

Research papers

Estimating streamflow permanence with the watershed Erosion Prediction Project Model: Implications for surface water presence modeling and data collection



Konrad C. Hafen^{a,b,*}, Kyle Blasch^c, Paul E. Gessler^d, Jason Dunham^e, Erin Brooks^f

^a U.S. Geological Survey, Idaho Water Science Center, 230 Collins Road, Boise, ID 83702, USA

^b University of Idaho, Water Resources Program, 875 Perimeter Drive MS, 1133, Moscow, ID 83844, USA

^c U.S. Geological Survey, Northwest - Pacific Islands Region, 230 Collins Road, Boise, ID 83702, USA

^d University of Idaho, Department of Forest, Rangeland, and Fire Sciences, 875 Perimeter Drive MS 1133, Moscow, ID 83844, USA

^e U.S. Geological Survey, Forest and Rangeland Ecosystem Science Center, 3200 SW Jefferson Way, Corvallis, OR 97331, USA

^f University of Idaho, Department of Soil and Water Systems, 875 Perimeter Drive MS 2060, Moscow, ID 83844, USA

ARTICLE INFO

Handling Editor: N. Basu

Keywords:

Stream permanence

Perennial

Non-perennial

Hydrologic model

ABSTRACT

Many data collection efforts and modeling studies have focused on providing accurate estimates of streamflow while fewer efforts have sought to identify when and where surface water is present and the duration of surface water presence in stream channels, hereafter referred to as streamflow permanence. While physically-based hydrological models are frequently used to explore how water quantity may be influenced by various climatic and basin characteristics at local, regional, national, and global extents they are less often used to explore streamflow permanence. Herein, the Watershed Erosion Prediction Project (WEPP) hydrological model is applied to watersheds in the humid H. J. Andrews Experimental Forest (HJA) and watersheds of the arid Willow and Whitehorse creeks (WW), both in Oregon, to simulate daily (WW) and annual (HJA and WW) streamflow permanence. One thousand parameter combinations were tested to calibrate WEPP to observed streamflow in the HJA watersheds and one hundred parameter combinations were tested to calibrate WEPP to observed surface water presence time series data in WW watersheds. When calibrated to observed streamflow, WEPP correctly classified annual streamflow permanence for 83 % of HJA stream reaches. In the WW, WEPP simulations correctly classified 63–93 % of daily streamflow permanence observations and 59–87 % of annual streamflow permanence classifications. Inclusion of a dry-day threshold (the maximum number of days a stream reach could be modeled 'dry' but still classified as permanent for the year) improved annual accuracy in three WW watersheds from 2 to 10 %. Parameter sets that produced the best daily accuracies in WW resulted in poor annual accuracies. Results highlight the importance of evaluating physically-based streamflow permanence models on both permanent and nonpermanent streams at daily and annual time scales to ensure evaluation metrics are appropriate for interpretation purposes. Additionally, results suggest that strategic collection of surface water presence observations and streamflow observations may support robust calibration of physically based models to simulate streamflow permanence moving forward.

1. Introduction

The majority of land surface area drains to headwater streams (Downing et al., 2012) so human land use practices and natural disturbances in headwater stream catchments have crucial implications for downstream water quality and quantity (Alexander et al., 2007). Historically, water resources research has focused on measuring and

modeling water quantity in larger rivers because they convey the majority of volumetric flow. More recently, the ecological importance of water resources has received attention (Fritz et al., 2020). Streams must have sufficient water quantity for organisms to survive. Further, whether or not a stream receives regulatory protection under the U.S. Clean Water Act (33 U.S.C. §§1251–1387) is determined by the binary presence or absence of surface water in a stream reach, not by the

* Corresponding author at: U.S. Geological Survey, Idaho Water Science Center, 230 Collins Road, Boise, ID 83702, USA.

E-mail address: khafen@usgs.gov (K.C. Hafen).

<https://doi.org/10.1016/j.jhydrol.2023.129747>

Received 27 December 2022; Received in revised form 17 April 2023; Accepted 25 May 2023

Available online 1 June 2023

0022-1694/Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

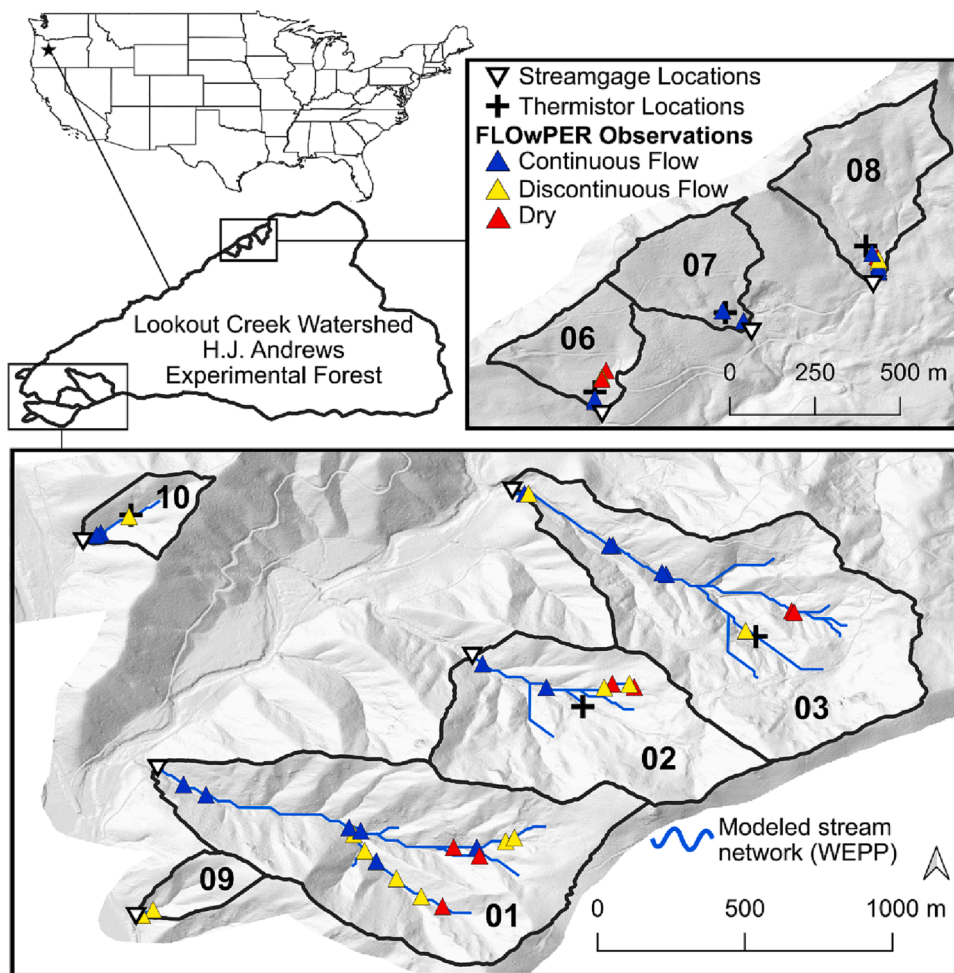


Fig. 1. Watersheds in the H.J. Andrews Experimental Forest where data collection and WEPP modeling occurred. Continuous streamflow data were available at the outlet of all watersheds, with additional information on continuous streamflow permanence conditions available from temperature data loggers (i.e., thermistors; + symbols) deployed in the summer and autumn of 2020, and one-time observations of streamflow permanence (triangles) using a flow permanence feature mapping application (FLOWPER; Jaeger et al., 2020) in summer and autumn of 2020.

relative quantity of water (Walsh and Ward, 2019).

In the United States (U.S.), the majority of data quantifying surface water in streams come from the U.S. Geological Survey (USGS) stream gage network (U.S. Geological Survey, 2016). The primary priorities of this network are to assess water availability and flood risk (DeWeber et al., 2014; Hester et al., 2006). As a result, larger rivers are disproportionately represented, while few data exist on smaller, headwater streams (DeWeber et al., 2014; Krabbenhoft et al., 2022; Zimmer et al., 2020). Experimental watersheds and long-term ecological research sites complement the USGS stream gage network by collecting long-term data over smaller spatial extents that are more representative of headwater systems (Knapp et al., 2012; Tetzlaff et al., 2017; Turner et al., 2003). Like the USGS stream gage network, these experimental stations are generally designed to quantify surface water by streamflow magnitude at gaged locations. The type of data collected at stream gages, and the geographic location of stream gages, have largely influenced model development by determining data and spatial scales available for model validation for applications including flood forecasting, peak and low flow statistics, and estimates of overall streamflow and streamflow timing (DeWeber et al., 2014; Krabbenhoft et al., 2022; Zimmer et al., 2020). Fewer data collection efforts and modeling studies have focused on identifying when and where surface water is present and the duration of surface water presence in stream channels, hereafter referred to as streamflow permanence (Hammond et al., 2020; Jaeger et al., 2019; Jensen et al., 2018; Williamson et al., 2015). The National Hydrography Dataset (NHD) (U.S. Geological Survey, 2019) is the most comprehensive dataset describing streamflow permanence in the U.S. (Nadeau and Rains, 2007). However, NHD streamflow permanence classifications (i.

e., perennial, intermittent, and ephemeral) are not well structured to represent the dynamic nature of streamflow presence and may exhibit error rates up to 50 % on headwater streams regarding the classification of permanent (perennial) versus nonpermanent (intermittent or ephemeral) stream reaches (Fritz et al., 2013; Hafen et al., 2020; Nadeau et al., 2015).

In the last decade, many data collection and modeling efforts have begun to focus on quantifying, estimating, and predicting streamflow permanence at varying spatial extents (Hafen et al., 2022; Jensen et al., 2017; Messenger et al., 2021; Sando and Blasch, 2015, 2015; Ward et al., 2018; Williamson et al., 2015). Regional and mesoscale (e.g., 1–10 km²) streamflow permanence modeling efforts have largely used statistical methods to identify relationships between climatic and physiographic variables that influence the presence of surface water in a particular location during a specific time period (Gendaszek et al., 2020a; Jaeger et al., 2019; Sando and Blasch, 2015). These methods identify variables that influence streamflow permanence, but the models are not readily adaptable to new locations and time periods. By contrast, implementation of physically-based models to estimate streamflow permanence has primarily occurred in a few small (<1 km²) watersheds (Ward et al., 2018; Williamson et al., 2015). Theoretically, by representing the physical processes that govern streamflow and streamflow permanence, physically-based models could be readily applied to new locations and conditions, though identification of optimal model parameters is often also necessary (Ward et al., 2018). Furthermore, many hydrological models are becoming operationalized on cloud-based platforms making them accessible for use by land and water managers. While physically based models have shown promise for representing streamflow

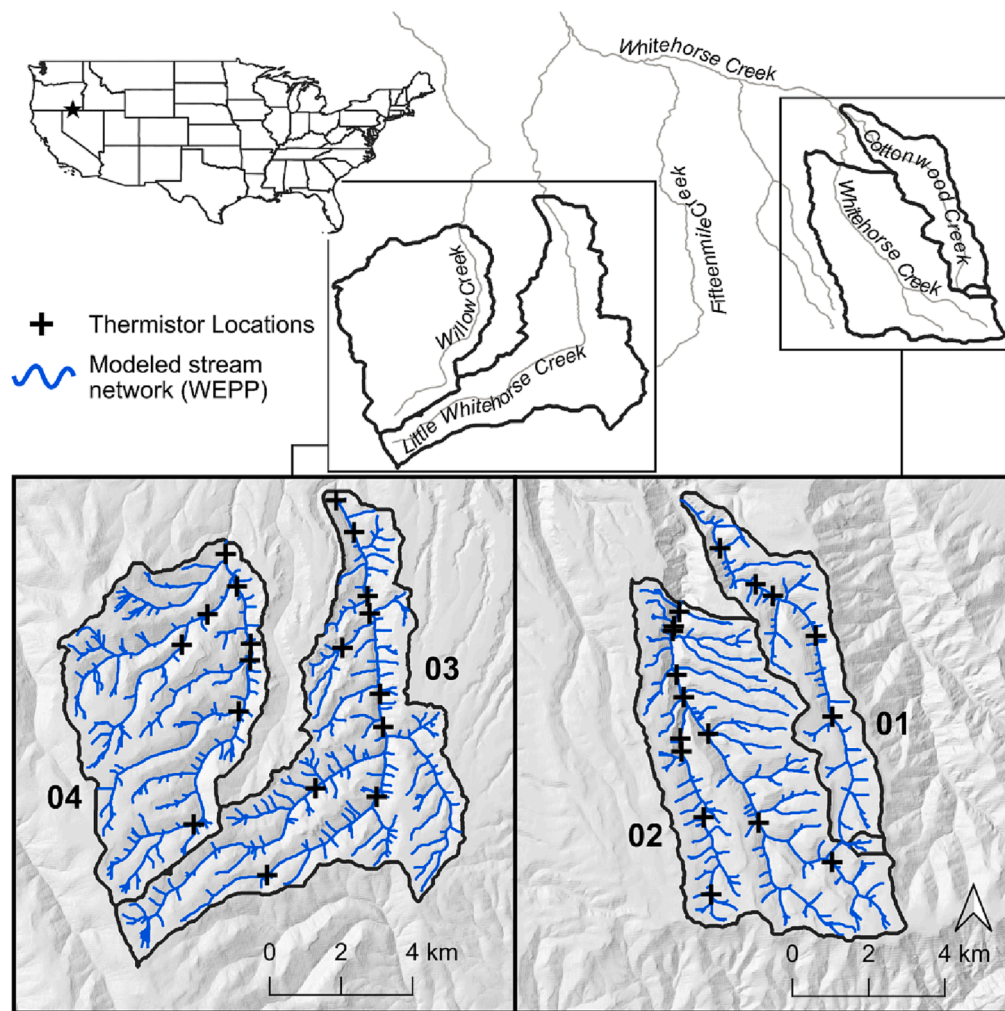


Fig. 2. Watersheds in the Willow-Whitehorse basin where temperature data loggers (i.e., thermistors; + symbols) were deployed and WEPP modeling occurred. In this watershed data on streamflow permanence were inferred from continuous records of stream temperature recorded by thermistors (+symbols) distributed throughout the stream networks (Arismendi et al., 2017; Gendaszek et al., 2020a).

permanence over small spatial extents, they have not been widely tested over larger extents. This is partially due to lack of time-series data that quantify streamflow permanence to validate performance of physically-based models over these larger extents (Jensen et al., 2018; Ward et al., 2018; Williamson et al., 2015).

The purpose of this work is to assess the performance of the cloud-based version (WEPPCloud; Lew et al., 2022) of the physically-based Watershed Erosion Prediction Project (WEPP) hydrological model (Flanagan et al., 2001; Flanagan and Nearing, 1995) to estimate streamflow permanence in both humid and arid environments of the western United States. WEPP was selected because it has shown potential to generate accurate streamflow estimates for high and low flows in small, ungaged watersheds (Brooks et al., 2016; Dobre et al., 2022) and has been implemented to model hydrology in a variety of environments and assess the hydrological effects of land surface disturbances (Dun et al., 2009; Srivastava et al., 2020, 2013; Zheng et al., 2020). Streamflow permanence estimates from WEPP were assessed with a combination of time-series data from sensors and direct observations of streamflow permanence in the arid Willow-Whitehorse watersheds (WW) in the Great Basin of southeast Oregon and in the humid H. J. Andrews Experimental Watershed (HJA) in the Cascade Mountains of western Oregon (Gendaszek et al., 2020a; Jones and Grant, 1996; Schultz et al., 2017; Ward et al., 2020). This study presents an evaluation of a physically based hydrological model (WEPP) to estimate streamflow permanence at the mesoscale.

2. Material and methods

2.1. Study areas

Two sets of watersheds in Oregon, representing both humid and arid climates, were selected for this study. The humid, coastal climate of the western Cascades was represented by eight gaged watersheds in the H. J. Andrews Experimental Forest (HJA) of northwestern Oregon (Fig. 1). The HJA has a rich history of hydrological experimentation, research, and available datasets (Jones and Grant, 1996; Thomas and Megahan, 1998; Ward et al., 2020). Streamflow for all eight watersheds has been continuously gaged since 1995. Drainage areas of the eight gaged watersheds range from 0.90 km² (HJA09) to 1.01 km² (HJA03). Despite their small size, all gaged watersheds support perennial flow at the stream gage sites, though portions of upstream channels and tributaries regularly lose streamflow in summer months (Ward et al., 2020). Elevations of the HJA watersheds range from 400 to 1200 m above sea level. Annual precipitation in the area averages 2300 mm and mean annual temperature is 9.2 deg. C (Ward et al., 2020). Vegetation in the HJA is representative of the Marine Regime Mountains Ecoregion (Bailey, 2016). Comprehensive descriptions of HJA vegetation, climate, morphology, logging experiments, and geology are well described by others (Dyrness, 1969; Frady et al., 2007; Jones et al., 2000; Swanson and Dyrness, 1975; Swanson and James, 1975).

The arid Willow and Whitehorse watersheds (WW) in southeastern

Oregon provide a strong climatological contrast to the more humid HJA watersheds. Four watersheds representing portions of Willow and Whitehorse creeks and their tributaries (WW01, WW02, WW03, WW04) were considered for this study (Fig. 2). Watersheds were selected to be small enough to balance WEPP computation processing requirements but large enough to represent as many data locations as possible (available data are described below). Areas of the modeled watersheds ranged from 15.94 km² (WW01) to 44.33 km² (WW03). Elevation in the WW watersheds ranged from 1,600–2,400 m and mean annual air temperature was 8.1 °C. Average annual precipitation in the WW is approximately 400 mm. Many WW streams are nonpermanent while a few maintain permanent flow each year (Schultz et al., 2017). Surface water presence was monitored at a number of sites in the WW watersheds from 2011 to 2017 as part of studies assessing Lahontan cutthroat trout (*Oncorhynchus clarkii henshawi*) habitat (Gendaszek et al., 2020a; Schultz et al., 2017). Vegetation is representative of the Temperate Desert Ecoregion (Bailey, 2016). Comprehensive details of climate, geology, and landcover of WW are provided elsewhere (Dunham et al., 2003; Gendaszek et al., 2020a; Schultz et al., 2017).

2.2. Streamflow data

Continuous streamflow data for all eight HJA watersheds considered in this study were recorded at 15-minute intervals for multiple decades by the U.S. Forest Service. Minimum, maximum, and mean daily streamflow values were calculated using the 15-minute time series. Based on the recommendation of HJA scientists, streamflow data prior to the year 2000 (when the stage-discharge relationships were most recently updated) were not used (S. Johnson and S. Wondzell, U.S. Forest Service, Pacific Northwest Research Station, personal communications). Measured streamflow data were not available for the WW watersheds.

2.3. Surface water presence data

2.3.1. H. J. Andrews data

During the summer and autumn of 2020 observations of surface water presence were made on stream reaches in gaged HJA watersheds using the USGS Flow Permanence data collection application (FLOWPER; Jaeger et al., 2020). Locations for FLOWPER observations were determined randomly and most locations were observed on two different dates. However, September wildfires prevented access to some HJA watersheds during autumn 2020. In FLOWPER observers designate a point on a stream reach to have 'continuous flow' if surface water is present in the 10 m upstream and downstream of the observer's location, 'discontinuous flow' if water is present but the stream channel also contains channel-spanning dry segments, or 'no flow' if there is no surface water present in the stream channel. Locations of FLOWPER observations are presented in Fig. 1.

In addition to FLOWPER observations, surface water presence was monitored with temperature loggers in HJA gaged watersheds during the summer and autumn of 2020 (Fig. 1). Two temperature loggers were deployed at each site; one in the stream channel, to record water temperature, and one adjacent to the stream channel to record air temperature. The instruments were placed at the deepest point of the stream channel cross-section to give the best indication of surface water presence. Data loggers recorded temperature at 1-hour intervals. Surface water presence was derived from the hourly temperature time series by comparing the magnitude and fluctuation of the in-channel temperature logger and out-of-channel temperature logger (where one exists), or the in-channel thermographs where an out-of-channel sensor was not deployed or malfunctioned (Arismendi et al., 2017; Blasch et al., 2002; Gendaszek et al., 2020a).

FLOWPER and temperature data within the gaged HJA watersheds were collected as part of larger monitoring effort throughout the entire HJA during the summer of 2020. FLOWPER data are available from

Heaston et al. (2022) and temperature data are available from Thorson et al. (2022a).

2.3.2. Willow-Whitehorse data

Surface water presence in the WW watersheds was recorded with thermistors which were deployed between 2011 and 2017. Thermistors were initially deployed to evaluate and model temperature and surface water presence for trout habitat in the watershed (Gendaszek et al., 2020a; Schultz et al., 2017). Surface water presence in WW watersheds was derived from temperature time series following the same methods used for HJA watersheds. Temperature time series and surface water presence data are available from Gendaszek et al. (2020b) for WW and Thorson et al. (2022b) for HJA.

2.3.3. Data processing

To avoid potential misclassifications of surface water presence from thermistor data due to frozen streams, only thermistor observations recorded between April 1 and October 31 were used in this study. Hourly thermistor time-series were converted to daily values. Any day where a thermistor location was determined to be dry for any hour was classified as 'dry', or absent surface water. Data were aggregated at the spatial scale of WEPP stream reaches (described below). All thermistor and FLOWPER data on each stream reach for each date were combined so that if any observation indicated absence of surface water at any location on any day the reach was classified as non-permanent, or 'dry' on that day. Any reach that had at least one 'dry' observation between April 1 and October 31 was classified as non-permanent for that year.

No temperature loggers in HJA recorded surface water absence. However, FLOWPER observations on the same stream reaches as temperature loggers (but at different locations within those reaches) recorded 'discontinuous flow' or 'no flow' conditions on these reaches on the same dates. These results do not indicate inaccuracy in either observation method but, rather, the complexity of streamflow dynamics in these stream systems. Within this same system, Ward et al., (2018) observed surface presence/absence to change multiple times along a stream reach based on channel substrate type and depth. For the most accurate annual classification of streamflow permanence in HJA, we considered all FLOWPER and temperature logger observations. Any stream reach where 'discontinuous flow' or 'no flow' was observed was classified as non-permanent for 2020. Because surface water presence time series from temperature loggers did not record any days without surface water, only annual streamflow permanence was considered in the HJA. Temperature loggers in the WW recorded sites which had both permanent and non-permanent surface water presence throughout a year. Both daily and annual streamflow permanence were considered for the WW. No FLOWPER data were collected in the WW.

2.4. WEPP modeling

WEPP models were generated using the University of Idaho's online implementation of WEPP, named WEPPCloud (Dobre et al., 2022; Lew et al., 2022). WEPPCloud automates acquisition and formatting of the topographic, soil, land cover, and climate data required by WEPP to create the WEPP input and run files. Individual WEPP models were established for each of the eight HJA watersheds and each of the four WW watersheds. For each watershed we used WEPP to generate mean daily streamflow estimates for the 2001–2020 water years (October 1–September 30). WEPP parameters were calibrated to streamflow data for the 2001–2019 water years. Calibrated parameters were then used to simulate streamflow for the 2020 water year.

To set up and run a WEPP simulation using WEPPCloud a user specifies the input digital elevation model (DEM) resolution, both 10 m and 30 m nationally available digital elevation model (DEM) products were available at the time of this study. This study used a 10 m DEM (USGS, 2016) to better represent topography in small, headwater catchments. After the DEM resolution is selected, and a study area is

Table 1

WEPP parameters and their sampled ranges for watersheds in the H. J. Andrews Experimental Forest and Willow-Whitehorse study areas in comparison to the default parameters. Parameter names are abbreviated as follows: baseflow coefficient = KB, deep seepage coefficient = KS, vertical conductivity of the restrictive layer = KR, and crop coefficient = KC.

Study Area	KB days ⁻¹	KS days ⁻¹	KR mm hr ⁻¹	KC -
HJA	0.0–0.1	0.0–0.1	0.0–0.5	0.8–1.2
WW	0.0–0.1	–	–	0.8–1.2
Default	0.04	0.0	144.0	0.95

located, WEPP implements the TOPAZ (Garbrecht and Martz, 2004) software to delineate channels. Only channels with a drainage area of at least 0.03 km² and greater than 70 m in length were modeled. These thresholds produced a stream network that closely matched the observations of Ward et al. (2018) and the High Resolution National Hydrography Dataset stream network (U.S. Geological Survey, 2019). The user then selects a watershed pour-point, after which TOPAZ delineates sub-catchments and hillslopes. Land cover (Dewitz, 2021) and soil data (from the Soil Survey Geographic Database (SSURGO; Soil Survey Staff, n.d.) or State Soil Geographic Database (STATSGO; Schwarz and Alexander, 1995) if SSURGO is not available) were summarized for each hillslope.

WEPPCloud allows users to alter other WEPP-specific parameters. These parameters were left at the default values for the initial WEPP run. After the initial run we downloaded WEPPCloud projects to a local machine for further parameter calibration and analysis with the WEPPY-win-bootstrap tool (Lew et al., 2022). The number of hillslopes that can be modeled with a WEPPCloud run is limited so servers are not overwhelmed with large model requests. Thus, we constrained the size and location of the four WW watersheds to be small enough to run with WEPPCloud while also modeling areas in the WW watersheds with the greatest sensor densities (Fig. 2). Limiting the size of WW watersheds was also necessary to efficiently test multiple parameter combinations for WEPP.

Climate inputs for WEPPCloud were also obtained from gridMET (Abatzoglou, 2013) and summarized by hillslope. The gridMET data (~4 km spatial resolution) were used for this study because gridMET was the only climate dataset available to WEPPCloud that had coverage for 2020, the year streamflow permanence data were collected in the HJA for this study. WEPPCloud default settings and parameters were maintained for HJA and WW simulations with the exceptions described above. For a full description of WEPPCloud capabilities readers are referred to Lew et al. (2022). For mathematical descriptions of WEPP processes readers are referred to Flanagan et al. (2001) and Flanagan and Nearing (1995).

2.4.1. WEPP calibration

2.4.1.1. H. J. Andrews. In the HJA watersheds, each WEPP model was calibrated to observed streamflow over the 2001–2019 water years and calibrated parameters were used to simulate streamflow for the 2020 water year. The purpose of this study was to evaluate the accuracy of streamflow permanence classifications generated by WEPP streamflow estimates (HJA streamflow data for the 2020 water year were not available at the time of analysis). To this end, we used the entire period of record for calibration to determine the accuracy of streamflow permanence estimates when the best streamflow calibration was used. Based on previous studies, and the results of initial WEPP runs, we identified four WEPP parameters to alter for streamflow calibration. The deep seepage coefficient (KS), which controls the amount of subsurface groundwater flow leaving the watershed without passing the stream gage, and baseflow coefficient (KB), which determines the rate of the linear baseflow recession curve, were found to be important in a

previous WEPP implementations in the northwest U.S. (Brooks et al., 2016; Srivastava et al., 2020). Based on observations from preliminary model runs that WEPP was underestimating annual water yield and flood peak magnitude we also adjusted the vertical hydraulic conductivity of the restrictive layer (KR) at the base of the root zone (e.g., bedrock or argillic horizon), and the crop coefficient (KC), a multiplier relating actual evapotranspiration (ET) to reference ET calculated with the Penman-Monteith equation (Monteith, 1965). These parameters have also been adjusted by calibration in previous WEPP applications (Srivastava et al., 2020). Parameter sampling ranges were established from parameter values determined in other studies (e.g., Brooks et al., 2016; Srivastava et al., 2020). Parameter units and ranges are presented in Table 1.

WEPP was run 1,000 times for each HJA watershed. For each run, the set of four parameters was randomly selected from a uniform distribution, bounded as indicated in Table 1. Percent bias (PBIAS), Nash-Sutcliffe Efficiency (NSE; Nash and Sutcliffe, 1970), and NSE for the natural log of daily streamflow estimates (NSE log Q), which gives a better metric for model fit during periods of low flows, were recorded for the results of each parameter set when compared with observed streamflow data.

A single parameter set was identified to represent streamflow for each HJA watershed by evaluating the agreement of each WEPP observation to simulations based on PBIAS, NSE, and NSE log Q. To identify the best parameter set, all model runs with PBIAS < 25 % and NSE > 0.3 were selected. From these runs the parameter set that produced the greatest value of NSE log Q was selected to represent streamflow for a watershed. PBIAS values < 25 % and NSE values > 0.3 were selected because previous modeling studies and reviews suggest these values indicate satisfactory accuracy for hydrological model results (Brooks et al., 2016; Foglia et al., 2009; Moriasi et al., 2007). While the NSE value of 0.3 is slightly lower than satisfactory values reported by Moriasi et al. (2007) we believe a lower threshold is justified because we are primarily focused on model accuracy during low-flow periods. NSE log Q was maximized because identifying non-permanent channels is dependent upon accurate simulated streamflow during low flow periods. In the event no parameter combinations fit values where PBIAS < 25 % and NSE > 0.3 the parameter set with the greatest value of NSE log Q was selected for streamflow modeling.

2.4.1.2. Willow-Whitehorse. WEPP calibration was conducted differently in the WW because there were no streamflow data available for calibration. Instead, WEPP was calibrated to best fit the daily and annual streamflow permanence observations, and only the KB and KC parameters were adjusted. Without streamflow data it is difficult to identify which parameters need to be adjusted to create a good model fit. The shape of the baseflow recession curve (KB) and the amount of ET (KC) are important parameters that could influence streamflow permanence predictions because they control the rate at which baseflow declines and the rate at which water exits the watershed to the atmosphere, respectively. Additionally, the larger WW watersheds required much more time for WEPP to run. By limiting calibration to two parameters the parameter space could be represented effectively with 100 model runs.

2.4.2. WEPP evaluation

WEPP model simulations produced estimates of mean daily streamflow for each modeled stream reach. To match the time period of streamflow permanence observations from FLOWPER and temperature loggers, only WEPP streamflow estimates between April 1 and October 31 were used. Daily streamflow estimates were converted to wet or dry classifications for each day. Any stream reach where the modeled streamflow value was zero was classified as dry. Any stream reach where the modeled daily streamflow was greater than zero was classified as wet. Any stream reach that was modeled as dry for at least one day from April 1 - October 31 was classified as non-permanent for that year.

Table 2

WEPP parameters determined by calibration to observed streamflow in H. J. Andrews watershed. The default WEPP parameters are also shown for reference. KC = crop coefficient, KR = vertical hydraulic conductivity of the restrictive layer, KS = deep seepage coefficient, and KB = baseflow recession coefficient.

WS	KC	KR	KS	KB
HJA01	0.912	0.206	0.0123	0.0072
HJA02	0.934	0.083	0.0067	0.0077
HJA03	1.029	0.180	0.0072	0.0113
HJA06	1.048	0.027	0.0090	0.0046
HJA07	1.091	0.111	0.0129	0.0037
HJA08	0.810	0.049	0.0013	0.0022
HJA09	0.946	0.074	0.0025	0.0011
HJA10	1.006	0.074	0.0089	0.0034

Streams where modeled flow was greater than 0 for all days from April 1 – October 31 were classified as permanent for that year. Classifications were made for both WEPP streamflow estimates using the best parameter set (as described above) and the default WEPPCloud parameters (Table 1).

2.4.2.1. H. J. Andrews. As described above, no dry observations were recorded at temperature logger locations in HJA watersheds. Because the daily time series from thermistors did not contain any dry observations, daily streamflow permanence was not modeled in HJA watersheds. However, at different locations on the same reaches, FLOWPER observations recorded dry conditions. Any reach in the HJA where a dry condition was observed in 2020 was classified as non-permanent. Any reach where a wet observation was made and a dry observation was not made was classified as permanent. WEPP streamflow permanence classifications were evaluated based on their agreement with streamflow permanence classifications made from FLOWPER and temperature logger observations. Agreement was determined by dividing the number of stream reaches that agree in classification by the total number of stream reaches. In total, streamflow permanence classifications were made for 18 stream reaches in the HJA.

2.4.2.2. Willow-Whitehorse. Annual streamflow permanence classifications from WEPP streamflow estimates were determined for WW in the same manner as HJA, as described above. Because surface water presence and absence were both recorded at many WW thermistor sites, daily accuracy of WEPP wet/dry classifications was analyzed for the WW in addition to annual accuracy. Since observed streamflow data were not available for WW watersheds, the best WEPP parameter set was identified by comparison to thermistor data. For results from each WEPP run the accuracy of WEPP results for dry observations, wet observations, and all observations were assessed. This assessment was done for both daily surface water presence observations and annual permanent and non-permanent classifications. In addition to the accuracy values an adjusted accuracy value, which ranged from –1.0 to 1.0, was calculated as

$$\text{AdjustedAccuracy} = \text{OverallAccuracy} - |\text{WetAccuracy} - \text{DryAccuracy}|$$

Where wet accuracy is the accuracy on days thermistors recorded surface water presence (or the accuracy on observed permanent streams when considered annually), dry accuracy is the accuracy on days thermistors recorded surface water absence (or the accuracy on observed non-permanent streams when considered annually), and overall accuracy is the total number of days, or years, the WEPP classification matched the observed classification. Adjusted accuracy was used to give equal weight to surface water presence and absence observations because overall accuracy is biased towards the category with more observations. For example, a site may be dry for 20 days out of a 200-day period. If the model predicts all 200 days to be wet the overall accuracy would be 90 % (or 0.9) but the model would not have correctly classified any dry observations. Adjusted accuracy penalizes model

results for over-predicting one category in relation to the other. The accuracy of WEPP daily and annual classifications was evaluated for the parameter sets with the best adjusted accuracy for daily and annual classifications.

A threshold for the number of dry days (dry-day threshold) allowed before a stream reach was classified as non-permanent was also tested for WW watersheds. For example, with a dry-day threshold of zero, WEPP would need to simulate water in a stream channel every day from April 1 to October 31 for a stream reach to be classified as permanent. With a dry-day threshold of three, a stream reach with three or fewer simulated dry days would still be classified as permanent. Other studies (e.g., Ward et al., 2018; Williamson et al., 2015) have implemented flow thresholds, where modeled flows below the threshold are classified as dry even when the model estimates water in the stream channel and serves to adjust models that overpredict the number of wet days. The dry-day threshold serves the opposite purpose, to adjust WEPP in the event it overpredicts too many streams to be annually non-permanent. The dry-day threshold only impacts the annual streamflow permanence classification. Dry-day thresholds of 0–20 days were tested against all parameter sets and the combination of the dry-day threshold value and parameter set that produced the highest adjusted accuracy was selected for further analysis. The dry-day threshold is useful to calibrate model outputs for annual streamflow permanence estimates when a flow threshold cannot be implemented.

3. Results

3.1. WEPP calibration

With calibrated parameters (Table 2) WEPP satisfactorily modeled streamflow (per our established accuracy benchmarks) in most HJA watersheds (Fig. 3). No parameter set for HJA08 and HJA09 met both the PBIAS and NSE constraints of 25 % and 0.3, respectively. WEPP underestimated annual water yield by at least 49 % in HJA08 and 42 % in HJA09. It is also apparent that WEPP underpredicted or missed flood peaks in several of the watersheds (likely because of the spatially coarse precipitation input data). However, the NSE log Q values indicate that WEPP satisfactorily modeled most low-flow periods, which are most important for identifying when (or if) surface flow may cease. The receding limbs of flood peaks also match relatively well between observed and modeled streamflow time series. Additionally, NSE values were greater than 0.3 for all watersheds except HJA09 over the 2001–2019 water years, indicating sufficient streamflow simulation (Foglia et al., 2009).

3.2. WEPP streamflow permanence classification accuracy

3.2.1. H. J. Andrews classification accuracy

When using the default WEPPCloud parameters (Table 1), streamflow permanence classifications from WEPP streamflow estimates were 39 % accurate (Fig. A2). In the upper reaches of the larger HJA watersheds (HJA01, HJA02, HJA03) and in three of the smaller watersheds (HJA06, HJA08, HJA10), WEPP predicted permanent conditions on non-permanent streams. Additionally, WEPP predicted the mainstem reaches of HJA01 and HJA02 (two segments) to be non-permanent when they were observed to be permanent. However, based on observations from other studies (Ward et al., 2018) the main stem reaches of HJA01 and HJA02 go dry in some places nearly every year (Sherri Johnson and Steve Wondzell, U.S. Forest Service, Pacific Northwest Research Station, personal communication). Assuming those stream segments also had dry patches in 2020 the accuracy of the WEPPCloud default parameters would be 56 %.

The WEPP model calibrated to observed streamflow performed considerably better for annual streamflow permanence classification than with the default WEPPCloud parameters, resulting in 61 % accuracy with observed streamflow permanence classifications (Fig. 4). Most

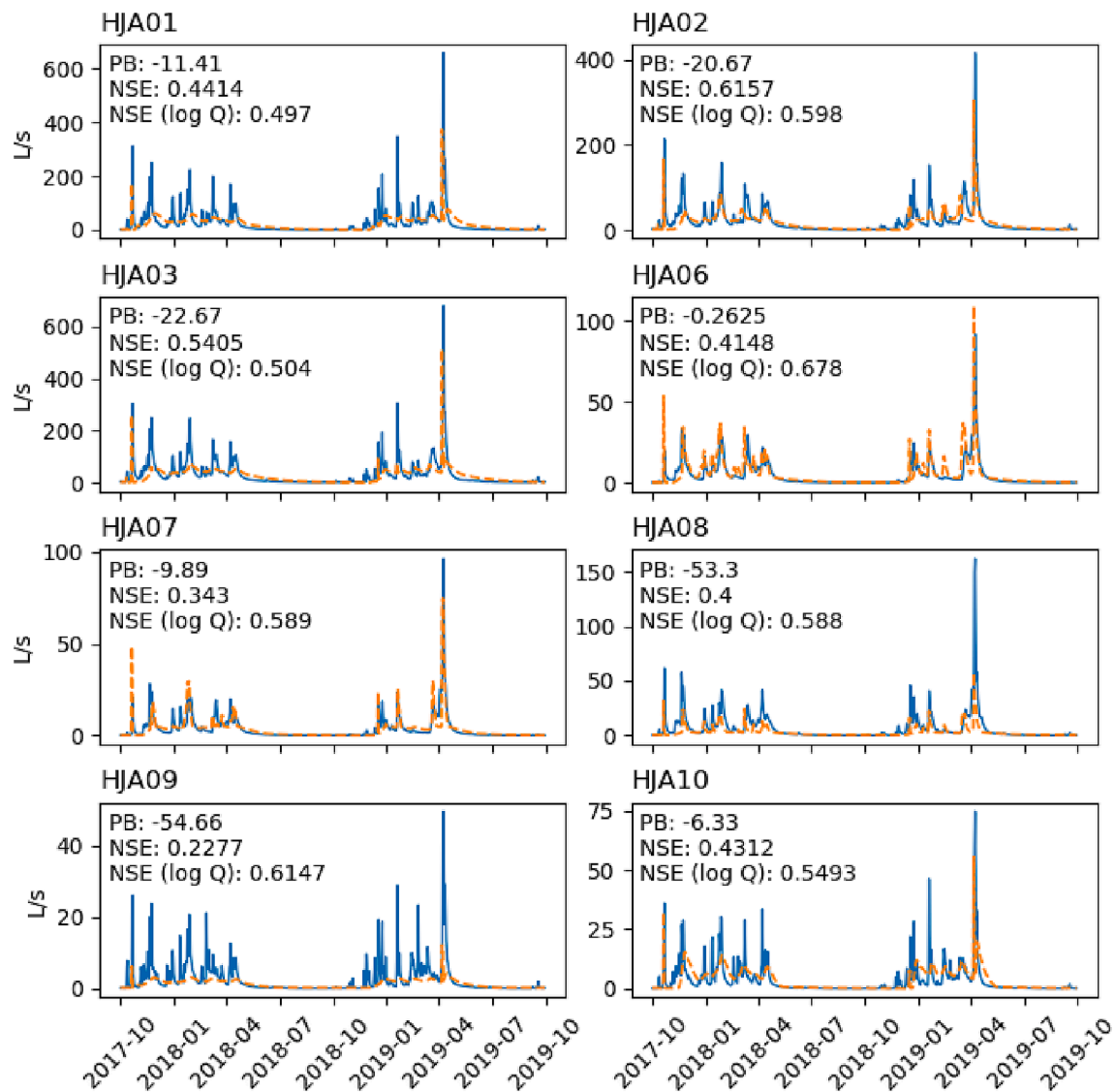


Fig. 3. Comparison of observed (blue) and modeled (dashed orange) streamflow for H. J. Andrews watersheds. Percent bias (PB), Nash Sutcliffe Efficiency (NSE), and NSE calculated on the natural log of streamflow values (log Q) represent model fits over the 2001 to 2019 water years. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

errors occurred in the smaller watersheds (HJA06, HJA07, and HJA08). Once again, the WEPP estimates classified the main stem reaches of HJA01 and HJA02 as non-perennial when no dry observations were made on those reaches in 2020. As noted above HJA scientists observe dry portions of these stream segments nearly every year. FLOWPER observations of these stream reaches did not include the entire reach, but just the conditions in the stream channel within 10 m of a point. Two FLOWPER observations were made on the mainstem reach of HJA01 and one observation on each of the two mainstem segments of HJA02 (Fig. 1). It is possible that the FLOWPER observations were made prior to a portion of the stream reach drying, or on a portion of the stream reach that did not dry, while portions upstream and/or downstream of the observation location were dry. Assuming these three stream segments were non-perennial during 2020, as indicated by previous observations of HJA scientists, the accuracy of the calibrated streamflow permanence estimates would be 83 %.

3.2.2. Willow-Whitehorse classification accuracy

Maximum accuracy of daily WEPP estimates differed by watershed and ranged from 70 % in WW02 to nearly 100 % in WW04 (Fig. 5). However, the maximum accuracies were inflated by a greater number of

wet observations (Fig. 5). Daily overall accuracies corresponding to the maximum adjusted accuracy were lower, ranging from 62 % in WW03 to 93 % in WW04. As indicated by Table 3, the adjusted accuracy selected parameter sets that maximized the accuracy for both wet and dry classifications. However, accuracy of annual (permanent and non-permanent) classifications was poor for parameter sets with the best daily adjusted accuracy with <50 % of permanent stream reaches being classified correctly at the annual time scale (Table 3).

Maximum accuracy of annual (non-permanent and permanent) WEPP estimates ranged from 60 % in WW03 to 95 % in WW04. As with daily accuracies, the maximum annual accuracies were inflated due to a greater number of observed permanent years than non-permanent years (Fig. 6). Overall annual accuracies corresponding to the maximum adjusted accuracy were lower, ranging from 56 % in WW03 to 85 % in WW01 (Table 4). WEPP parameters corresponding to the best daily accuracies (Table 3) and best annual accuracies (Table 4) were different for all WW watersheds.

Inclusion of the dry-day threshold was supported for three of the four WW watersheds. WW02 achieved its greatest annual adjusted accuracy when permanent streams were represented as streams with eight or fewer dry days, WW03 with three or fewer dry days, and WW04 with

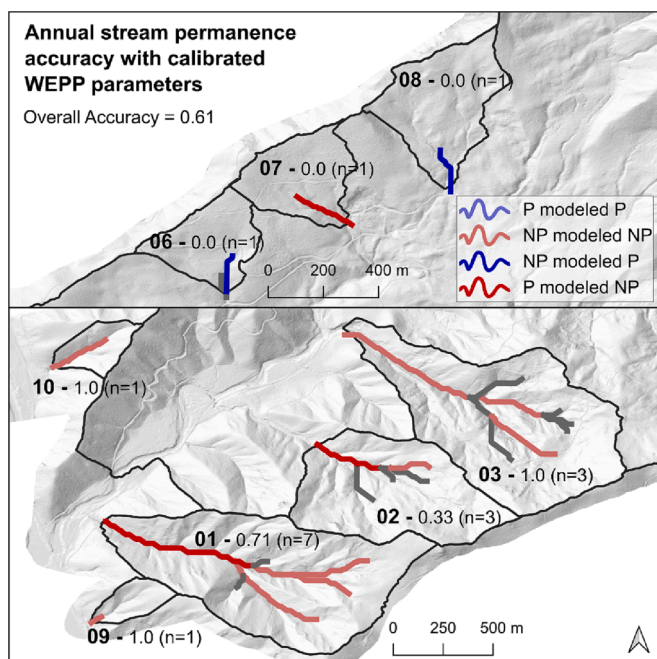


Fig. 4. Accuracy of modeled (WEPP) permanent (P) and nonpermanent (NP) streams for each H. J. Andrews watershed with WEPP parameters calibrated to observed streamflow. The number of modeled stream reaches in each watershed is described by n . Grey lines indicate stream reaches where no observational data were available for validation.

two or fewer dry days (Fig. 7). WW01 achieved greatest annual adjusted accuracy with no dry-day threshold. Inclusion of the dry-day threshold improved accuracy only marginally for WW02 and WW03 (2 % and 3 %, respectively). However, the dry-day threshold improved accuracy for WW04 by 10 % (Table 4). This finding is similar to other studies (Ward et al., 2018; Williamson et al., 2015) that determined a daily streamflow threshold was necessary to eliminate incorrectly classified dry observations and indicates that WEPP (and potentially other models) may require an opposing threshold to adjust for permanence classifications at an annual time step.

The WEPP parameterizations that produced the best annual accuracies for WW04 stayed consistent with addition of the dry-day threshold. However, WW02 and WW03 parameters changed with inclusion of the dry day threshold (parameters were selected from the initial set of calibration runs). KC and KB both increased slightly for WW02 while KC had a slight decrease and KB a slight increase for WW03 (Table 4).

Analysis of daily and annual accuracies presented above indicates that high daily accuracies do not always result in correct permanent and non-permanent classifications for a year. Fig. 8 (and Fig. A3, A4, and A5) show the annual and daily accuracies at each observed reach in the WW watersheds for each observed year. These results show that time periods and locations with high daily accuracy may still be classified incorrectly annually (e.g., Fig. 8) while time periods and locations with low accuracy may still be classified correctly (e.g., Fig. A3). It is important to note that the level of daily accuracy for a correct non-permanent classification varies. For example, a non-permanent stream reach could be dry for 10 or 100 days during a year. The model only needs to simulate enough dry days to be greater than the dry-day threshold (if one is used) to correctly simulate the annual condition. On the other hand, permanent stream reaches require the model to correctly simulate all wet days (if no dry-day threshold is used) to produce the correct annual classification.

4. Discussion

Better understanding of the patterns of streamflow permanence may be useful for aquatic habitat evaluations and regulatory determinations. To this end a host of statistical and process-based models (including physically based models) have been developed and applied (Gendaszek et al., 2020a; Jaeger et al., 2019; Jensen et al., 2018; Sando and Blasch, 2015; Ward et al., 2020, 2018; Williamson et al., 2015; Yu et al., 2018). This study presents an example of streamflow permanence modeling using a physically based model (the WEPP model) in both humid and arid study areas. We implemented a unique approach by using surface water presence observations to calibrate WEPP in the WW watersheds where streamflow data were not available. Results indicate the importance of evaluating model simulations based on both daily and annual accuracy, assessing model performance on both permanent and non-permanent streams, and the usefulness for targeted data collection to accurately describe the permanence condition of entire stream reaches, not just discrete points.

NHD streamflow permanence classifications are currently the most comprehensive source of streamflow permanence data for the United States. Overall accuracy of the NHD classifications is approximately 80 % but is much lower (50–60 %) for headwater streams (Fritz et al., 2013; Hafen et al., 2020; Nadeau et al., 2015). Accuracy of WEPP streamflow permanence estimates ranged from 59 to 87 % for our modeling application. WEPP streamflow permanence estimates are not directly comparable to the NHD classifications because the networks defined by WEPP and NHD differ. More importantly, the NHD streamflow permanence designations are generally static through time, which makes it difficult to conduct an objective comparison with the WEPP results which are dynamic through time and attempt to model stream reaches that may alternate between permanent and non-permanent each year. In the HJA, the only misclassifications of annual streamflow permanence made with the WEPP approach were on headwater (first order) streams. Overall, WEPP correctly classified annual streamflow permanence for 6 of 9 (67 %) HJA first-order streams and the only misclassifications occurred on headwater stream reaches in the WEPP network (i.e., there were no WEPP misclassifications on stream reaches represented by NHD). Because WW results covered multiple years and the WW network included more headwater reaches than the NHD it is more difficult to make accuracy comparisons. However, the WW accuracy range of 59–87 % is similar to the NHD accuracy reported by other studies (Fritz et al., 2013; Hafen et al., 2020; Nadeau et al., 2015). Though WEPP accuracy was variable between watersheds and through time, the annually dynamic WEPP streamflow permanence classifications can provide more insight about how specific climatic conditions may impact streamflow permanence for a given study area. Additionally, this physically-based approach could be used to test for nonstationary trends in streamflow permanence through time (Milly et al., 2008).

Previous studies that examined the utility of physically-based models to simulate streamflow permanence have focused intensive data collection efforts on a small number of nonpermanent stream reaches (Jaeger et al., 2014; Ward et al., 2018; Williamson et al., 2015). Data for this study represented 40 unique stream reaches (18 in HJA and 32 in WW) but did not describe reaches in the same detail as previous studies. One advantage of our approach is that both permanent and non-permanent streams are represented. Without data for both permanent and nonpermanent streams model utility is uncertain because a single daily miscalculation on a permanent stream reach can result in an annual classification error. As indicated by our results, a model parameterization designed to maximize daily accuracy may not maximize annual accuracy and annual permanence classifications may be incorrect even when high accuracy against daily values is achieved. When comparing the calibrated parameters that produced the best daily accuracies (Table 3) and the best annual accuracies (Table 4) for streamflow permanence the annual calibrations decreased ET (i.e., decreased the KC parameter resulting in increased streamflow) and

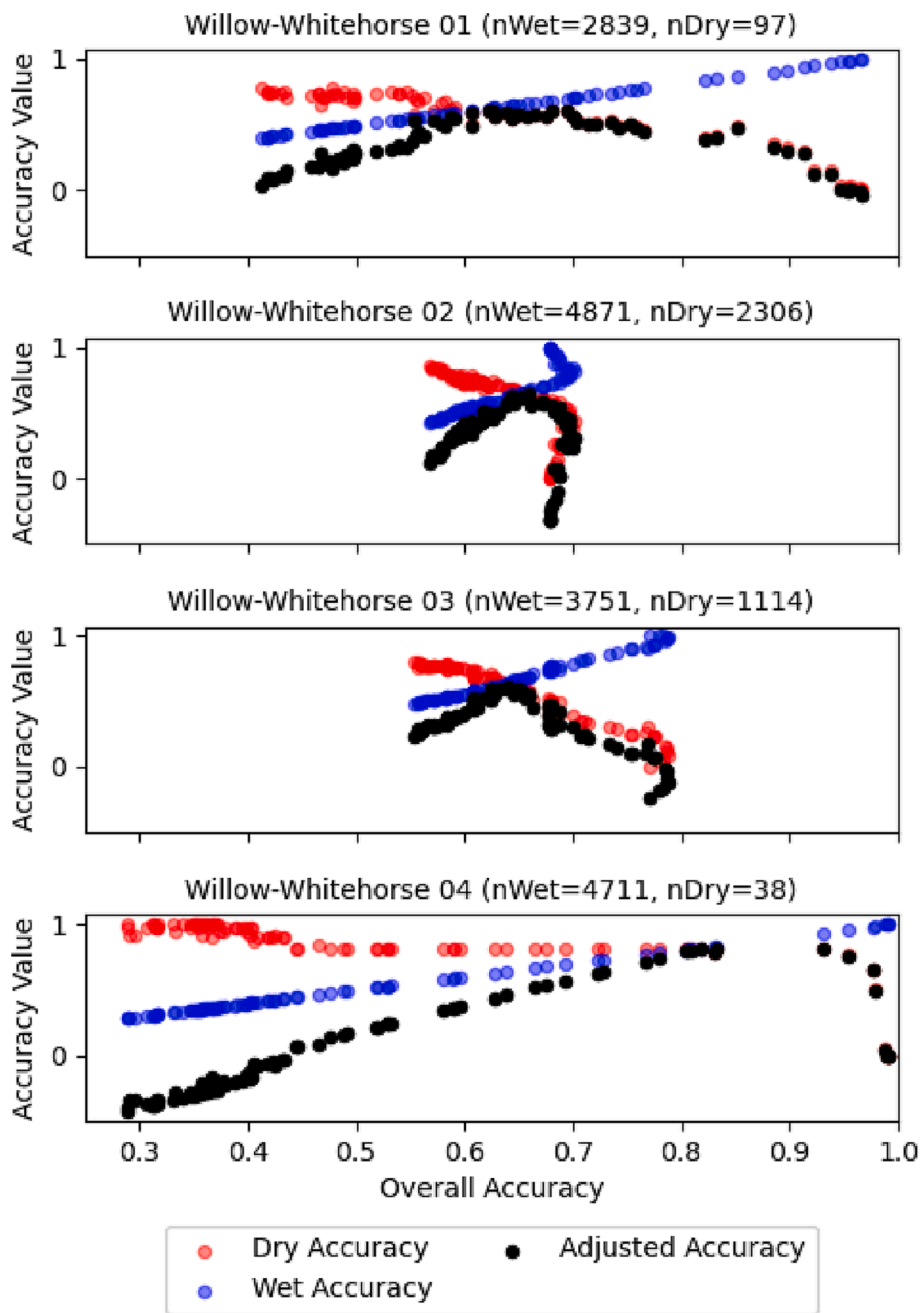


Fig. 5. Daily accuracy of modeled (WEPP) surface water presence with different parameter sets when compared with observed surface water presence at each Willow-Whitehorse watershed where nWet and nDry are the total number of wet and dry days, respectively, observed at all thermistor sites for a watershed.

Table 3

Daily and annual accuracy of the WEPP parameter sets with the highest adjusted accuracy values for each watershed. WAcc and DAcc describe the modeled accuracy when compared to wet and dry observations, respectively. PAcc and NPAcc describe annual modeled accuracy when compared to permanent and nonpermanent locations, respectively.

WS	Daily Accuracy			Annual Accuracy			WEPP params	
	Acc	WAcc	DAcc	Acc	PAcc	NPAcc	KC	KB
WW01	0.85	0.86	0.48	0.20	0.00	1.00	1.139	0.030
WW02	0.66	0.65	0.66	0.67	0.07	1.00	1.076	0.052
WW03	0.62	0.59	0.71	0.47	0.00	1.00	1.153	0.064
WW04	0.93	0.93	0.81	0.60	0.57	1.00	1.105	0.014

decreased KB (i.e., lengthened baseflow recession resulting in more elevated streamflow following runoff). Assessment of annual permanence classification accuracy is also important to temper daily accuracy

assessment which can be biased from unequal occurrence of wet and dry days. The model parameters that produced the best daily and annual accuracies in the WW watershed were different. Thus, the timestep at

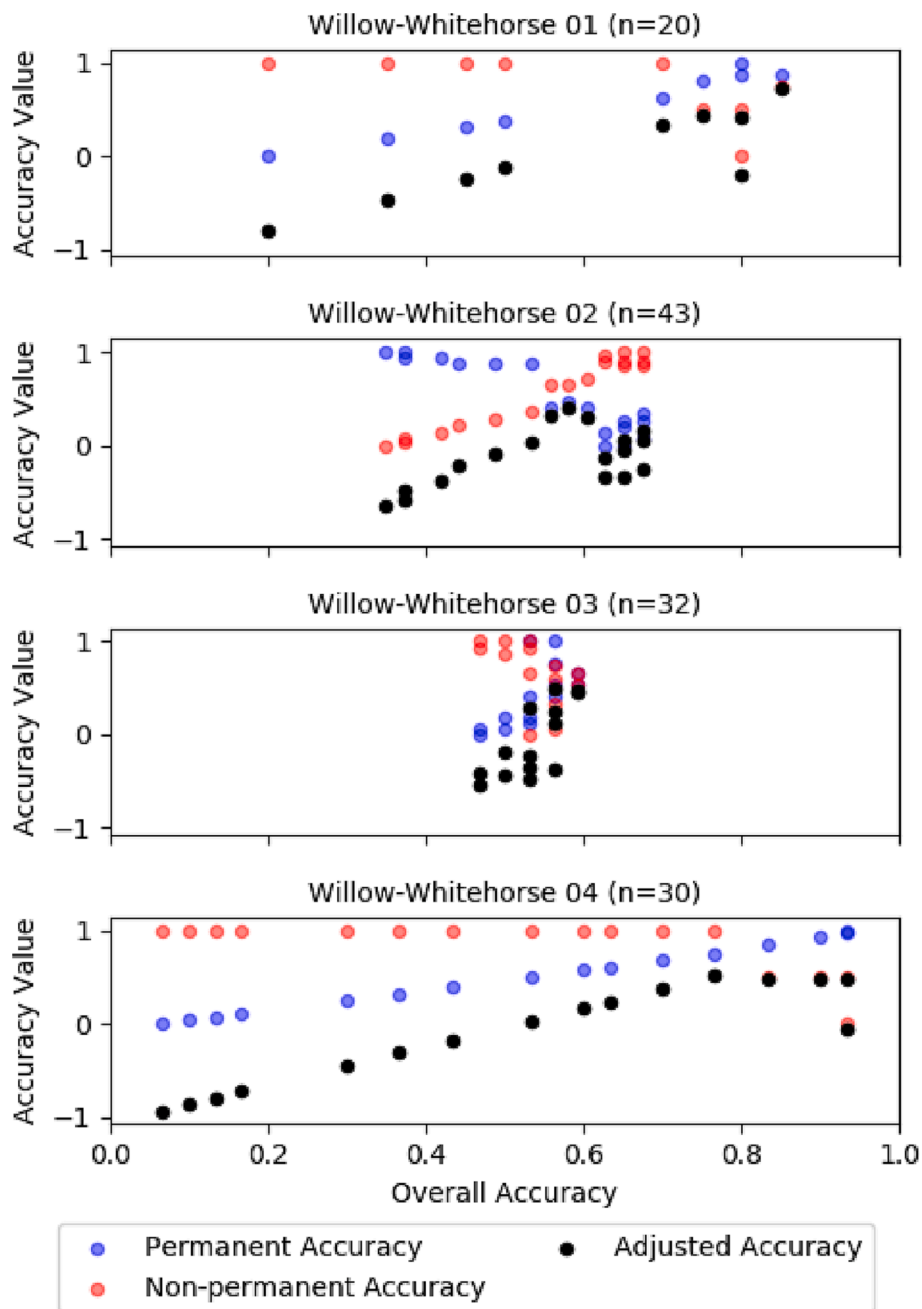


Fig. 6. Annual analysis of WEPP accuracy for all tested parameter sets in Willow-Whitehorse watersheds where n is the summed number of years with observed data across all observation locations in each watershed.

Table 4

Annual WEPP accuracy (Acc) in each Willow-Whitehorse watershed (WS) with and without a threshold for the minimum number of dry days (Days) required for a stream to be classified as non-permanent and the corresponding values for crop coefficient (KC) and baseflow recession coefficient (KB).

WS	Without Dry-day Threshold			With Dry-day Threshold			
	Acc	KC	KB	Acc	Days	KC	KB
WW01	0.85	1.018	0.026	0.85	0	1.018	0.026
WW02	0.58	0.869	0.029	0.60	8	0.913	0.030
WW03	0.56	1.087	0.025	0.59	3	0.906	0.029
WW04	0.77	0.828	0.018	0.87	2	0.828	0.018

which streamflow permanence models are validated should be a point of consideration for future studies because accuracy of model timesteps does not necessarily represent accuracy at meaningful application timesteps. For example, correct annual classifications can be modeled even with lower daily accuracies and high daily accuracies can result in annual misclassifications.

Other studies that use streamflow estimates from process-based models to determine streamflow permanence at sub-annual timesteps often implement streamflow thresholds below which a stream is classified as non-permanent (Hafen et al., 2022; Ward et al., 2018; Williamson et al., 2015). For example, with a flow threshold of 10 L/s stream reaches with a modeled streamflow <10 L/s would be classified as non-permanent, or dry, even when the model predicted positive flow below the threshold. Our results show support for an opposing dry-day threshold when using daily streamflow permanence estimates to infer

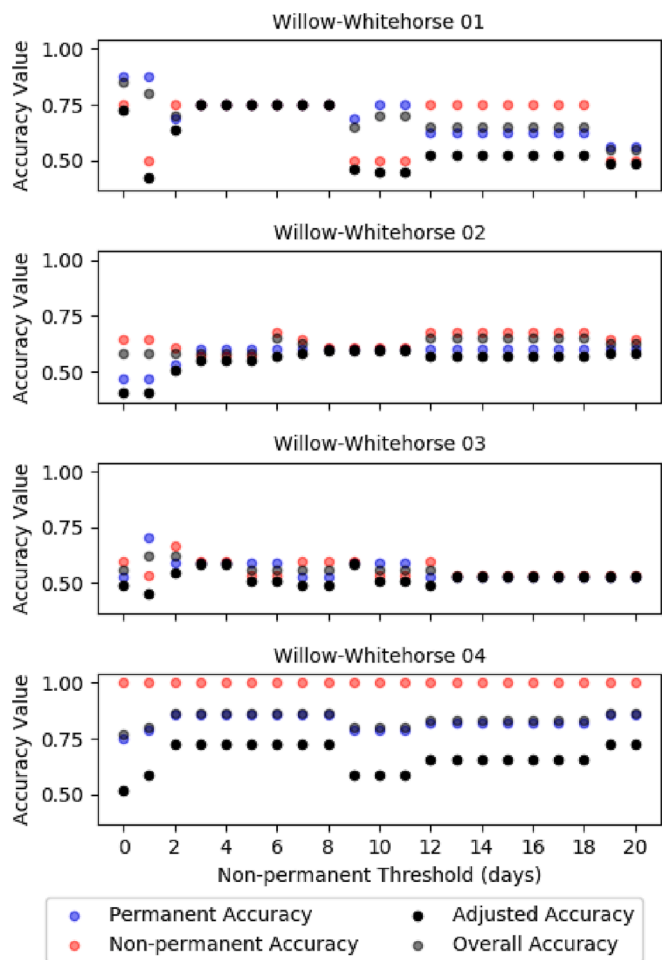


Fig. 7. Annual accuracy values of the best WEPP parameterization for each dry-day threshold value in each Willow-Whitehorse watershed.

annual streamflow permanence. With a dry-day threshold, the annual classification of stream reach is non-permanent only when the number of modeled dry days exceeds the threshold. This threshold was useful for the WW watersheds where streamflow data were not available for model

calibration.

In the HJA watersheds WEPP performance increased by 34 % after streamflow estimates were calibrated to observed streamflow than with the default WEPPCloud parameterization. However, very few areas have the stream gage density of HJA, potentially making calibration to observed streamflow a challenging approach. Calibrated parameters were relatively similar between the gaged HJA watersheds but did display some variation and modeled streamflow permanence accuracy in HJA with the default WEPP parameters was 56 %. The low streamflow permanence classification accuracy for the uncalibrated model indicates that calibration is necessary to achieve suitable results, but the similarity of calibrated variables between the HJA watersheds suggest that calibration may be conducted using observed data from larger gaged basins. One limitation of this study is a direct link between daily streamflow estimates calibrated to streamflow and daily observed surface water presence. In the HJA watersheds, where daily streamflow data were available for WEPP calibration, daily surface water presence records from temperature sensors did not accurately describe surface water presence along a full stream reach. Because WEPP provides streamflow estimates at the reach scale it is difficult to identify relationships between WEPP estimates and surface water presence. Strategically placing sensors at stream-reach locations that are most likely to dry first will provide better a better indicator of cumulative surface water presence along an entire reach and provide more value for model evaluation.

It is important to note potential limitations imposed by the input climate data. The gridMET data had a horizontal spatial resolution of 4 km (1 pixel represents 16 km²), which is much coarser than the size of the modeled HJA watersheds. We also developed models with daily data. The coarse spatial and temporal resolution of the input climate drivers likely influence the ability of WEPP to accurately capture flood peaks. Brooks et al. (2016) observed that even with hourly climate inputs WEPP underestimated flood peaks. Therefore, the availability of detailed climate data with which to drive WEPP may be a limiting factor to accurately implementing this approach in ungaged areas. Data from a single weather station were available for the HJA study area, however, we chose to use gridMET because the data are available nationwide and to provide a more robust comparison to the arid WW watersheds where weather station data were not available. Previous studies have also identified microclimate variation in the HJA that are not represented by the weather station (Daly et al., 2010; 2007).

Misclassification of four non-permanent stream reaches, two in HJA01 and two in HJA02, point to the uncertainty associated with using

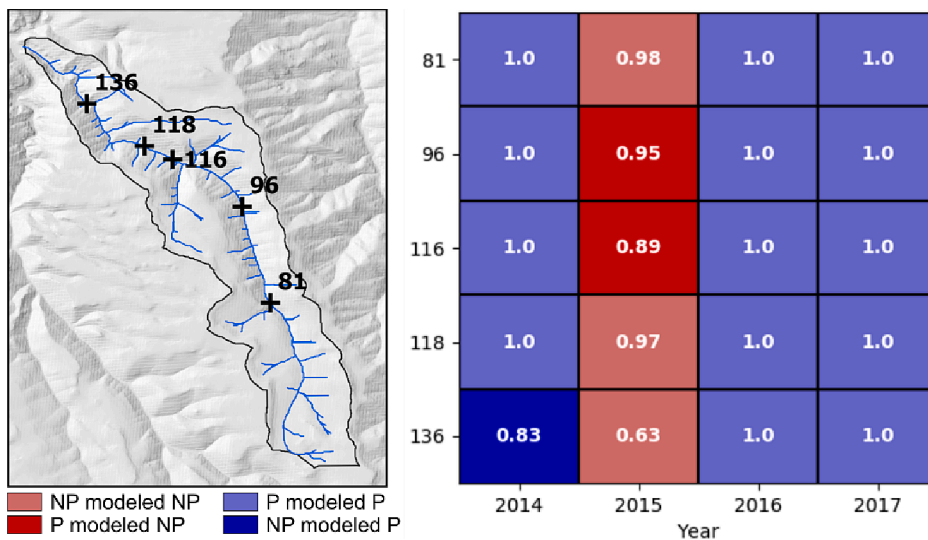


Fig. 8. Annual accuracy of permanent (P) and non-permanent (NP) classifications from the parameter set with the best annual accuracy for WW01. Numbers inside the grid show the corresponding daily accuracy (proportion) of wet and dry observations. Numbers on the Y-axis correspond to the reach identifiers on which the thermistors are located (map).

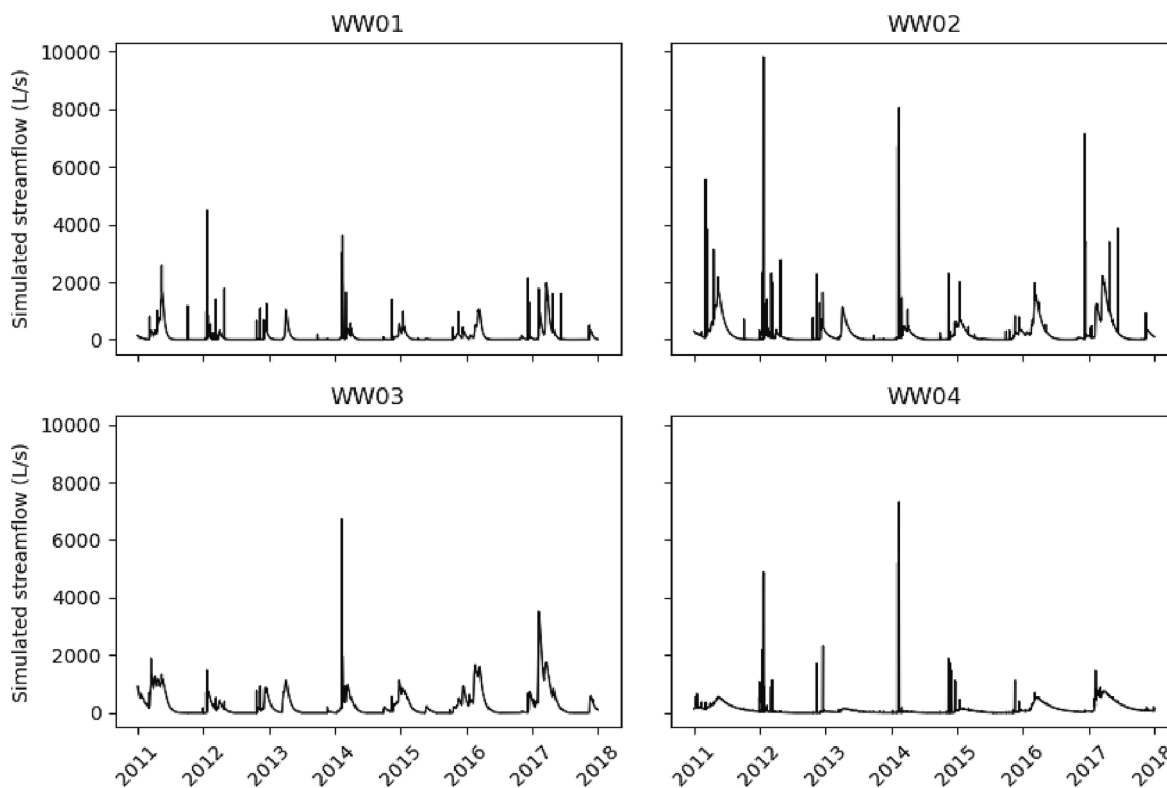


Fig. A1. Simulated streamflow (L/s) from the WEPP model using parameters calibrated to daily water presence observations (Table 3) at the outlet of each of the Willow-Whitehorse watersheds.

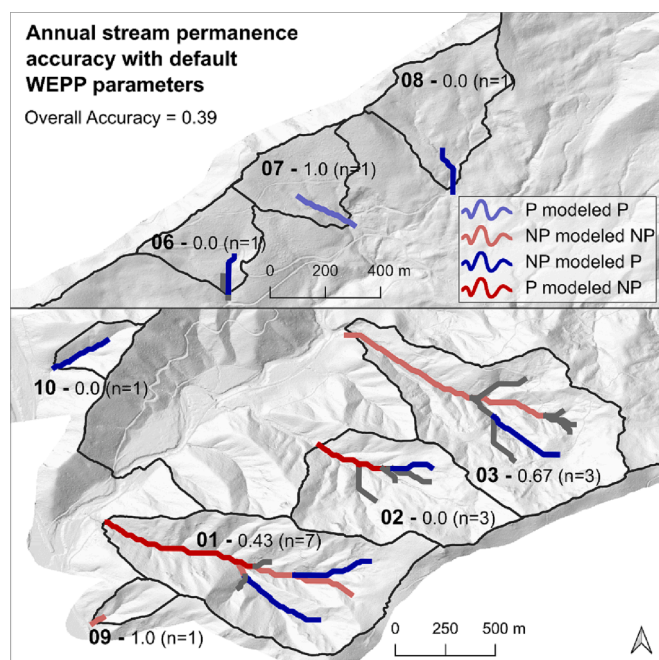


Fig. A2. Accuracy of modeled (WEPP) permanent (P) and nonpermanent (NP) streams for each H. J. Andrews watershed with default WEPP parameters. The number of modeled stream reaches in each watershed is described by *n*. Grey lines indicate stream reaches where no observational data were available for validation.

data from point observations to make reach-scale classifications. These examples indicate that data collection for classifying streamflow permanence will be most effective when focused in areas where stream

reaches are most likely to go dry (though these locations can be difficult to determine without observational data). Dry observations indicate, with certainty, a stream reach was not permanent (for a given time period) while wet observations only serve to support the hypothesis that the stream is permanent but cannot confirm this hypothesis unless spatially and temporally continuous observations are made. This asymmetry is analogous to species detections; when a species is not observed it simply indicates the species was not detected (Kéry and Schmid, 2004; McCarthy et al., 2013), it does not indicate the species was not present. Detecting a non-permanent stream depends on 1) timing of the observation relative to the time when drying may be expected and 2) the duration and possible frequency of drying – the number of days that the stream is dry and, thus, the number of chances available to make a dry observation. Observation error, when a dry stream is classified as wet (or vice versa) by an observer or instrument, is another factor to consider. In the case of streamflow permanence observation, observation error is likely most prevalent for instrumented data. For example, changes in channel morphology can result in temperature loggers becoming covered with sediment (so the observed temperature does not represent air or water temperature), no longer occurring in the deepest part of a channel cross-section (resulting in dry observation when there is water in a channel), or scour around the sensor can result in a pool that persists even when upstream and/or downstream portions of the stream channel are dry. Examples from this study accentuate the importance to account for detectability in future streamflow permanence modeling applications.

Overall, the calibration produced good results in WW01 and WW04 and moderate results in WW02 and WW03 for both daily and annual accuracy. Calibration to surface water presence observations could be useful for future development of physically based streamflow models because collection of surface water presence/absence data is less time consuming and costly than continuous records of streamflow data. The drawback to this method, as previously mentioned, is that surface water sensors record the condition at a point and may not necessarily be

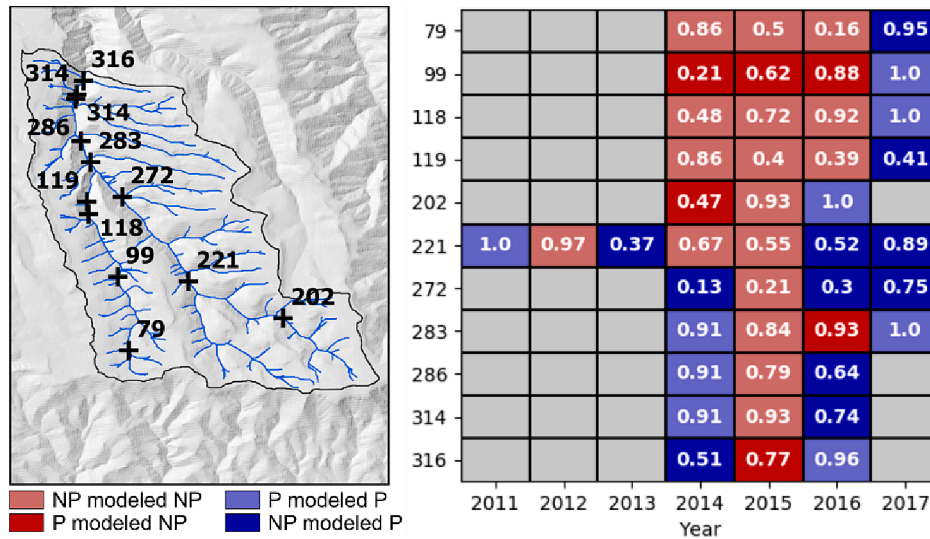


Fig. A3. Annual accuracy of permanent (P) and non-permanent (NP) classifications from the parameter set with the best annual accuracy for WW02. Numbers inside the grid show the corresponding daily accuracy (proportion) of wet and dry observations. Numbers on the Y-axis correspond the reach identifiers on which the thermistors are located (map).

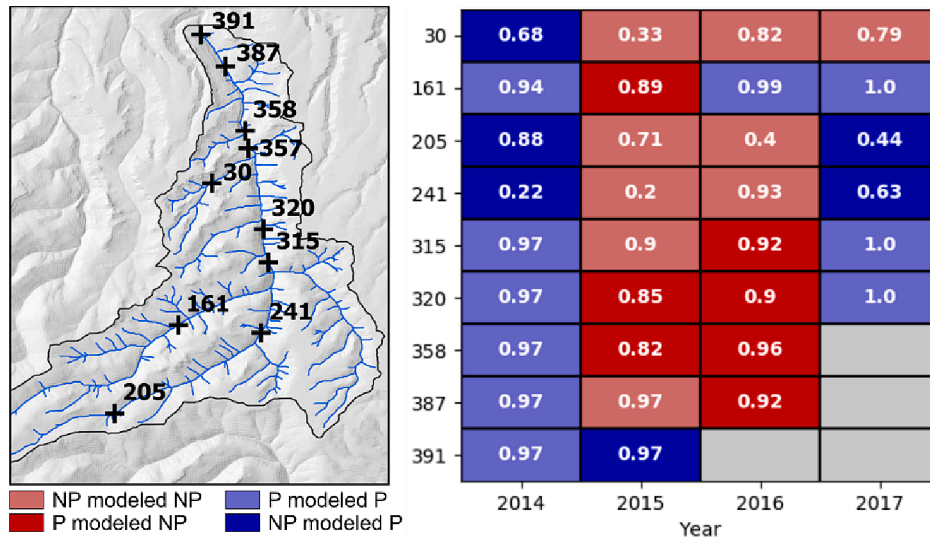


Fig. A4. Annual accuracy of permanent (P) and non-permanent (NP) classifications from the parameter set with the best annual accuracy for WW03. Numbers inside the grid show the corresponding daily accuracy (proportion) of wet and dry observations. Numbers on the Y-axis correspond the reach identifiers on which the thermistors are located (map).

indicative of a stream reach (Kampf et al., 2021), the spatial scale that is important for regulatory determinations (Walsh and Ward, 2019). Focusing data collection at locations that are more representative of a stream reach (as mentioned above), along with reach scale classification from on-the-ground surveys and remotely sensed products may help narrow this knowledge gap.

This study showed good agreement between WEPP streamflow estimates and surface water presence in the humid HJA watersheds and arid WW watersheds. Overall accuracy of the WEPP streamflow permanence estimates was high in the humid HJA, however fewer data over a shorter period of record were available when compared to the WW. More variable accuracy was observed in the arid WW watersheds where more surface water presence data were available. Similarly, in the application of the WEPP model to multiple watersheds across the Lake Tahoe Basin the agreement between simulated and observed streamflow was generally very good with the exception of a few smaller and drier watersheds that expressed unique subsurface hydrogeology (Brooks

et al., 2016). Based on these examples, the question may be whether there is enough geophysical information available for process-based models to realistically simulate low summer flows, and thus accurate streamflow permanence.

5. Conclusion

This study implemented a unique approach by using surface water presence observations to calibrate WEPP in the WW watersheds where streamflow data were not available. A more traditional approach, where WEPP was calibrated to observed streamflow, was also implemented. Accuracy of annual streamflow permanence classifications generated from the WEPP model ranged from 59 to 87 %. These accuracies are comparable to the overall accuracy of NHD streamflow permanence classifications but have better accuracy on headwater streams and are dynamic through time to account for climate conditions.

Different WEPP parameterizations produced the best accuracies

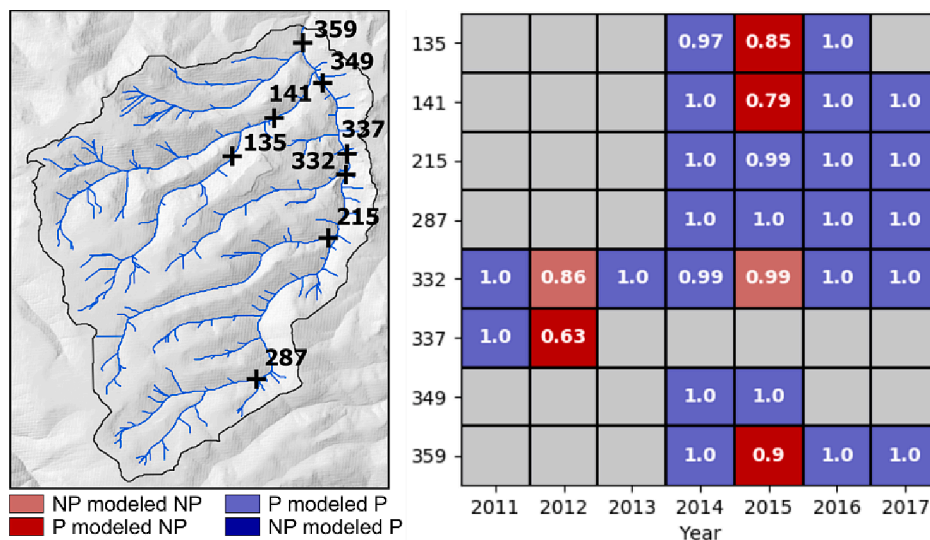


Fig. A5. Annual accuracy of permanent (P) and non-permanent (NP) classifications from the parameter set with the best annual accuracy for WW04. Numbers inside the grid show the corresponding daily accuracy (proportion) of wet and dry observations. Numbers on the Y-axis correspond the reach identifiers on which the thermistors are located (map).

based on the classification period (daily vs annual). The best daily parameterizations did not result in the best annual accuracies. Because regulatory determinations are based on an annual classification of streamflow permanence it is important to validate model results at an annual time step.

This study used streamflow and surface water presence data from intensively monitored watersheds. Study results point to some of the difficulties that arise when using point-based surface water observations to represent reach-scale streamflow permanence. Namely, observation and measurement of surface water at one location does not indicate that the entire stream reach contains surface water. This is an important consideration for design of future surface water observation campaigns.

Overall, this study indicates that physically based models have potential for modeling streamflow permanence. However, implementation of these models may be limited by validation data that describe streamflow permanence at appropriate spatial and temporal scales. This study further emphasizes the potential limits of process-based models to simulate the hydrology of small headwater watersheds in ungaged basins due to limited information on unique local topographic and geologic controls not well captured or available in ungaged catchments. Designing data collection to match the domain of model outputs will improve calibration and validation of streamflow permanence models.

CRedit authorship contribution statement

Konrad C. Hafen: Conceptualization, Methodology, Formal analysis, Data curation, Writing – original draft, Writing – review & editing. **Kyle Blasch:** Project administration, Funding acquisition, Supervision, Writing – review & editing. **Paul E. Gessler:** Project administration, Supervision, Writing – review & editing. **Jason Dunham:** Project administration, Supervision, Writing – review & editing. **Erin Brooks:** Project administration, Supervision, Writing – review & editing.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Konrad Hafen reports financial support was provided by National Science Foundation.

The remaining authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data have been published by the U.S. Geological Survey and are cited in the manuscript.

Acknowledgements

Streamflow data were provided by the HJ Andrews Experimental Forest and the Long-Term Ecological Research program, administered cooperatively by the U.S. Department of Agriculture Forest Service Pacific Northwest Research Station, Oregon State University, and the Willamette National Forest. This material is based upon work in the HJ Andrews Experimental Forest supported by the National Science Foundation, United States (Grant Number DEB-1440409). Funding was provided by the National Science Foundation Integrative Graduate Education and Research Training Program (Award Number 1249400) and the U.S. Geological Survey’s Water Availability and Use Science Program. We thank Andy Gendaszek, Emily Heaston, Nate Chelgren, and Justin Thorson for processing, collecting, and providing surface water presence data. Sherri Johnson and Steve Wondzell provided important, invaluable information about the HJ Andrews forest that helped guide study design along with Chris Konrad. Mariana Dobre aided with WEPP parameterization and automation. Marc Stewart and Pat Haluska were instrumental in designing and implementing data collection. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Appendix A

References

Abatzoglou, J.T., 2013. Development of gridded surface meteorological data for ecological applications and modelling. *Int. J. Climatol.* 33, 121–131.
 Alexander, R.B., Boyer, E.W., Smith, R.A., Schwarz, G.E., Moore, R.B., 2007. The role of headwater streams in downstream water quality. *JAWRA J. Am. Water Resour. Assoc.* 43, 41–59. <https://doi.org/10.1111/j.1752-1688.2007.00005.x>.
 Arismendi, I., Dunham, J.B., Heck, M.P., Schultz, L.D., Hockman-Wert, D., 2017. A statistical method to predict flow permanence in dryland streams from time series of stream temperature. *Water* 9, 946.
 Bailey, R.G., 2016. Bailey’s ecoregions and subregions of the United States, Puerto Rico, and the U.S. Virgin Islands (Forest Service Research Data Archive). U.S. Forest Service. <https://doi.org/10.2737/RDS-2016-0003>.

- Blasch, K.W., Ferré, T.P., Christensen, A.H., Hoffmann, J.P., 2002. New field method to determine streamflow timing using electrical resistance sensors. *Vadose Zone J.* 1, 289–299.
- Brooks, E.S., Dobre, M., Elliot, W.J., Wu, J.Q., Boll, J., 2016. Watershed-scale evaluation of the Water Erosion Prediction Project (WEPP) model in the Lake Tahoe basin. *J. Hydrol.* 533, 389–402. <https://doi.org/10.1016/j.jhydrol.2015.12.004>.
- Daly, C., Smith, J.W., Smith, J.L., McKane, R.B., 2007. High-resolution spatial modeling of daily weather elements for a catchment in the Oregon Cascade Mountains, United States. *J. Appl. Meteorol. Climatol.* 46, 1565–1586.
- Daly, C., Conklin, D.R., Unsworth, M.H., 2010. Local atmospheric decoupling in complex topography alters climate change impacts. *Int. J. Climatol.* 30, 1857–1864.
- DeWeber, J.T., Tsang, Y.-P., Krueger, D.M., Whittier, J.B., Wagner, T., Infante, D.M., Whelan, G., 2014. Importance of understanding landscape biases in USGS gage locations: implications and solutions for managers. *Fisheries* 39, 155–163. <https://doi.org/10.1080/03632415.2014.891503>.
- Dewitz, J., 2021. National Land Cover Database (NLCD) 2019 Products. <https://doi.org/10.5066/P9KZC5M4>.
- Dobre, M., Srivastava, A., Lew, R., Deval, C., Brooks, E.S., Elliot, W.J., Robichaud, P.R., 2022. WEPPcloud: an online watershed-scale hydrologic modeling tool. Part II. Model performance assessment and applications to forest management and wildfires. *J. Hydrol.* 610, 127776.
- Downing, J.A., Cole, J.J., Duarte, C., Middelburg, J.J., Melack, J.M., Prairie, Y.T., Kortelainen, P., Striegl, R.G., McDowell, W.H., Tranvik, L.J., 2012. Global abundance and size distribution of streams and rivers. *Inland Waters* 2, 229–236.
- Dun, S., Wu, J.Q., Elliot, W.J., Robichaud, P.R., Flanagan, D.C., Frankenberger, J.R., Brown, R.E., Xu, A.C., 2009. Adapting the water erosion prediction project (WEPP) model for forest applications. *J. Hydrol.* 366, 46–54. <https://doi.org/10.1016/j.jhydrol.2008.12.019>.
- Dunham, J., Schroeter, R., Rieman, B., 2003. Influence of maximum water temperature on occurrence of Lahontan cutthroat trout within streams. *N. Am. J. Fish Manag.* 23, 1042–1049.
- Dyrness, C., 1969. Hydrologic properties of soils on three small watersheds in the western Cascades of Oregon. USDA For. Ser. Res. Note PNW 111, 17.
- Flanagan, D.C., Nearing, M.A., 1995. USDA-Water Erosion Prediction Project: Hillslope profile and watershed model documentation. NSERL report.
- Flanagan, D.C., Ascough, J.C., Nearing, M.A., Lafen, J.M., 2001. The water erosion prediction project (WEPP) model, in: *Landscape Erosion and Evolution Modeling*. Springer, pp. 145–199.
- Foglia, L., Hill, M.C., Mehl, S.W., Burlando, P., 2009. Sensitivity analysis, calibration, and testing of a distributed hydrological model using error-based weighting and one objective function. *Water Resour. Res.* 45 <https://doi.org/10.1029/2008WR007255>.
- Frad, C., Johnson, S., Li, J., 2007. Stream macroinvertebrate community responses as legacies of forest harvest at the HJ Andrews Experimental Forest, Oregon. *For. Sci.* 53, 281–293.
- Fritz, K.M., Hagenbuch, E., D'Amico, E., Reif, M., Wigington, P.J., Leibowitz, S.G., Comeleo, R.L., Ebersole, J.L., Nadeau, T.-L., 2013. Comparing the extent and permanence of headwater streams from two field surveys to values from hydrographic databases and maps. *JAWRA J. Am. Water Resour. Assoc.* 49, 867–882. <https://doi.org/10.1111/jawr.12040>.
- Fritz, K.M., Nadeau, T.-L., Kelso, J.E., Beck, W.S., Mazor, R.D., Harrington, R.A., Topping, B.J., 2020. Classifying streamflow duration: the scientific basis and an operational framework for method development. *Water* 12, 2545.
- Garbrecht, J., Martz, L.W., 2004. TOPAZ user manual.
- Gendaszek, A.S., Dunham, J.B., Torgersen, C.E., Hockman-Wert, D.P., Heck, M.P., Thorson, J., Mintz, J., Allai, T., 2020a. Land-cover and climatic controls on water temperature, flow permanence, and fragmentation of Great Basin stream networks. *Water* 12, 1962.
- Gendaszek, A.S., Hockman-Wert, D.P., Dunham, J.B., Torgersen, C.E., 2020b. Stream Temperature and Water Presence Models of Willow/Whitehorse and Willow/Rock Watersheds, Oregon and Nevada (U.S. Geological Survey data release), Data release. <https://doi.org/10.5066/P9ZPBQVH>.
- Hafen, K.C., Blasch, K.W., Rea, A., Sando, R., Gessler, P.E., 2020. The influence of climate variability on the accuracy of NHD perennial and nonperennial stream classifications. *JAWRA J. Am. Water Resour. Assoc.* 56, 903–916.
- Hafen, K.C., Blasch, K.W., Gessler, P.E., Sando, R., Rea, A., 2022. Precision of headwater stream permanence estimates from a monthly water balance model in the Pacific Northwest, USA. *Water* 14, 895. <https://doi.org/10.3390/w14060895>.
- Hammond, J.C., Zimmer, M., Shanfield, M., Kaiser, K., Godsey, S.E., Mims, M.C., Zipper, S.C., Burrows, R.M., Kampf, S.K., Dodds, W., 2020. Spatial patterns and drivers of non-perennial flow regimes in the contiguous US. *Geophys. Res. Lett.* 2020GL090794.
- Heaston, E.D., Jaeger, K.L., Burnett, J., Wondzell, S.M., Chelgren, N., Dunham, J.B., Johnson, S., 2022. FLOWPER Database: StreamFLOW PERmanence field observations, April 2020 - Nov 2020, U.S. Geological Survey data release. <https://doi.org/10.5066/P9WOF1X6>.
- Hester, G., Carsell, K., Ford, D., 2006. Benefits of USGS streamgaging program—users and uses of USGS streamflow data (Report for the National Hydrologic Warning Council). David Ford Consulting Engineers, Inc.
- Jaeger, K.L., Burnett, J., Heaston, E.D., Wondzell, S.M., Chelgren, N., Dunham, J.B., Johnson, S., Brown, M., 2020. FLOWPER user guide—For collection of FLOWPERmanence field observations (Open-File Report No. 2020–1075). U.S. Geological Survey. <https://doi.org/10.3133/ofr20201075>.
- Jaeger, K.L., Olden, J.D., Pelland, N.A., 2014. Climate change poised to threaten hydrologic connectivity and endemic fishes in dryland streams. *Proc. Natl. Acad. Sci.* 111, 13894–13899.
- Jaeger, K.L., Sando, R., McShane, R.R., Dunham, J.B., Hockman-Wert, D.P., Kaiser, K.E., Hafen, K., Risley, J.C., Blasch, K.W., 2019. Probability of Streamflow Permanence Model (PROSPER): a spatially continuous model of annual streamflow permanence throughout the Pacific Northwest. *J. Hydrol. X* 2, 100005. <https://doi.org/10.1016/j.jhydroa.2018.100005>.
- Jensen, C.K., McGuire, K.J., Prince, P.S., 2017. Headwater stream length dynamics across four physiographic provinces of the Appalachian Highlands. *Hydrol. Process.* 31, 3350–3363. <https://doi.org/10.1002/hyp.11259>.
- Jensen, C.K., McGuire, K.J., Shao, Y., Andrew Dolloff, C., 2018. Modeling wet headwater stream networks across multiple flow conditions in the Appalachian Highlands. *Earth Surf. Proc. Land.* 43, 2762–2778.
- Jones, J., Grant, G.E., 1996. Peak flow responses to clear-cutting and roads in small and large basins, western Cascades, Oregon. *Water Resour. Res.* 32, 959–974.
- Jones, J.A., Swanson, F.J., Wemple, B.C., Snyder, K.U., 2000. Effects of roads on hydrology, geomorphology, and disturbance patches in stream networks. *Conserv. Biol.* 14, 76–85.
- Kampf, S.K., Dwire, K.A., Fairchild, M.P., Dunham, J., Snyder, C.D., Jaeger, K.L., Luce, C.H., Hammond, J.C., Wilson, C., Zimmer, M.A., 2021. Managing nonperennial headwater streams in temperate forests of the United States. *For. Ecol. Manag.* 497, 119523.
- Kéry, M., Schmid, H., 2004. Monitoring programs need to take into account imperfect species detectability. *Basic Appl. Ecol.* 5, 65–73.
- Knapp, A.K., Smith, M.D., Hobbie, S.E., Collins, S.L., Fahey, T.J., Hansen, G.J., Landis, D.A., La Pierre, K.J., Melillo, J.M., Seastedt, T.R., 2012. Past, present, and future roles of long-term experiments in the LTER network. *Bioscience* 62, 377–389.
- Krabbenhof, C.A., Allen, G.H., Lin, P., Godsey, S.E., Allen, D.C., Burrows, R.M., DelVecchia, A.G., Fritz, K.M., Shanfield, M., Burgin, A.J., 2022. Assessing placement bias of the global river gage network. *Nat. Sustainability* 5, 586–592. <https://doi.org/10.1038/s41893-022-00873-0>.
- Lew, R., Dobre, M., Srivastava, A., Brooks, E.S., Elliot, W.J., Robichaud, P.R., Flanagan, D.C., 2022. WEPPcloud: an online watershed-scale hydrologic modeling tool. Part I. Model description. *J. Hydrol.* 608, 127603.
- McCarthy, M.A., Moore, J.L., Morris, W.K., Parris, K.M., Garrard, G.E., Vesk, P.A., Rumpff, L., Giljohann, K.M., Camac, J.S., Bau, S.S., 2013. The influence of abundance on detectability. *Oikos* 122, 717–726.
- Messenger, M.L., Lehner, B., Cockburn, C., Lamouroux, N., Pella, H., Snelder, T., Tockner, K., Trautmann, T., Watt, C., Detry, T., 2021. Global prevalence of non-perennial rivers and streams. *Nature* 594, 391–397. <https://doi.org/10.1038/s41586-021-03565-5>.
- Milly, P.C., Betancourt, J., Falkenmark, M., Hirsch, R.M., Kundzewicz, Z.W., Lettenmaier, D.P., Stouffer, R.J., 2008. Stationarity is dead: Whither water management? *Science* 319, 573–574.
- Monteith, J., 1965. Evaporation and environment. *Symp. Soc. Exp. Biol.* 205–234.
- Moriassi, D.N., Arnold, J.G., Van Liew, M.W., Bingner, R.L., Harmel, R.D., Veith, T.L., 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Trans. ASABE* 50, 885–900.
- Nadeau, T.-L., Leibowitz, G.S., Wigington, J.P., Ebersole, L.J., Fritz, M.K., Coulombe, A.R., Comeleo, R.L., Blocksom, A.K., 2015. Validation of rapid assessment methods to determine streamflow duration classes in the Pacific Northwest, USA. *Environ. Manag.* 56, 34–53.
- Nadeau, T.-L., Rains, C.M., 2007. Hydrological connectivity between headwater streams and downstream waters: how science can inform policy. *J. Am. Water Resour. Assoc.* 43, 118–133. <https://doi.org/10.1111/j.1752-1688.2007.00010.x>.
- Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I—A discussion of principles. *J. Hydrol.* 10, 282–290.
- Sando, R., Blasch, K.W., 2015. Predicting alpine headwater stream intermittency: a case study in the northern Rocky Mountains. *Ecohydrol. Hydrobiol.* 15.
- Schultz, L., Heck, M., Hockman-Wert, D., Allai, T., Wenger, S., Cook, N., Dunham, J.B., 2017. Spatial and temporal variability in the effects of wildfire and drought on thermal habitat for a desert trout. *J. Arid Environ.* 145, 60–68.
- Schwarz, G.E., Alexander, R., 1995. State soil geographic (STATSGO) data base for the conterminous United States (USGS Numbered Series No. 2331–1258), Open-File Report. <https://doi.org/10.3133/ofr95449>.
- Soil Survey Staff, n.d. Soil Survey Geographic (SSURGO) Database [WWW Document]. Natural Resources Conservation Service, United States Department of Agriculture. URL <https://sdmdataaccess.sc.egov.usda.gov/> (accessed 3.1.20).
- Srivastava, A., Dobre, M., Wu, J.Q., Elliot, W.J., Bruner, E.A., Dun, S., Brooks, E.S., Miller, I.S., 2013. Modifying WEPP to improve streamflow simulation in a Pacific Northwest watershed. *Trans. ASABE* 56, 603–611.
- Srivastava, A., Brooks, E., Dobre, M., Elliot, W., Wu, J., Flanagan, D., Gravelle, J., Link, T., 2020. Modeling forest management effects on water and sediment yield from nested, paired watersheds in the interior Pacific Northwest, USA using WEPP. *Sci. Total Environ.* 701, 134877.
- Swanson, F.J., James, M.E., 1975. Geology and geomorphology of the HJ Andrews Experimental Forest, western Cascades, Oregon.
- Swanson, F.J., Dyrness, C., 1975. Impact of clear-cutting and road construction on soil erosion by landslides in the western Cascade Range, Oregon. *Geology* 3, 393–396.
- Tetzlaff, D., Carey, S.K., McNamara, J.P., Laudon, H., Soulsby, C., 2017. The essential value of long-term experimental data for hydrology and water management. *Water Resour. Res.* 53, 2598–2604.
- Thomas, R.B., Megahan, W.F., 1998. Peak flow responses to clear-cutting and roads in small and large basins, western Cascades, Oregon: a second opinion. *Water Resour. Res.* 34, 3393–3403.
- Thorson, J.M., Chelgren, N., Heaston, E.D., Haggerty, P.K., Dunham, J.B., 2022a. Air and stream temperature and pressure for HJ Andrews Experimental Forest, Western

- Oregon, 2020. U.S. Geological Survey data release. <https://doi.org/10.5066/P94NYDHFV>.
- Thorson, J.M., Dunham, J.B., Heck, M.P., Hockman-Wert, D.P., Mintz, J.M., 2022b. Stream Temperature in the Northern Great Basin region of Southeastern Oregon, 2016-2021, U.S. Geological Survey data release. <https://doi.org/10.5066/P924MOCB>.
- Turner, M.G., Collins, S.L., Lugo, A.L., Magnuson, J.J., Rupp, T.S., Swanson, F.J., 2003. Disturbance dynamics and ecological response: the contribution of long-term ecological research. *Bioscience* 53, 46–56.
- U.S. Geological Survey, 2016. National Water Information System (USGS Water Data for the Nation) [WWW Document]. URL <http://waterdata.usgs.gov/nwis/>.
- U.S. Geological Survey, 2019. National Hydrography Dataset (ver. USGS National Hydrography Dataset Best Resolution for Hydrologic Unit 17) [WWW Document]. URL <https://www.usgs.gov/core-science-systems/ngp/national-hydrography/access-national-hydrography-products> (accessed 11.1.19).
- USGS, 2016. National Elevation Data [WWW Document]. US Geological Survey. URL <http://nationalmap.gov/elevation.html> (accessed 3.1.16).
- Walsh, R., Ward, A.S., 2019. Redefining clean water regulations reduces protections for wetlands and jurisdictional uncertainty. *Front. Water* 1, 1. <https://doi.org/10.3389/frwa.2019.00001>.
- Ward, A.S., Schmadel, N.M., Wondzell, S.M., 2018. Simulation of dynamic expansion, contraction, and connectivity in a mountain stream network. *Adv. Water Resour.* 114, 64–82.
- Ward, A.S., Wondzell, S.M., Schmadel, N.M., Herzog, S.P., 2020. Climate change causes river network contraction and disconnection in the HJ Andrews Experimental Forest, Oregon, USA. *Front. Water* 2, 1–10. <https://doi.org/10.3389/frwa.2020.00007>.
- Williamson, T.N., Agouridis, C.T., Barton, C.D., Villines, J.A., Lant, J.G., 2015. Classification of ephemeral, intermittent, and perennial stream reaches using a TOPMODEL-based approach. *JAWRA J. Am. Water Resour. Assoc.* 51, 1739–1759. <https://doi.org/10.1111/1752-1688.12352>.
- Yu, S., Bond, N.R., Bunn, S.E., Xu, Z., Kennard, M.J., 2018. Quantifying spatial and temporal patterns of flow intermittency using spatially contiguous runoff data. *J. Hydrol.* 559, 861–872. <https://doi.org/10.1016/j.jhydrol.2018.03.009>.
- Zheng, F., Zhang, X.J., Wang, J., Flanagan, D.C., 2020. Assessing applicability of the WEPP hillslope model to steep landscapes in the northern Loess Plateau of China. *Soil Tillage Res.* 197, 104492.
- Zimmer, M.A., Kaiser, K.E., Blaszcak, J.R., Zipper, S.C., Hammond, J.C., Fritz, K.M., Costigan, K.H., Hosen, J., Godsey, S.E., Allen, G.H., 2020. Zero or not? Causes and consequences of zero-flow stream gage readings. *Wiley Interdiscip. Rev. Water* 7, e1436.