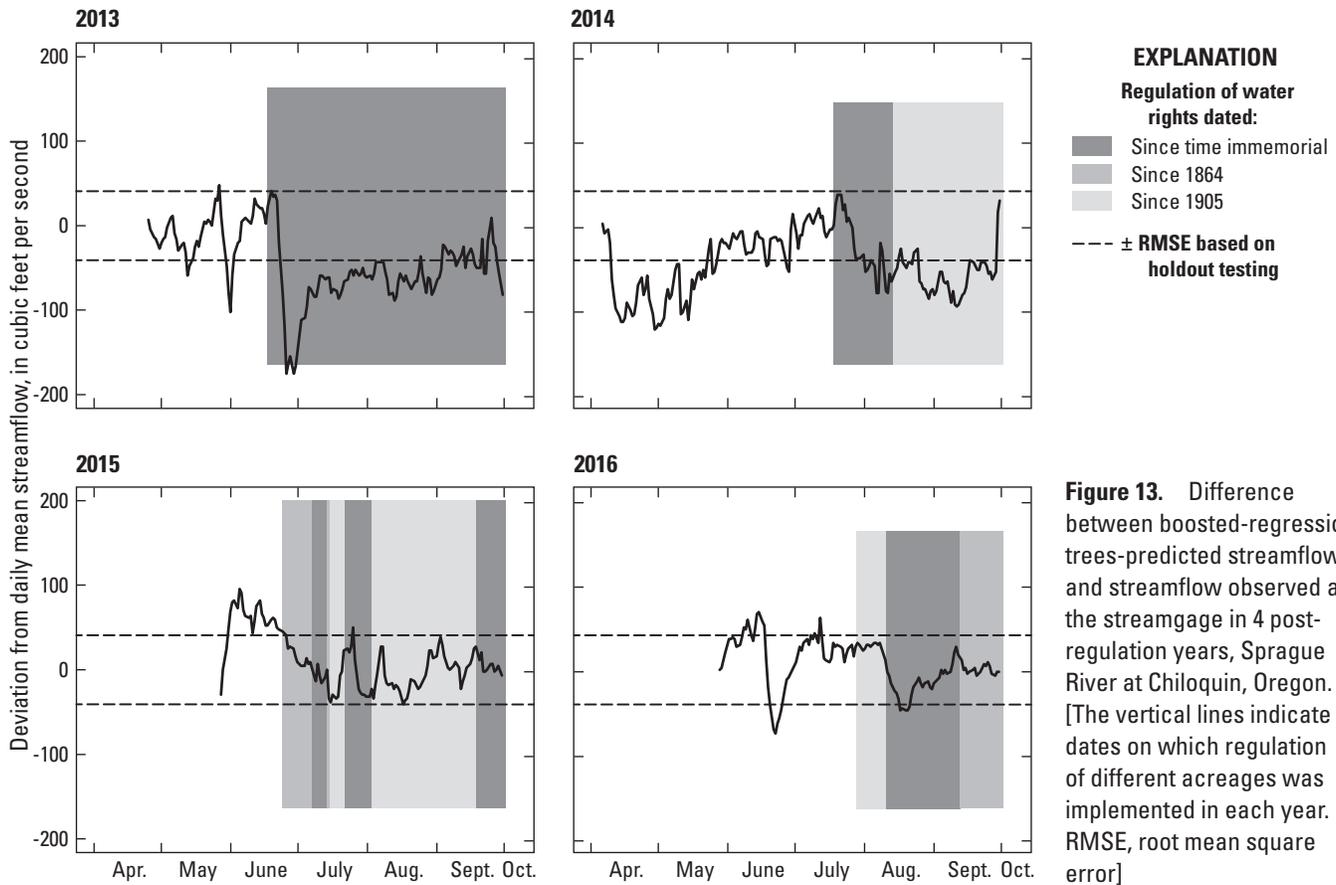


Figure 13. Difference between boosted-regression-trees-predicted streamflow and streamflow observed at the streamgage in 4 post-regulation years, Sprague River at Chiloquin, Oregon. [The vertical lines indicate dates on which regulation of different acreages was implemented in each year. RMSE, root mean square error]

Generally, however, it is difficult to identify deviations in daily streamflow that correspond to regulation activities, because of the inherent variability in water management decisions made by individual users, inconsistencies in regulation compliance, and regulation actions that are variable in time and location. The effect of regulation is more reliably identified as an accumulation of difference with respect to the baseline over time, which is shown in figure 14 for the 4 regulation years. The results shown are broadly consistent with what is known about how regulation was imposed in the 4 years. Regulation of all surface water rights began earliest in 2013 (June 17) and was maintained for the rest of the water year, which resulted in the greatest accumulation of RIF by the end of the water year—12,600±8,400 acre-ft (table 4). In 2014, regulation of groundwater rights was added to the regulation of surface-water rights, but started weeks later on (July 19) and therefore the accumulation by the end of the water year was only 6,900±6,200 acre-ft. In 2015, regulation

started in June as in 2013 (June 24), but acreage was regulated on and off throughout the season, and the regulation of the maximum acres (to time immemorial) started in July and was intermittent. In 2016, regulation started the latest of the 4 years (July 27), and regulation to time immemorial started on August 11 and lasted only 19 days, so the accumulation of RIF was minimal. The cumulative RIF in 2015 and 2016 was 78±7,860 and 73±5,630 acre-ft, respectively (table 4). In order to assess the reliability of these estimates, the cumulative RIF can be compared with what is known about the accumulation of model error (fig. 11). The results are provided in table 4, where it is shown that in 2013 and 2014, the cumulative instream flow from irrigation curtailment exceeded the 95 percent confidence threshold for quantifying the flow. Under maximally regulated conditions, the RIF is a significant fraction of the total streamflow at the streamgage—31 percent in 2013 and 29 percent in 2014. The values in 2015 and 2016 were not distinguishable from model error (table 4).

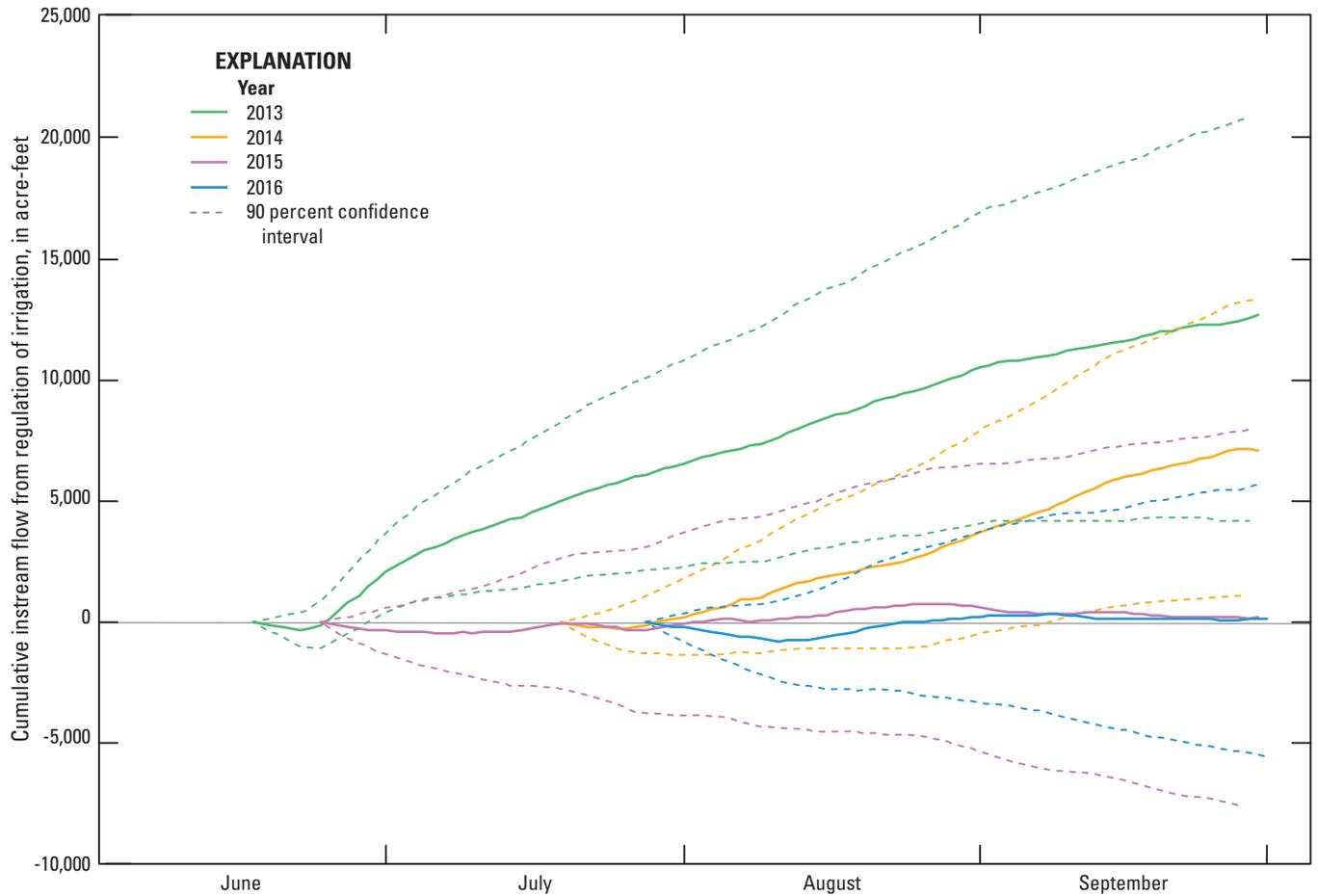


Figure 14. Model estimate of the cumulative instream flow made available through the end of September by irrigation curtailment in post-regulation years, Sprague River at Chiloquin, Oregon. [The accumulation begins each year on the date on which regulatory action started in that year.]

Table 4. Cumulative instream flow from irrigation curtailment (RIF) at the Chiloquin gage on the Sprague River, calculated in 4 post-regulation years.

[Also shown is the total volume of flow past the gage during the regulation time period.]

| Year | Days under regulation | Volume in acre-feet | Confidence threshold (percent) | Total volume in acre-feet |
|------|-----------------------|---------------------|--------------------------------|---------------------------|
| 2013 | 105 | 12,600 | 95 | 40,566 |
| 2014 | 73 | 6,900 | 95 | 23,532 |
| 2015 | 98 | 78 | na | 31,260 |
| 2016 | 65 | 73 | na | 20,787 |

Conclusion

The quantification of the instream flow resulting from the curtailment of irrigation in the upper Klamath Basin is a challenging problem, but not unsolvable. Ideally, the volume of water should be calculated relative to what streamflow would have been, in exactly the same weather and climate context, and under exactly the same antecedent conditions, without limits placed on irrigation (the baseline conditions). While there are several approaches to the problem, most require that the baseline conditions be taken from observations made in hydrologically similar and un-regulated years in the past, which means that the baseline conditions cannot be entirely correct for the regulated year in the present. Boosted regression trees offer the option of re-creating a unique baseline for any given year in current, regulated conditions, without requiring spatially explicit land cover and soil characteristics, including crop types and associated irrigation practices, that are critical to deterministic calculation of evapotranspiration.

We were able to demonstrate that BRT can be used to calculate the RIF, but a crucial component is the assessment of model error. This was done by creating an ensemble of holdout tests based on each individual year in our training data, which mimicked the way that the model will be used to predict a baseline in a given regulated year. From this ensemble of holdout tests, the distribution of the accumulation of model errors over time could be calculated, and the threshold for identification of RIF could be determined at a desired level of confidence. For example, at 30, 60, 90, and 120 days from the start of regulation, the threshold for identifying RIF was calculated to be 3,140, 5,420, 7,440, and 9,520 acre-ft, respectively, with a confidence of 95 percent. RIF in 2 regulation years—2013 and 2014—exceeded the 95 percent threshold and made up about one third of the total flow past the streamgage during the regulation period, in spite of the fact that irrigated acres make up less than 7 percent of the acres in the basin.

The relative influence of the model predictors underscores that the streamflow in the Sprague River during irrigation season responds more strongly to the climate signals that influence baseflow—the information embedded in the depth-to-water at a well in the basin not influenced by pumping, and the streamflow in the Fall River, a groundwater-dominated system in an adjacent basin—than to precipitation-induced runoff. This is the case even though the Sprague River

is runoff-dominated at other times of the year and has a strong seasonal snowmelt signal. The relative influence of the depth-to-water in the well was 19.24 (out of 100), which made it the predictor with the most influence on the model results. The streamflow in Fall River had a combined relative influence, based on two derived predictors, of 12.11, making it the third most influential predictor. The second most influential predictor was air temperature with a combined relative influence of 16.85, indicating primarily the strong influence of temperature on evapotranspiration.

The BRT approach provides information that can inform decisions about management of water on an operational basis. All of the predictors used are readily available and updated continuously. The model is created in the R open-source computing environment.

References Cited

- Abatzoglou, J.T., 2013, Development of gridded surface meteorological data for ecological applications and modelling: *International Journal of Climatology*, v. 33, no. 1, p. 121–131. [Also available at <https://doi.org/10.1002/joc.3413>.]
- Beebee, R.A., and Manga, M., 2004, Variation in the relationship between Snowmelt Runoff in Oregon and ENSO and PDO: *Journal of the American Water Resources Association*, v. 40, no. 4, p. 1011–1024. [Also available at <https://doi.org/10.1111/j.1752-1688.2004.tb01063.x>.]
- Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.J., 1984, *Classification and regression trees*: Boca Raton, Florida, Chapman & Hall/CRC, 353 p.
- Cannon, A.J., and Whitfield, P.H., 2002, Downscaling recent streamflow conditions in British Columbia, Canada using ensemble neural network models: *Journal of Hydrology (Amsterdam)*, v. 259, no. 1, p. 136–151. [Also available at [https://doi.org/10.1016/S0022-1694\(01\)00581-9](https://doi.org/10.1016/S0022-1694(01)00581-9).]
- Cleveland, W.S., and Devlin, S.J., 1988, Locally weighted regression—An approach to regression analysis by local fitting: *Journal of the American Statistical Association*, v. 83, no. 403, p. 596–610. [Also available at <https://doi.org/10.1080/01621459.1988.10478639>.]

- Daly, C., Halbleib, M., Smith, J.I., Gibson, W.P., Doggett, M.K., Taylor, G.H., Curtis, J., and Pasteris, P.P., 2008, Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States: *International Journal of Climatology*, v. 28, no. 15, p. 2031–2064. [Also available at <https://doi.org/10.1002/joc.1688>.]
- De'ath, G., 2007, Boosted trees for ecological modeling and prediction: *Ecology*, v. 88, no. 1, p. 243–251. [Also available at [https://doi.org/10.1890/0012-9658\(2007\)88\[243:BTFEMA\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2007)88[243:BTFEMA]2.0.CO;2).]
- Elith, J., Leathwick, J.R., and Hastie, T., 2008, A working guide to boosted regression trees: *Journal of Animal Ecology*, v. 77, no. 4, p. 802–813. [Also available at <https://doi.org/10.1111/j.1365-2656.2008.01390.x>.]
- Erdal, H.I., and Karakurt, O., 2013, Advancing monthly streamflow prediction accuracy of CART models using ensemble learning paradigms: *Journal of Hydrology (Amsterdam)*, v. 477, p. 119–128. [Also available at <https://doi.org/10.1016/j.jhydrol.2012.11.015>.]
- Fienen, M.N., Nolan, B.T., and Feinstein, D.T., 2016, Evaluating the sources of water to wells—Three techniques for metamodeling of a groundwater flow model: *Environmental Modelling & Software*, v. 77, p. 95–107. [Also available at <https://doi.org/10.1016/j.envsoft.2015.11.023>.]
- Friedman, J.H., 2001, Greedy Function Approximation—A Gradient Boosting Machine: *Annals of Statistics*, v. 29, no. 5, p. 1189–1232. [Also available at <https://doi.org/10.1214/aos/1013203451>.]
- Friedman, J.H., 2002, Stochastic gradient boosting: *Computational Statistics & Data Analysis*, v. 38, no. 4, p. 367–378. [Also available at [https://doi.org/10.1016/S0167-9473\(01\)00065-2](https://doi.org/10.1016/S0167-9473(01)00065-2).]
- Gannett, M.W., Lite, K.E., Jr., La Marche, J.L., Fisher, B.J., and Polette, D.J., 2007, Ground-water hydrology of the upper Klamath Basin, Oregon and California: U.S. Geological Survey Scientific Investigations Report, p. 84., [Also available at <https://doi.org/10.3133/sir20075050>.]
- Ghosh, S., and Mujumdar, P.P., 2008, Statistical downscaling of GCM simulations to streamflow using relevance vector machine: *Advances in Water Resources*, v. 31, no. 1, p. 132–146. [Also available at <https://doi.org/10.1016/j.advwatres.2007.07.005>.]
- Greenwell, B., Boehmke, B., Cunningham, J., and GBM Developers, 2018, *gbm—Generalized boosted regression models (version 2.1.4)*: R Foundation, accessed June 2018 at <https://CRAN.R-project.org/package=gbm>.
- Hastie, T., Tibshirani, R., and Friedman, J.H., 2001, *The Elements of statistical learning—Data mining, inference, and prediction*: New York, Springer-Verlag.
- Hess, G.W., and Stonewall, A.J., 2014, Comparison of historical streamflows to 2013 streamflows in the Williamson, Sprague, and Wood Rivers, Upper Klamath Lake Basin, Oregon—2014-1128: U.S. Geological Survey Open-File Report 2014-1128, [Also available at <https://doi.org/10.3133/ofr20141128>.]
- Humphrey, G.B., Gibbs, M.S., Dandy, G.C., and Maier, H.R., 2016, A hybrid approach to monthly streamflow forecasting—Integrating hydrological model outputs into a Bayesian artificial neural network: *Journal of Hydrology (Amsterdam)*, v. 540, p. 623–640. [Also available at <https://doi.org/10.1016/j.jhydrol.2016.06.026>.]
- Joshi, D., St-Hilaire, A., Daigle, A., and Ouarda, T.B.M.J., 2013, Databased comparison of sparse Bayesian learning and multiple linear regression for statistical downscaling of low flow indices: *Journal of Hydrology (Amsterdam)*, v. 488, p. 136–149. [Also available at <https://doi.org/10.1016/j.jhydrol.2013.02.040>.]
- Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Leetmaa, A., Reynolds, R., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K.C., Ropelewski, C., Wang, J., Jenne, R., and Joseph, D., 1996, The NCEP/NCAR 40-Year Reanalysis Project: *Bulletin of the American Meteorological Society*, v. 77, no. 3, p. 437–472. [Also available at [https://doi.org/10.1175/1520-0477\(1996\)077<0437:TNYRP>2.0.CO;2](https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2).]
- Kennedy, A.M., Garen, D.C., and Koch, R.W., 2009, The association between climate teleconnection indices and Upper Klamath seasonal streamflow—Trans-Niño Index: *Hydrological Processes*, v. 23, no. 7, p. 973–984. [Also available at <https://doi.org/10.1002/hyp.7200>.]
- Lima, A.R., Cannon, A.J., and Hsieh, W.W., 2016, Forecasting daily streamflow using online sequential extreme learning machines: *Journal of Hydrology*, v. 537, p. 431–443. [Also available at <https://doi.org/10.1016/j.jhydrol.2016.03.017>.]

- McCabe, G.J. and Dettinger, M.D., 2002, Primary modes and predictability of year-to-year snowpack variations in the Western United States from teleconnections with Pacific Ocean climate: *Journal of Hydrometeorology*, v. 3, no. 1, p. 13-25. [Also available at [https://doi.org/10.1175/1525-7541\(2002\)003<0013:PMAPOY>2.0.CO;2](https://doi.org/10.1175/1525-7541(2002)003<0013:PMAPOY>2.0.CO;2).]
- Nasseri, M., Zahraie, B., and Forouhar, L., 2017, A comparison between direct and indirect frameworks to evaluate impacts of climate change on streamflows—Case study of Karkheh river basin in Iran: *Journal of Water and Climate Change*, v. 8, no. 4, p. 652–674. [Also available at <https://doi.org/10.2166/wcc.2017.043>.]
- National Oceanic and Atmospheric Administration, 2017, Daily summaries at Crater Lake National Park Headquarters: NOAA National Centers for Environmental Information web interface, accessed August, 2017, at <https://www.ncdc.noaa.gov/cdo-web/datasets/GHCND/stations/GHCND:USC00351946/detail>.
- Nolan, B.T., Fienen, M.N., and Lorenz, D.L., 2015, A statistical learning framework for groundwater nitrate models of the Central Valley, California, USA: *Journal of Hydrology*, v. 531, p. 902–911. [Also available at <https://doi.org/10.1016/j.jhydrol.2015.10.025>.]
- Oregon Water Resources Department, 2018, Water rights in Oregon—An introduction to Oregon’s Water Laws (Aqua Book): Oregon Water Resources Department, 52 p.
- Oregon Water Resources Department, 2018a, Groundwater levels for KLAM0002145: Oregon Water Resources Department web interface, accessed February, 2018, at https://apps.wrd.state.or.us/apps/gw/gw_info/gw_hydrograph/Hydrograph.aspx?gw_logid=KLAM0002145.
- Oregon Water Resources Department, 2018b, Regulatory status: Oregon Water Resources Department, accessed April, 2018, at https://apps.wrd.state.or.us/apps/misc/dsb_area_status/Default.aspx?wm_district=17.
- R Core Team, 2018, R—A language and environment for statistical computing version 3.5.0: R Foundation for Statistical Computing, Vienna, Austria. accessed June 2018 at <http://www.r-project.org/>.
- Rasouli, K., Hsieh, W.W., and Cannon, A.J., 2012, Daily streamflow forecasting by machine learning methods with weather and climate inputs: *Journal of Hydrology*, v. 414–415, p. 284–293. [Also available at <https://doi.org/10.1016/j.jhydrol.2011.10.039>.]
- Redmond, K.T., and Koch, R.W., 1991, Surface climate and streamflow variability in the western United States and their relationship to large-scale circulation indices: *Water Resources Research*, v. 27, no. 9, p. 2381–2399. [Also available at <https://doi.org/10.1029/91WR00690>.]
- Reitz, M., Senay, G., and Sanford, W., 2017, Combining remote sensing and water-balance evapotranspiration estimates for the conterminous United States: *Remote Sensing*, v. 9, no. 12, p. 1181. [Also available at <https://doi.org/10.3390/rs9121181>.]
- Rice, J.S., Emanuel, R.E., and Vose, J.M., 2016, The influence of watershed characteristics on spatial patterns of trends in annual scale streamflow variability in the continental US: *Journal of Hydrology*, v. 540, p. 850–860. [Also available at <https://doi.org/10.1016/j.jhydrol.2016.07.006>.]
- Risley, J.C., Gannett, M.W., Lea, J.K., and Roehl, Jr, E.A., 2005, An analysis of statistical methods for seasonal flow forecasting in the Upper Klamath River basin of Oregon and California: U.S. Geological Survey 2005-5177, 43 p., [Also available at <https://pubs.usgs.gov/sir/2005/5177/>.]
- Rutledge, A.T., 1998, Computer programs for describing the recession of groundwater discharge and for estimating mean groundwater recharge and discharge from streamflow records—Update: U.S. Geological Survey Water Resources Investigations Report 98-4148, 52 p., [Also available at <https://pubs.usgs.gov/wri/wri984148/>.]
- Sachindra, D.A., Huang, F., Barton, A., and Perera, B.J.C., 2013, Least square support vector and multi-linear regression for statistically downscaling general circulation model outputs to catchment streamflows: *International Journal of Climatology*, v. 33, no. 5, p. 1087–1106. [Also available at <https://doi.org/10.1002/joc.3493>.]
- Senay, G.B., Friedrichs, M., Singh, R.K., and Velpuri, N.M., 2016, Evaluating Landsat 8 evapotranspiration for water use mapping in the Colorado River Basin: *Remote Sensing of Environment*, v. 185, p. 171–185. [Also available at <https://doi.org/10.1016/j.rse.2015.12.043>.]

- Snelder, T.H., Lamouroux, N., Leathwick, J.R., Pella, H., Sauquet, E., and Shankar, U., 2009, Predictive mapping of the natural flow regimes of France: *Journal of Hydrology* (Amsterdam), v. 373, no. 1, p. 57–67. [Also available at <https://doi.org/10.1016/j.jhydrol.2009.04.011>.]
- State of Oregon, 2018, Klamath River Basin general stream adjudication: State of Oregon, accessed October 15, 2018, at https://www.oregon.gov/OWRD/programs/WaterRights/Adjudications/KlamathAdj/KBA_ACFOD_00001.PDF.
- Tisseuil, C., Vrac, M., Lek, S., and Wade, A.J., 2010, Statistical downscaling of river flows: *Journal of Hydrology* (Amsterdam), v. 385, no. 1-4, p. 279–291. [Also available at <https://doi.org/10.1016/j.jhydrol.2010.02.030>.]
- Umut, O., and Gul, I., 2015, Statistical downscaling of monthly reservoir inflows for Kemer watershed in Turkey—Use of machine learning methods, multiple GCMs and emission scenarios: *International Journal of Climatology*, v. 35, no. 11, p. 3274–3295. [Also available at <https://doi.org/10.1002/joc.4206>.]
- U.S. Geological Survey, 2017, USGS 1150100 Sprague River Near Chiloquin, OR: U.S. Geological Survey, National Water Information System web interface, accessed June, 2017, at https://waterdata.usgs.gov/or/nwis/uv/?site_no=11501000.
- U.S. Geological Survey, 2018, Geo data portal: U.S. Geological Survey, web, accessed March, 2018, at <https://cida.usgs.gov/gdp/>.
- Xia, Y., and others, 2012, Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2)—1. Intercomparison and application of model products: *Journal of Geophysical Research, Atmospheres*, v. 117, D3, [Also available at <https://doi.org/10.1029/2011jd016048>.]
- Yaseen, Z.M., Jaafar, O., Deo, R.C., Kisi, O., Adamowski, J., Quilty, J., and El-Shafie, A., 2016, Stream-flow forecasting using extreme learning machines—A case study in a semi-arid region in Iraq: *Journal of Hydrology*, v. 542, p. 603–614. [Also available at <https://doi.org/10.1016/j.jhydrol.2016.09.035>.]

For more information concerning the research in this report, contact the
Director, Oregon Water Science Center
U.S. Geological Survey
2130 SW 5th Avenue
Portland, Oregon 97201
<https://www.usgs.gov/centers/or-water>

Publishing support provided by the U.S. Geological Survey
Science Publishing Network, Tacoma Publishing Service Center

