Journal of Hydrology 589 (2020) 125121

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Research papers

Insights on expected streamflow response to land-cover restoration

P. James Dennedy-Frank*,[1], Steven M. Gorelick

Stanford University, Department of Earth System Science, 473 Via Ortega, Room 140, Stanford, CA 94305, USA

ARTICLE INFO

Keywords:
Watershed services
Land-cover change
Restoration
Graph cluster analysis
Watershed simulation
Soil depth

ABSTRACT

Ecosystem service approaches to watershed management have grown quickly, increasing the importance of understanding the streamflow response to realistic land-cover change. Previous work has investigated the relationship between watershed characteristics and streamflow in catchments around the world, but little has focused on systematic relationships between watershed characteristics and streamflow *change* after land-cover restoration. To address this gap, we simulate streamflow responses to restoring 10% of watershed area from agricultural land to forest and natural pasture in 29 watersheds around the world. This change is consistent with that performed in watershed-service programs. We calculate the change in a broad array of streamflow indices for each site and use a graph-connectedness approach to cluster the sites based on the sign of the index value changes. We find three primary clusters with distinct responses to restoration. Permutation tests and effect sizes demonstrate the difference in watershed characteristics and streamflow indices across clusters. The *low-flow intensifying* sites have shallower soils and smaller saturated soil volumes. After restoration, simulated streamflow in these sites increases during relatively dry periods and declines during high-flow periods. The *high-flow intensifying* sites have larger saturated soil volumes. After restoration, simulated dry-season flow in these sites decreases. The *high-flow enhancing* sites have larger soil hydraulic conductivities than the high-flow intensifying sites. After restoration, simulated dry-season flow in these sites decreases less than in high-flow intensifying sites. The soil depth and hydraulic conductivity appear to be the characteristics that determine clusters, as clusters are not statistically related to climate, watershed location, proximity, size and shape, elevation, or pre-existing land cover. This study provides valuable understanding of land-cover restoration and the watershed characteristics that most impact streamflow *change*.

1. Introduction

Scientifically supported watershed management requires predicting how changes in land cover will affect a watershed's response to precipitation. Land-cover change may alter watershed discharge and water balance and thus have significant consequences for cities, agriculture, and the environment (Bennett and Ruef, 2016; Damania et al., 2017). However, hydrologic science has not identified clear and simple rules and hydrologic understanding to predict the impact of land-cover change on streamflow in many important contexts (Dennedy-Frank et al., 2016; Dennedy-Frank and Gorelick, 2019; Guswa et al., 2014; Hamel et al., 2017). Site-specific models can estimate the hydrologic response to land-cover changes (Khoi and Suetsugi, 2014a,b; Piniewski et al., 2014; Strauch and Volk, 2013) but that approach is not easily extended to regions where data, scientific capacity, and funding are limited (Bremer et al., 2016; Daily et al., 2009; Guswa et al., 2014; Naeem et al., 2015).

Scientifically supported watershed management is becoming particularly important as ecosystem services become a major conservation driver and watershed services are widely recognized (Bennett and Ruef, 2016; Bremer et al., 2016; Vogl et al., 2016b; Vogl et al., 2017). It is noteworthy that watershed service investments have often occurred opportunistically, rather than by targeting areas where they provide the most value. For example, early water funds in Colombia and Ecuador have led to further such funds in Brazil and Mexico, but only recently has work has tried to assess where such investments could provide the most benefit to people (Chaplin-Kramer et al., 2019; Dennedy-Frank and Gorelick, 2019).

Conservation planners and water managers would benefit from better tools to understand and quantify how watershed characteristics affect streamflow responses to land-cover change. There is a pressing need for watershed service analyses that can support decisions about

* Corresponding author.
E-mail addresses: pjdf@lbl.gov (P. James Dennedy-Frank), gorelick@stanford.edu (S.M. Gorelick).
[1] Present address: Lawrence Berkeley National Laboratory, Earth & Environmental Sciences Area, 1 Cyclotron Road, M.S. 74R-316C, Berkeley, CA 94720, United States.

https://doi.org/10.1016/j.jhydrol.2020.125121
Received 10 March 2020; Received in revised form 16 May 2020; Accepted 26 May 2020
Available online 04 June 2020
0022-1694/ Published by Elsevier B.V.

restoring land cover to conditions with less human impact by reliably estimating the consequent changes in hydrologic behavior. Even simple rules about the effects of land-cover change could help target more in-depth analyses. For example, suppose land-cover restoration decreases baseflow in sites where the warm and wet season are out of phase and increases baseflow where these seasons are in phase. Then restoration to enhance dry-season flow could be targeted to sites where the seasons are in-phase. These would serve the same role as the Indicators of Hydrologic Alteration have for many ecological stream integrity analyses (Kennen et al., 2007; Richter et al., 1996). Such analyses and rules are needed particularly in arid and tropical regimes where watershed services management approaches have seen the strongest adoption (Bremer et al., 2016; LACC & TNC, 2013; Vogl et al., 2016b). These needs are not well-addressed by classic hydrologic approaches to understanding watershed behavior that focus on: 1) predicting the discharge response to precipitation, often without historical gauge records (Brauman, 2015; Carrillo et al., 2011; Laaha et al., 2013; Sawicz et al., 2011); and 2) temperate and cold climates (Duan et al., 2006; Newman et al., 2014, 2015; Peel et al., 2000, 2010).

We present an analysis of predicted streamflow change after restoring land cover to pre-development conditions in 29 sites around the world. Site hydrologic response to simulated restoration is represented by the percent change in a broad set of streamflow indices, and a graph-connectedness approach is employed to identify similarities and differences in hydrologic responses across sites. Finally, we discuss how the identified responses compare with those seen in paired catchment and catchment classification studies. Although quantifying the effects of land-cover change on streamflow is an important topic to science and society, previous hydrologic studies have not systematically studied such effects.

Scientists have used computational hydrologic models in a wide variety of cases. The models have varied in complexity from very simple (Hamel et al., 2017; Zhang et al., 2001) to very complex (Condon and Maxwell, 2017; Moffett et al., 2012). Such models have been run at spatial scales from local (Moffett et al., 2012; Van Liew et al., 2007) to global (Döll et al., 2016; Wada et al., 2011) and at temporal scales from sub-daily to multi-annual. These models have assessed land-cover change with different levels of sophistication, from inspection of scenarios to trend analysis based on machine learning and cellular automata (Fisher et al., 2017; Kepner et al., 2012; Khoi and Suetsugi, 2014b; Kim et al., 2017; Logsdon and Chaubey, 2013). Similarly, models have integrated important socioeconomic elements with different levels of complexity, from deeply interdisciplinary studies to watershed modeling exercises (Akhavan et al., 2010; Suliman et al., 2015; Yoon, 2017). Many hydrologic models have been effectively used for watershed management and planning by reflecting human impacts on watersheds (Ficklin et al., 2013; Kroeger et al., 2019; Zhang et al., 2015), but to our knowledge no work has synthesized the results across such studies.

In particular, previous hydrologic simulations have not systematically identified the streamflow response to land-cover change, though some studies have found important site-specific effects of land cover on local water resources (Khoi and Suetsugi, 2014b; Kroeger et al., 2017; Vogl et al., 2016b). Many land management studies rely on computational hydrologic simulators to estimate the effects of land-cover change, but they: 1) are strongly biased towards the cold and temperate watersheds of the United States and Europe (e.g., in Gassman et al., 2007), and 2) sometimes suffer from sufficient inaccuracies or lack of robustness to make their use in decision-making questionable (e.g., Zaherpour et al., 2018). The computational hydrologic simulators used by academic researchers are more complex and more advanced (Kollet and Maxwell, 2006; Strauch and Volk, 2013; Therrien and Sudicky, 1996; White et al., 2011), but are not easily applied for management. They have not led the way to simpler, broadly pertinent rules that describe the hydrologic impacts of land-cover change.

The alternative catchment classification approach (Wagener et al., 2007) has similarly failed to find strong relationships between streamflow and watershed characteristics, and has often found that such relationships cannot be extended beyond the study region (Hamel et al., 2017). This approach seeks to statistically relate runoff characteristics with watershed characteristics across a broad range of sites (e.g., Carrillo et al., 2011; Laaha et al., 2013; Sawicz et al., 2011; Smith et al., 2018). In doing so, it does not isolate the streamflow *change* after shifts in land cover, instead looking solely at the relationships between land cover—and many other variables—and streamflow before any shift. These studies too are generally biased towards temperate and cold climates.

This work addresses the gap between readily applicable rules needed by land managers and the computationally intensive simulations or sophisticated statistical investigations performed in research settings that have not previously been simplified or generalized. Our effort seeks to use moderately complex hydrologic simulations in many sites to assess the effects of land-cover restoration on streamflow *change* for a broad suite of environments, a novel approach in the hydrology literature. This differs from the detailed computational simulations by investigating many sites with a simpler model. It differs from the catchment classification approach in focusing on streamflow *change* after land-cover shifts and isolating these shifts in a more detailed analysis of fewer catchments. Instead, by comparing *changes* across many sites to understand hydrologic drivers, our approach combines the ability of hydrologic simulation to isolate the effects of land-cover shifts on streamflow *change* and the discriminatory power through comparison of catchment classification. We search for simple rules that describe the effects of land-cover restoration in these well-controlled simulations to address this long-standing challenge.

2. Methods

We present a systematic study of the effects of land-cover restoration on hydrologic behavior *changes* at 29 sites from around the world and across climatic zones. From the peer-reviewed literature, we collected site-specific watershed models, obtaining data and rigorously recalibrating the simulations to improve their performance. We simulated all site models under consistent and locally realistic scenarios of restoring pre-development land cover. We then calculated the percent change in 141 streamflow indices describing different aspects of the hydrograph for each site. A graph-based clustering algorithm was used to group sites with similar hydrologic responses based on a minimally-correlated subset of streamflow indices; we call our algorithm a graph-connectedness approach. Permutation tests were then used to relate these clusters to the pre-restoration streamflow indices and watershed characteristics.

2.1. Hydrologic modeling

The selected watersheds were simulated using the Soil and Water Assessment Tool (SWAT), a model of moderate complexity often used to quantify the hydrologic impacts of land-cover change (Arnold et al., 1998; de Bressiani et al., 2015; Gassman et al., 2007). SWAT represents many hydrologic processes, including splitting runoff and infiltration, flow through many soil layers and an unsaturated zone, and baseflow from a shallow aquifer as well as resource-responsive plant growth and time-varying plant-water use. To do so, SWAT includes characteristics such as the depth, hydraulic conductivity, and porosity of up to 10 soil layers, a recession constant for a shallow aquifer, and plant growth and leaf area parameters.

Our approach integrates local high-resolution data and hydrologic understanding from carefully constructed site models across the globe in a suite of geographic and climatic zones. SWAT was selected for this

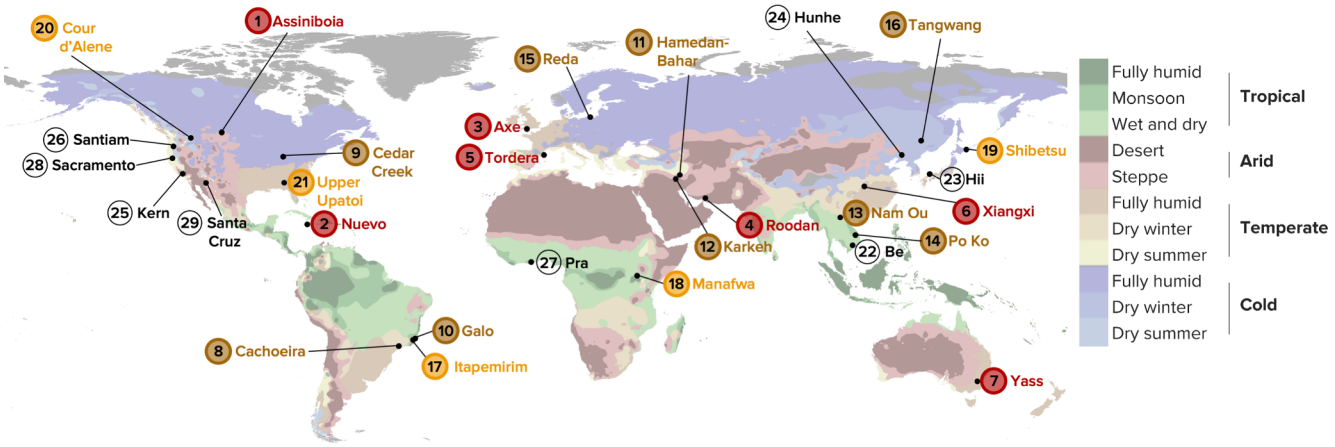

Fig. 1. *29 hydrologic simulations on 6 continents in 11 second-order Köppen-Geiger Climate Zones*. The sites are mapped here against the Köppen-Geiger Climate Zone map from Peel et al. (2007) with desatured colors for picture clarity. The names and numbers correspond to Table 1, and are color coded as the clusters in the rest of the paper (low-flow intensifying in red, high-flow intensifying in brown, and high-flow enhancing in yellow).

study because its widespread use for water resource and ecosystem service assessment in many contexts (Francesconi et al., 2016; Gassman et al., 2007) indicates acceptance by the water and land management communities, and allows the collection of consistent simulations from basins around the globe. SWAT includes two key features to investigate land-cover change effects: explicit land-cover representation that can be easily changed (Arnold et al., 2010) and a phenological vegetation model that responds to both climate and nutrient stresses (Williams, 1995). The watersheds in this study cover a broad range of land cover types, pedologic environments, and management contexts. This novel use of existing local, high-resolution site models represents a middle-way between poorly-constrained global hydrologic models (Döll et al., 2016; Wada et al., 2011) and site-specific assessments that provide strong conceptual understanding and predictive power for a single site (Vogl et al., 2016b; Yoon, 2017). The watershed simulations allow us to isolate the impacts of land-cover change on streamflow, while accounting for differences in climate, soil properties, and topography.

The site models were obtained from the authors of peer-reviewed studies, after selecting a subset from over 2500 site studies (Texas A&M; USDA ARS, 2017) based on a number of key characteristics, discussed here. Sites were chosen such that they cover Köppen-Geiger Climate Zones (Peel et al., 2007) to avoid oversampling temperate and cold sites in the US and Europe. The watersheds lie on 6 continents and in 11 second-order Köppen-Geiger Climate Zones (Fig. 1). Other simulation criteria included an area between 50 km^2 and 20,000 km^2 and at least five years of calibration data. A Nash-Sutcliffe efficiency (Nash and Sutcliffe, 1970) of 0.5 for daily discharge data or 0.7 for monthly discharge data, and a bias of $<20\%$ accounting for groundwater abstraction and known point sources ensured satisfactory model performance per Moriasi et al. (2007). Most sites had significantly more than 5 years of data, and only 3 sites had solely monthly data, with the others having daily calibration data available from the authors or publicly-accessible government sites. Table 1 provides details about the site models (for references see Supplement SI-1).

All watershed site models were rebuilt in SWAT2012 rev 637 to eliminate differences in model structure. Calibration was performed via a split-sample approach, with approximately half of the data being used for simulation training and the other half reserved for verification. Recalibrations typically reduced over-parameterization of spatially variable parameter values, while conforming to the hydrologic processes formulated by the original site model authors. For sites for which calibration was performed on a monthly basis but for which daily data were available, the daily calibration statistics were checked, and these models were re-calibrated to improve daily statistics. More detail on the calibration statistics can be found in Supplement SI-1. We refer to these re-calibrated models as baseline scenarios throughout the paper.

2.2. Restoration scenario representation

Land-cover change was simulated as restoration of 10% of the watershed area to pre-development conditions. Pre-development land cover was selected based on known ecological suitability in the watershed from current land-cover and previous estimates (Ramankutty and Foley, 1999). Land-cover change was restricted to increasing one land cover and decreasing a second (see Supplement SI-2 for details). If 10% of the land cover could not be changed, index value changes were linearly extrapolated to 10% from a baseline of 0% index change for 0% land-cover change. The sites were required to have at least 4% land-cover change to avoid over-interpreting this extrapolation.

2.3. Clustering with streamflow indices

We assess the similarities and differences between the watersheds' responses to restoration using a graph-connectedness approach to define clusters of watersheds that respond in similar ways. To do so, we build a graph that has watersheds as the nodes and defines these nodes' distance from each other by the similarity of the change in numerous streamflow index values after simulated restoration. In the resulting graph, watersheds that are close to each other have a similar response to restoration, and those far from each other different responses to restoration.

That graph-connectedness approach is detailed in the following subsections. First, we define a broad set of streamflow indices and calculate the index value changes between baseline and restoration scenarios. We develop measures to reduce the effect of a small number of anomalously large index value changes. Second, we use a dimension reduction approach to select a set of minimally correlated index value changes that still have clear meanings. Finally, we construct a graph such that the watersheds with more similar responses to restoration are closer to each other. We designate groups of close watersheds a cluster.

2.3.1. Streamflow indices

A set of 141 streamflow indices is gathered from published literature that investigated changes in river discharge, tested the similarities and differences in behavior across catchments, and assessed local water resources for planning purposes. These indices include many of those from the Indicators of Hydrologic Alteration (Richter et al., 1997; Richter et al., 1996) and extensions to that work (Gao et al., 2009;

Table 1
List of the simulated watersheds, their basic characteristics, and simulation periods
Color coding and table grouping follows the scheme of Fig. 2 for easy reference.

Site	Country	Climate zone	Area [km^2]	Latitude [°]	Longitude [°]	Simulation years
1. Assiniboia Catchment (Asb)	Canada	Arid steppe	50	49.6	-105.9	1978-2003
2. Rio Nuevo (RN)	Jamaica	Tropical rainforest	95	18.3	-77.0	2002-2007
3. River **Axe** (Axe)	United Kingdom	Temperate, no dry season	400	50.8	-3.0	1988-1997
4. Roodan Watershed (Roo)	Iran	Arid desert	10333	27.7	57.3	1989-2008
5. Tordera Catchment (Tor)	Spain	Temperate, dry summer	858	41.8	2.6	1996-2004
6. Xiangxi River (Xia)	China	Temperate, dry winter	3175	31.2	110.8	1981-1993
7. Yass River Catchment (Yas)	Australia	Temperate, no dry season	1597	-36.0	147.5	1993-2011
8. Cachoeira River Basin (Ca)	Brazil	Temperate, dry winter	291	-23.0	-46.3	1987-2001
9. Cedar Creek Watershed (CC)	U.S.	Cold, no dry season	706	43.0	-90.0	1993-2013
10. Galo Creek (Gal)	Brazil	Tropical monsoon	942	-20.3	-40.7	1997-2003
11. Hamedan-Bahar Watershed (H-B)	Iran	Arid steppe	2387	34.9	48.5	1993-2008
12. Karkeh Watershed (Krk)	Iran	Arid desert	51954	32.5	48.0	1990-2002
13. Nam Ou Basin (NO)	Laos	Cold, dry winter	26180	21.2	102.3	1994-2003
14. Po Ko Catchment (PK)	Vietnam	Tropical monsoon	3232	16.0	107.5	2000-2011
15. Reda Catchment (Rda)	Poland	Cold, no dry season	483	54.6	18.1	1998-2006
16. Tangwang Catchment (Tw)	China	Cold, dry winter	20807	46.7	129.9	1965-2000
17. Itapemirim River (Ita)	Brazil	Tropical monsoon	2237	-20.5	-41.5	1993-2000
18. Manafwa River Basin (Mnf)	Uganda	Tropical savannah	353	0.9	34.2	1956-1961
19. Shibetsu Watershed (Shi)	Japan	Cold, no dry season	672	43.6	144.9	2004-2008
20. Cour d'Alene South Fork (CdA)	U.S.	Cold, dry summer	733	47.5	-116.1	1991-2010
21. Upper Upatoi Watershed (UU)	U.S.	Temperate, no dry season	878	32.5	-84.7	2004-2011
22. Be River (Be)	Vietnam	Tropical savannah	7347	11.8	107.1	1981-1989
23. Hii River Basin (Hii)	Japan	Temperate, no dry season	914	35.2	132.9	1988-2006
24. Hunhe Catchment (Hun)	China	Cold, dry winter	9285	46.8	132.0	1958-1965
25. Kern River (Krn)	U.S.	Arid desert	6234	35.6	-118.4	1950-2005
26. North **Santiam** River (NS)	U.S.	Cold, dry summer	558	44.6	-121.9	1950-2005
27. Pra River Basin (Pra)	Ghana	Tropical savannah	19584	6.4	-1.4	1964-1991
28. Sacramento River (Sac)	U.S.	Cold, dry summer	18836	41.3	-120.9	1950-2005
29. Santa Cruz Watershed (SC)	U.S.	Arid steppe	9090	31.8	-110.8	1987-2006

Kennen et al., 2007; Olden and Poff, 2003); catchment classification work (Carrillo et al., 2011; Sawicz et al., 2011); and current guidance that summarizes classic water resources approaches used by the USGS and WMO (Hortness, 2006; Martin et al., 2016; Risley et al., 2008; World Meteorological Organization, 2009; Ziegeweid et al., 2015). These indices are selected because of their use in water resource planning and in describing flow differences across watersheds. A full list of the indices and their literature citations are presented in Supplement SI-3.

For each site, the percent change in each of the 141 streamflow index values is calculated after calculating the index for both the existing land-cover and the pre-development land-cover restoration scenario. The percent change normalizes the effects of total discharge magnitude and watershed size, and it readily permits statistical tests and other simple analyses. Percent change is defined such that an increase in the index value results in a positive percent change, and vice versa.

Preliminary analyses with these index value changes found that a small percentage (~0.5%) of them had anomalously large errors that dominated regression or dimension reduction approaches. We address this by adopting two measures that remove the effect of these anomalously large changes in index values. To build the graph (discussed presently) we use the index sign change, i_s, which simply applies the sign function (±) to the index value change, with a tolerance that sets

very small values to 0 (see Supplement SI-4); hereafter, we simply reference the sign of the index value change. For statistical tests of the indices across different clusters in the graph, we use a truncated index value change, $\boldsymbol{i_t}$. $\boldsymbol{i_t}$ simply takes the percent change of the index value and truncates it to ±25%, so that all index values are within the range [−25%, 25%]. This approach reduces the magnitude of the ~0.5% of anomalously large index value changes to reduce their influence.

2.3.2. Selecting a minimally correlated subset of indices

Many of the streamflow index values and their changes are strongly correlated to one another (Gao et al., 2009; Olden and Poff, 2003). Therefore we use principle component analysis (PCA) (Pedregosa et al., 2011) on the index sign change, $\boldsymbol{i_s}$, to avoid overweighting components in the graph analysis. Components were selected to cover 99% of the variance to ensure that important components are not ignored. The principal components and their major contributing indices are shown in SI-5.

Because the PCA components are linear combinations of the streamflow index value changes, these components are difficult to interpret. Thus, they do not provide useful guidance for predicting hydrologic response to land-cover change. To address the challenge of interpreting the PCA components we use varimax rotation. Varimax rotation finds the set of streamflow index value changes that maximizes the sum of the variance of the square loadings. That is, it seeks to have each component represented by one (or a few) very large loadings from streamflow index value changes and no loadings from other streamflow index value changes. These strongly loaded streamflow index value changes, of the same dimension as the PCA, represent the set of index value changes most similar to the purely orthogonal PCA components. Varimax rotation is performed with Python's factor-rotation package (van der Schans, 2015). These index value changes represent the same variance in the system as the PCA components but provide a clearer interpretation at the cost of increased correlation between the indices. Here, these reduced dimension index value changes are used to construct the graph.

The 26 streamflow index value changes selected through PCA and varimax rotation are shown in Table 2, along with basic definitions and their primary source. We group them into 6 families: low flow, monthly flow, monthly flow variance, integrated signal, daily flow, and high flow. The low-flow family describes the changes in magnitude, frequency, duration, and variability of low flows. "Low flow" here is defined alternately as all flows less than: 1) the 75% exceedance flow, 2) the minimum 30-day average flow, or 3) the minimum 7-day average flow. In addition, a digital filter is run to estimate the baseflow recession (Arnold and Allen, 1999), and the resulting change in recession constant is included. The monthly flow family describes the change in maximum or minimum flow in a given month. The monthly flow variance family describes the changes in variability in the monthly average flow for a given month, or the variance among monthly flows. The integrated signal family brings together multiple years of flow through the flow duration curve. The daily flow family deals with either the daily change of flow or with particular days when criteria are met. The high-flow family describes the changes in the magnitude, timing, and duration of high flow periods, defined as larger than 3× or 7× median, 25% exceedance, or the average maximum daily or 7-day flow.

2.3.3. Clustering approach

We use a graph-connectedness approach with a spring-based visualization (Kamada and Kawai, 1989) to cluster and investigate the similarities and differences among the watersheds' streamflow *changes* after simulated land-cover restoration. The approach is then extended to help identify watershed characteristics that cluster in the same manner as the changes in streamflow index values, and thus are likely related to the hydrologic responses to land-cover change. Although graph-based clustering has been used in other fields (e.g., Hartuv et al., 2000), to our knowledge our application of the graph-connectedness approach in searching for watershed similarity is novel. As shown, it yields useful multi-dimensional similarity in clusters based on relatively few assumptions. We discuss the graph-connectedness approach to clustering here, and statistical tests of features' correlation with the graph-based clusters in Section 2.4. Note that the term "features" as used here refers to both streamflow index values (not their changes) and watershed characteristics.

The graph-connectedness approach builds a network with each site serving as a node. The distance between the nodes is set by a measure of similarity between the change in streamflow index values at each site and implemented as a graph link (sometime referred to as an "edge" in other applications) between the nodes with a length proportional to this distance. We define the similarity as $\boldsymbol{l} = 1 - \boldsymbol{r}$, where $\boldsymbol{r}$ is the correlation of the signs of the observed change in streamflow index values, discussed presently. Thus, smaller distances represent more similar sites and larger distances less similar sites. The graph shape is determined by a model that minimizes the "energy in a spring" (Kamada and Kawai, 1989). This model is analogous to a physical system in which springs with different resting lengths are linked between nodes, with the energy in each spring increasing as the square of the link's length away from the resting (0-energy) distance, consistent with a simple spring obeying Hooke's Law. Once all the springs are linked, the system is set free, and it eventually comes to rest in a minimum energy state (Hagberg et al., 2010). This enables easy visualization of the relationships between nodes, and the delineation of clusters.

The graph is built using the sign-correlation as the similarity measure. This is the broadest measure of index value change similarity we could determine, and was selected to reduce the effects of the ~0.5% of anomalously large index value changes that otherwise dominate measures of similarity. The sign-correlation simply represents the percentage of streamflow indices showing value changes of the same sign between two sites under land-cover restoration. The sign-correlations can take any value between 0% (no index value changes have the same sign between two sites) to 100% (all index value changes have the same sign between the two sites). Three signs are possible for each index value change, 1 is an increase, −1 is a decrease, and 0 is no change (see SI-4 for details). We assume that sites that have the same direction of change of an index value are more likely to behave similarly.

A graph with links between each pair of nodes would include near-random connections between many sites, which we address by including only links that connect sites more strongly/weakly than would be expected by chance. To determine the chance level of connection between the streamflow index value changes for two watersheds we develop a simple null model with the same structure, discussed in Supplement SI-6.

2.4. Permutation tests and effect size to identify watershed characteristics related to clusters

Permutation tests are employed to investigate which features might be associated with which clusters. Permutation tests are a non-parametric statistical test (Good, 2005). Consider a group of watershed feature metric values, $\boldsymbol{v}$, split into two clusters with sizes $\boldsymbol{c}_1$ and $\boldsymbol{c}_2$. The difference between the mean value of the $\boldsymbol{v}$ in each cluster is calculated. This difference is then checked against clusters of the same size created from permutations of all of the values, $\boldsymbol{v}$. From a large number of permutations, we estimate the probability that the mean feature values are statistically distinct across the observed clusters. We use the null hypothesis that the means of the features in the two clusters are the same. The test provides a probability that an observed difference of feature metric values across clusters is statistically significant, suggesting that the feature is related to the clusters.

Permutation tests are run on three different features against the clusters determined through the graph-connectedness approach: 1) the changes in streamflow index values as determined previously, the signs of which are used to build the graph; 2) the watershed characteristics,

Table 2
Indices used for graph construction, along with a brief description and the source. * 1. Arnold and Allen, 1999; 2. Clausen and Biggs, 2000; 3. Haines et al., 1988; 4. Hughes and James, 1989; 5. Poff, 1996; 6. Puckridge et al., 1998; 7. Richter et al., 1996; 8. Sawicz et al., 2011; 9. Wood et al., 2000; 10. New metric developed for this project.

Index name	Description	Source*
Low-flow indices		
Low flow pulse count	Average annual number of pulses with flow below 75% exceedance flow	7
Variability in low flow pulse duration	Standard deviation of average annual duration of flow pulses below 75% exceedance flow	7
Annual minima of 30-day discharge/median flow	Average lowest 30-day flow for each year, divided by all-time median flow	2
Baseflow recession constant, β, from filter	Baseflow recession constant estimate from automated digital filter	1
Variability in baseflow index 1	Standard deviation of annual 7-day minimum flow divided by that year's mean flow (non-parametric skew)	7
Monthly flow		
Mean of February maximum flow	Mean value of maximum daily flow of each February in record	9
Mean of June minimum flow	Mean value of minimum daily flow of each June in record	9
Monthly flow variability		
CV of February, March, May, August, September, December flow	Standard deviation of average monthly February, March, May, August, September, December flow divided by mean February, March, May, August, September, December flow	7
Skewness of monthly flows	Mean of average monthly flows minus median of average monthly flows divided by median of average monthly flows (non-parametric skew)	6
Integrated flow measures		
Concavity index	Natural log of 33% exceedance flow - natural log of 66% exceedance flow divided by 0.33	8
Daily flow measures		
Variability in annual reversals in flow direction	Median number of days each year when the change in flow from one set of days is a different direction than the preceding set of days	7
Variability in Julian date of annual minimum	Coefficient of variation of Julian date of minimum flow each year, computed using circular components to address cross-year transfer	7
Median difference in $\log_{10}$(flow) for two consecutive days of increasing flow	Calculate difference in $\log_{10}$(flow) for each set of days, and take the median of all values that are positive	2
Peak flow		
High peak flow at 3× median flow	Average flow for all flow events greater than 3× full-record median flow divided by full-record median flow	2
High flow volume at 3× median flow	Median volume of flow while flow is above 3× full-record median each year, divided by full-record median	2
Average duration of flow events above 7× median flow	Median across years of total number of days flow is above 7× median flow divided by the total number of such events	7
CV of magnitude of maximum annual flows at 7 days	Standard deviation of mean average flow during annual maximum 7-day flow divided by mean of mean average flow during annual maximum 7-day flow	7
CV of high flow pulse count at 25% exceedance flow	Standard deviation of annual number of flow events above 25% exceedance flow divided by the mean number of such flow events	4
Variability of high flow pulse duration	Standard deviation across years of total number of days flow is above 75% exceedance flow divided by the total number of such events divided by the mean of total number of days flow is above 75% exceedance flow divided by the total number of such events	7
Categorization of discharge with strong peak months	Fraction of flow that occurs in 3-month period with highest flow	3
% annual discharge in 30 days before peak flow		10

and 3) the streamflow index values based on pre-restoration land-cover. Permutation tests on the changes in streamflow index values help determine which indices drive the clustering. The Python package permute (Stark et al., 2015) is used for permutation tests.

The effect sizes of the feature differences across clusters indicate the degree to which clusters are distinguishable by that feature. Distinguishable features likely contribute to watershed responses to land-cover change. We use Cohen's $\boldsymbol{d}$, a standard effect size metric (Welkowitz et al., 2012) that reflects the difference in the means of the clusters:

$$\boldsymbol{d} = \frac{\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2}{\mathscr{S}} \tag{1}$$

with $\bar{\boldsymbol{x}}_{\boldsymbol{c}}$ the mean of cluster $\boldsymbol{c}$, and $\mathscr{S}$ the pooled standard deviation between the two clusters. $\mathscr{S}$ is defined as:

$$\mathscr{S} = \sqrt{\frac{(\boldsymbol{m}_1 - 1)\boldsymbol{s}_1^2 + (\boldsymbol{m}_2 - 1)\boldsymbol{s}_2^2}{\boldsymbol{m}_1 + \boldsymbol{m}_2 - 2}} \tag{2}$$

with $\boldsymbol{m}_{\boldsymbol{c}}$ the number of members of cluster $\boldsymbol{c}$, and $\boldsymbol{s}_{\boldsymbol{c}}^2$ the sample variance for cluster $\boldsymbol{c}$. We search for effect sizes greater than 1, so that the means differ by more than the pooled standard deviation, which we expect indicates that such a feature is substantially different and thus distinct across the clusters. This indicates that the feature likely contributes to differences in hydrologic behavior. Cohen's $\boldsymbol{d}$ is similar to the inverse of the coefficient of variation, seeking to make sure that the difference in means is large relative to the standard deviation.

3. Results

We first discuss the graph and define the three identified clusters of watershed response to land-cover change: the *low-flow intensifying* sites, the *high-flow intensifying* sites, and the *high-flow enhancing* sites. Next, permutation test results are presented for three categories: a) changes in streamflow index values after land-cover change, b) watershed characteristics, and c) pre-restoration streamflow index values. In the *low-flow intensifying* sites, simulated streamflow increases during low-flow periods and decreases during high-flow periods. The opposite occurs in the *high-flow intensifying* sites. The *low-flow intensifying* sites contain shallower soils and smaller soil water volumes than the *high-flow intensifying* sites. The *high-flow enhancing* sites have soil depths similar to the *high-flow intensifying* sites but also larger hydraulic conductivity. As a result, the *high-flow enhancing* sites change in the same direction as the *high-flow intensifying* sites, but with a smaller magnitude. Finally, we explain additional features of different clusters of watersheds for which streamflow responds similarly to land-cover change.

3.1. Graph-connectedness clustering

3.1.1. Sign-correlation of restoration scenario shows three coherent clusters

Fig. 2 shows the graph built from the sign-correlation of the changes in streamflow index values for 10% restoration of pre-development

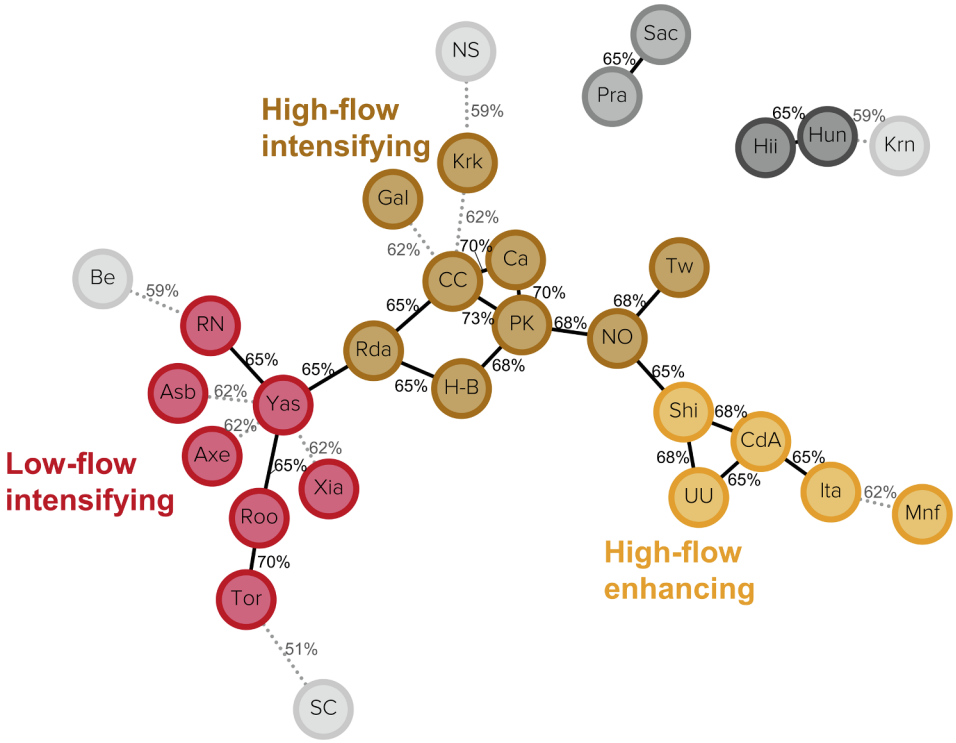

Fig. 2. Three primary clusters are apparent using the graph-connectedness approach to cluster the sites in this study with the sign-correlation of changes in streamflow indices after restoration to pre-development land cover as the distance metric (described in methods). The abbreviations used here are noted in Table 1.

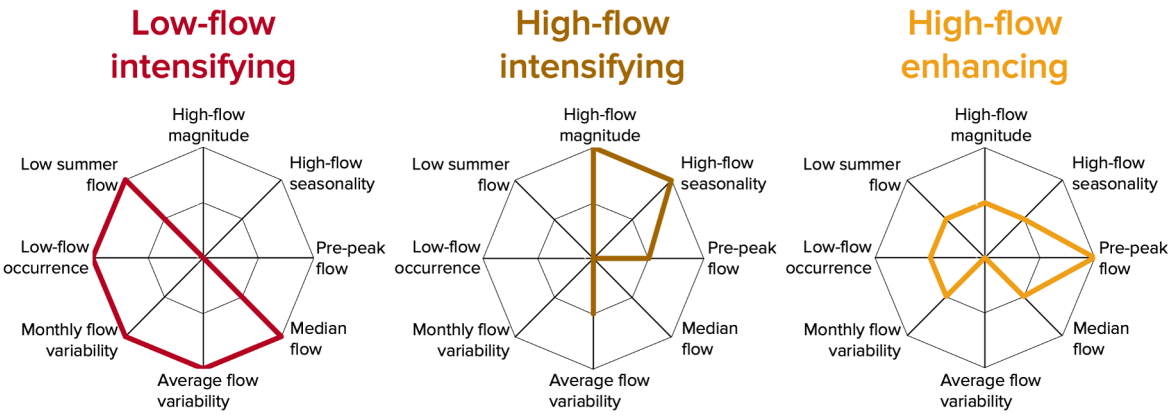

Fig. 3. The watersheds in the three clusters show distinctly different responses to land-cover change. Radar graphs plot the rank of a set of streamflow indices, with the outermost line the largest increase (smallest decrease), the inner line the second-largest increase (second-smallest decrease), and the center point the smallest increase (largest decrease).

land. There are three primary clusters, shown as red, brown, and yellow. In addition, there are small clusters consisting of 2 or 3 sites, and several sites that are not in a cluster with weak similarity to any other site, which we designate in dark grey shades and light grey shades, respectively.

We define the three primary clusters, apparent by visual inspection (Fig. 2): *low-flow intensifying*, *high-flow enhancing*, and *high-flow intensifying*. The clusters are named after the behavior of related sets of indices, as shown in Fig. 3 and discussed below.

The clusters and graph shape are determined by the sign-correlations, noted next to the links in Fig. 2, as discussed in the methods. Clusters were identified as those watersheds showing sign-correlations with a significance level, α, of <0.05 in the random null model. This corresponds to a sign-correlation of $>61\%$. Watersheds with sign-correlations that were all $<61\%$ were set aside as outside the three primary clusters (see Supplement SI-6). There are 7, 9, and 5 sites in the low-flow intensifying, high-flow intensifying and high-flow enhancing clusters, respectively.

The graph as shown uses a stronger significance level, $\alpha = 0.025$, for clarity. Links stronger than the random null model with significance level $\alpha = 0.025$, a sign-correlation $>63\%$, are shown as solid black lines. Links with a sign-correlation $<63\%$ but still included per the methods section criteria are noted with dotted grey lines. In addition, sign-correlations weaker than the random null model with $\alpha = 0.025$, $<22\%$, push away dissimilar watersheds in the graph, but are not shown for clarity. They are included in Supplement SI-7, which lists all sign-correlations used to construct the graph.

Sites falling in between clusters that met the 61% sign-correlation threshold were assigned to only one of the clusters based on their highest sign-correlation to all members of the competing clusters. We identified the Yass River Catchment, in Australia, as a member of the low-flow intensifying cluster based on its status as a hub with the

Table 3
Mean changes in streamflow for specific indices that show significant differences across clusters from pre-restoration to restoration land-cover change. [1]See Table 2 for details on index calculation; [2]Negative values for these indices are good for regulating flow; [3]These indices are not included in Fig. 3; [4]Digital filter used for automatic baseflow separation from Arnold and Allen (1999); [5]Median flow is not selected by varimax rotation, but is included because it is useful to understand the clusters.

Generalized index name	Index name[1]	Low-flow intensifying	High-flow enhancing	High-flow intensifying
Low-flow indices				
Low-flow occurrence[2]	Low-flow pulse count	−4%	+2%	+10%
Baseflow recession[3,4]	Baseflow recession constant, β, from digital filter	+2%	−5%	−5%
Monthly flow				
Low summer flow	Mean minimum June flow	+1%	-<1%	−8%
Monthly flow variability				
Monthly flow variability[2]	CV of monthly flow: February, March, May, August, September, December	−1%	~0%	+1%
Integrated flow measures				
Median flow[5]	Median flow[5]	+4%	−2%	−6%
Average flow variability	Concavity index	+2%	-<1%	-<1%
High flow measures				
High-flow magnitude[2]	Average flow above 3x median flow	−6%	+1%	+5%
Pre-peak flow[2]	Percent discharge in 30 days before peak flow	−3%	+ <1%	−1%
High-flow seasonality[2]	Percent flow in highest-flow 3 months	−1%	+ <1%	+1%
High-flow volume[2,3]	Volume of flow above 3x median flow	−6%	−1%	−5%

strongest connection for many sites. The Reda Catchment in Poland was identified as a member of the high-flow intensifying cluster based on its links to two high-flow intensifying cluster members of about the same strength as its connection to the Yass. The Nam Ou, in Laos, was identified as a member of the high-flow intensifying cluster because of its stronger link to the Po Ko in Vietnam than the Shibetsu in Japan. The Shibetsu was selected as a member of the high-flow enhancing cluster because of its stronger connections to the Upper Upatoi in Georgia, US and the S. Fork of the Cour d'Alene in Idaho, US.

3.1.2. Cluster similarities and membership

The differences in the clusters are fundamentally multivariate, and there is no single streamflow index that individually defines them. However, there are a number of consistent streamflow responses that vary across clusters. Table 3 shows the mean change of a set of streamflow index values that are statistically different across clusters.

Fig. 3 shows that the watersheds in each cluster have different responses to land-cover change. The figure shows the rank order of values for a subset (8 of 10) of the indices from Table 3, including the *median flow* value, as radar plots. The signs of the index values were reversed for *low-flow occurrence* and *monthly flow variability* because a reduction in these index values leads to increasing low-flow. The values were ranked from smallest to largest as 1 to 3. The outer rim of the radar plot axis is 3, and the center point is 1. Ties resulted in repeated rank values. The *baseflow recession* and *high-flow volume* were omitted because they are unintuitive and do not change the results. The *median flow* is one of the 141 pre-dimension reduction indices, and was included because it provides a simple measure of central tendency.

Comparing the pre-restoration simulation to the restoration land-cover simulation, the low-flow intensifying cluster generally sees a rise in the flow during relatively dry and average periods, and a decline in the discharge during high-flow periods. This is a regulating behavior that is typically desired for watershed services, shifting discharge from high-flow periods to low-flow periods, both to reduce the chance of flood damage and to provide consistent dry season flow. This can be seen clearly by the largest rank in low-flow index value changes including *low-flow occurrence* and increase in *low summer flow,* and the smallest rank in high-flow indices such as *high-flow seasonality*.

The high-flow intensifying cluster, in contrast, sees a decrease in discharge during low-flow periods and an increase in the relative high flows as well as flow variability. The high-flow value changes show a large rank, including the largest rank for *high-flow magnitude* and *high-flow seasonality,* and the second largest for *pre-peak flow*. This cluster generally has the smallest rank in the low-flow and intermediate flow indices, including 10% more *low-flow occurrences* and 8% less *low summer flow*. This reduction in low-flow is sometimes a concern when considering the hydrologic effects of restoration.

The high-flow enhancing cluster sees similar responses to the high-flow intensifying cluster, but the magnitudes are smaller so the enhancing cluster generally has the middle rank. It does have the largest value change for *pre-peak flow,* and is the only cluster that has a positive mean value change for that index. However, it has the middle rank for *high-flow magnitude, high-flow seasonality,* and *low-flow occurrence.*

3.1.3. Indices across clusters

The clusters are further distinguishable by their marked differences in mean value changes of specific indices. First, consider the *high-flow magnitude* response to land-cover restoration. The low-flow intensifying cluster has a decline of 6% of the *high flow magnitude* (see Table 3), which is statistically different ($\alpha = 0.05$) from the high-flow intensifying and high-flow enhancing clusters that show rises of 5% and 1%, respectively. The permutation tests show particularly low p-values ($p < 0.0025$; see methods section for null hypothesis), as well as effect sizes of 2.4 and 1.7 for the low-flow intensifying cluster compared to the high-flow intensifying and high-flow enhancing clusters. Since the difference between the rise in the *high-flow magnitude* in the high-flow intensifying and low-flow intensifying clusters is more than twice their variability, we argue they represent fundamentally different responses.

Second, changes in the average flow also highlights the hydrologic response differences of the three clusters resulting from restoration land-cover change. Consider the *median flow,* or 50% exceedance flow, of the sites within each cluster (from the list of 141 indices). The *median flow* value in the high-flow intensifying and high-flow enhancing clusters show a decline of 6% and 2%, respectively, while the low-flow intensifying cluster shows a rise of 4%. The high-flow intensifying and low-flow intensifying clusters have statistically different means with $p = 0.0003$, and the high-flow intensifying and high-flow enhancing clusters have different means with $p = 0.008$. These clusters have effect sizes of 1.7 and 1.5, respectively. It is noteworthy that the change in median flow is found to be statistically different across the three clusters, but was not selected by the dimension reduction approach.

Third, index value changes reflecting land-cover restoration impact on low flow show statistically significant differences across clusters (see Table 3). Specifically, comparing to the pre-restoration case, the low-

flow intensifying cluster has reduced mean of *low-flow occurrences* rather than the increased mean of *low-flow occurrences* in the high-flow intensifying cluster, with p-value = 0.0026 and effect size 1.9. The high-flow enhancing cluster has a mean change that differs from the high-flow intensifying cluster with a p-value of 0.03 and effect size of 1.2.

Together, the differences and similarities across clusters reveal a consistent set of themes, as shown in Fig. 3. In the low-flow intensifying cluster, the high-flow decreases, while in the high-flow intensifying cluster, and less so in the high-flow enhancing cluster, high flow increases relative to the median flow in each watershed. In the low-flow intensifying cluster watersheds, the water that went to high flow in the pre-restoration land-cover case, is instead stored and released during low flow periods. In the high-flow intensifying and high-flow enhancing clusters, restoration affects the modeled flow in one of two ways: 1) the water that comprises low flow and average flow are each diminished because that water is routed more quickly to the high-flow regime, and 2) greater evapotranspiration lowers both high flow and median flow. In both of these cases, there is a higher proportion of high flow to median flow.

3.2. Soil characteristics and slope of flow duration curve are strongly related to clusters

We explored which watershed characteristics and streamflow index values based on pre-restoration land-cover are most strongly related to the three clusters as determined by permutation tests. The most obvious characteristics, such as geographic location and climate zone, do not show strong relationships to the clusters, as shown in Fig. 1. For example, the high-flow intensifying cluster includes sites from Vietnam, Iran, Brazil, Poland, and Indiana. The high-flow enhancing cluster includes sites from Brazil, Japan, Idaho, and Georgia, USA.

Fig. 4 shows the values of watershed characteristics and streamflow indices for each site, while Table 4 presents the results of statistical tests and effect sizes. In each section of Fig. 4, each column represents a site, and the columns are grouped according to the clusters. Each box shows the value of the feature, defined on the left, at that site through both color and the printing of the value in the box. The color scale on the right side of the boxes corresponds to these values. These color scales are sometimes truncated to better represent the distribution. Table 4 provides the permutation test p-value and the effect size across clusters.

3.2.1. Importance of soil depth, hydraulic conductivity, and water content

Soil characteristics are the only watershed characteristics that predict cluster membership (Fig. 4a). The soil depth and saturated water content are both significantly smaller ($\alpha = 0.05$) in the low-flow intensifying sites than in either the high-flow intensifying or high-flow enhancing sites, with effect sizes greater than 1. These results show that in our simulations high flow is reduced and low-flow increased in only those sites that have relatively thin soils, which might be improved through restoration of land cover. Soil clay percentage and sand percentage differ statistically ($\alpha = 0.05$) between the low-flow intensifying and high-flow enhancing clusters. The high-flow enhancing cluster has sandier soils and the low-flow intensifying cluster has more clay-rich soils. This suggests that in our simulations increasing infiltration in low-flow intensifying cluster sites may lead to a diminution in high flow and rise in low-flow under land-cover restoration. These soil characteristics are not independent, but all are presented here to show the differences and similarities.

Soil hydraulic conductivity, in contrast, is significantly smaller ($\alpha = 0.05$) in the high-flow intensifying cluster than the high-flow enhancing cluster, and is the only physical characteristic that separates these two clusters. This suggests that the primary difference between the sets of watersheds in their respective clusters is simply how fast water moves through the soils. In the high-flow enhancing cluster sites, the water moves quickly through the soil and recharges groundwater, so the response to land-cover restoration is lower than that in the high-flow intensifying cluster sites. The high-flow intensifying cluster sites,

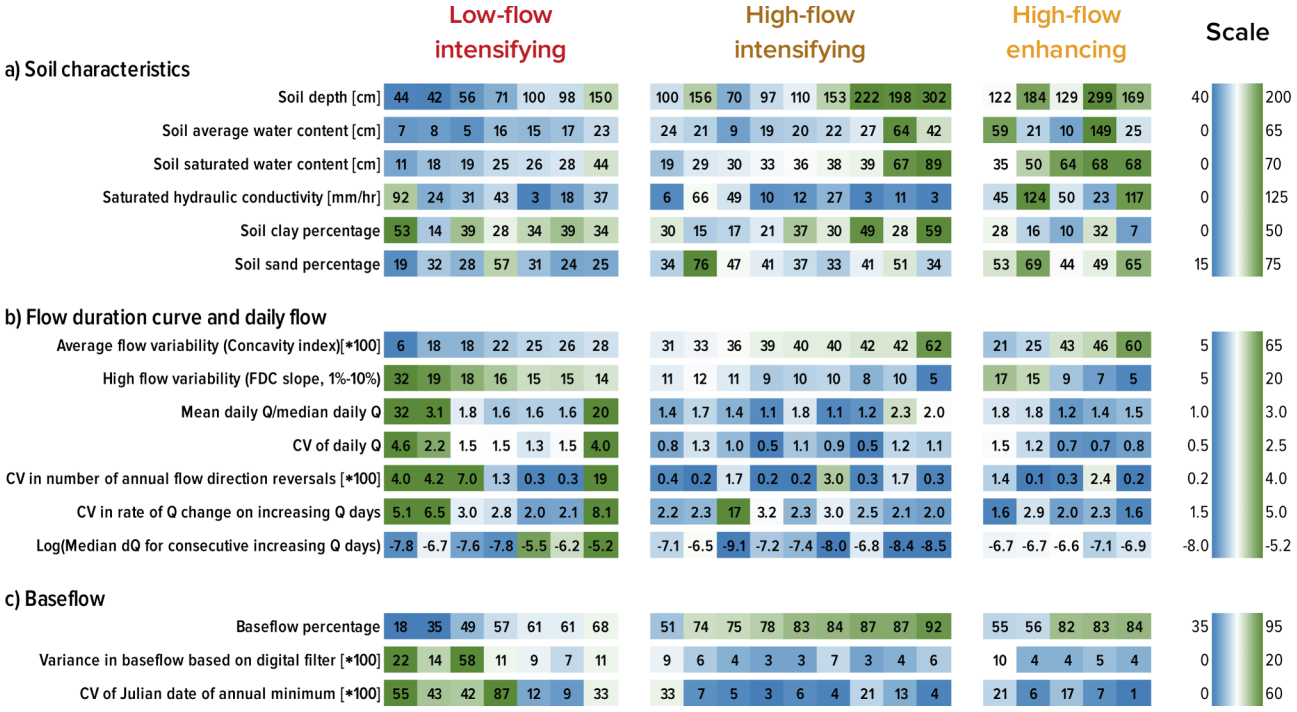

Fig. 4. The clusters show differences across watersheds in a few features: a) soil depth and saturated water content; b) pre-restoration average- and high-flow variability index values; and c) pre-restoration baseflow index values. These differences are significant and substantial (Table 4). Each box represents a watershed feature (listed on the left, with appropriate units) at a site included in the clusters, with the feature's color scale to the right. The color scales provide a realistic sense of the distribution of features across clusters; in some cases, feature values extend beyond the color-bar limits to visually emphasize the distribution. Note that the site ordering in the columns is not the same in the three figure sections.

Table 4
Results of statistical tests for discriminating watershed features across clusters.[1]p-value for the two-sample permutation test with null hypothesis that feature's mean value across clusters is the same; [2]Effect size (Cohen's d) for the difference between the means; Statistically significant p-values ($\alpha = 0.05$) and substantial effect sizes (>1) bolded for emphasis.

	Low-flow intensifying vs. High-flow intensifying		Low-flow intensifying vs. High-flow enhancing		High-flow intensifying vs. High-flow enhancing	
	p-value[1]	*Effect size*[2]	*p-value*[1]	*Effect size*[2]	*p-value*[1]	*Effect size*[2]
a) Soil characteristics						
Soil depth [mm]	**0.02**	**1.32**	**0.007**	**2.06**	0.55	0.36
Soil average water content [mm]	**0.01**	**1.17**	**0.05**	**1.22**	0.32	0.79
Soil saturated water content [mm]	**0.04**	**1.05**	**0.005**	**2.90**	0.23	0.78
Saturated hydraulic conductivity [mm/hr]	0.28	0.62	0.11	**1.11**	**0.03**	**1.75**
Soil clay percentage	0.67	0.23	**0.03**	**1.55**	0.09	**1.06**
Soil sand percentage	0.06	**1.50**	**0.01**	**2.36**	0.11	**1.04**
b) Flow duration curve and daily flow						
Average flow variability (Concavity index)	**0**	**2.64**	**0.03**	**1.77**	0.81	0.15
High flow variability (FDC slope, 1%-10%)	**0.0001**	**2.19**	**0.03**	**1.49**	0.61	0.32
Mean daily Q/median daily Q	**0.04**	0.98	**0.05**	0.83	0.93	0.06
CV of daily Q	**0**	**1.70**	**0.005**	**1.41**	0.85	0.12
CV in number of annual flow direction reversals	**0.03**	**1.06**	0.07	0.91	0.99	0.02
CV in rate of Q change on increasing Q days	0.91	0.04	**0.05**	**1.24**	0.14	0.53
Log(Median dQ for consecutive increasing Q days)	0.07	**1.06**	0.88	0.12	**0.03**	**1.29**
c) Baseflow						
Baseflow fraction	**0.002**	**2.15**	**0.04**	**1.46**	0.39	0.59
Variance in baseflow based on digital filter	**0.0008**	**1.23**	**0.005**	**1.00**	0.69	0.23
CV of Julian date of annual minimum	**0.007**	**1.66**	**0.02**	**1.52**	0.99	0.04

in contrast, have slower flow through the soil, which leads to a larger low-flow decrease and smaller high-flow decline as the restored land cover can transpire water that has a longer soil residence time. Many other characteristics were tested, including watershed shape, elevation and slope, climate, and pre-restoration land cover, but none of these showed a significant difference among the clusters.

3.2.2. Hydrologic behavior under pre-restoration land-cover shows strong relationships to restored watershed response

In addition to the watershed characteristics, we also analyzed pre-restoration streamflow index values to determine which indices might correlate with the graph-based clusters and thus provide useful information about the similarities and differences in watershed behavior across these clusters.

Fig. 4b and c show the pre-restoration streamflow index values, arranged by cluster, and Table 4 the results of permutation tests and effect sizes across the clusters. There are significant differences between clusters in the hydrologic response as measured by the flow duration curve ("duration curve"), which shows how often flow exceeds any given value and thus serves as an integrated measure of the likelihood of experiencing flow conditions in the long-term. Here we represent the duration curve by the *average flow variability* (technically the concavity index, which is the slope of the line connecting the 33% and 66% exceedance daily flows) and the slope of the duration curve in the range from 1% to 10% exceedance (*high-flow duration index)*. A high value of *average flow variability* means that there is large mid-range flow variability of the duration curve. A duration curve with a large *high-flow duration index* has a smaller number of extremely large events and fewer moderately large events.

We found that *average flow variability* in the low-flow intensifying cluster is significantly smaller than in the high-flow intensifying cluster (Fig. 4b, Table 4b). There are 11,440 possible combinations of the 16 *average flow variability* samples comprised by the seven watersheds in the low-flow intensifying cluster and the nine watersheds in the high-flow intensifying cluster. We found a perfect sorting with the high-flow intensifying cluster having the largest values of *average flow variability* and the low-flow intensifying cluster having the smallest values, and an effect size of 2.6 between the two. This perfect sorting is one of the two most extreme cases of the 11,440 combinations. The *average flow variability* value does not have the same perfect sorting between the low-flow intensifying and high-flow enhancing clusters, but the difference is still significant ($\alpha = 0.05$) with effect size 1.8.

Regarding the *high-flow duration index*, there were extreme differences between the low-flow intensifying and high-flow intensifying clusters, again with perfect sorting. The effect size for the *high-flow duration index* is slightly smaller than that for the *average flow variability*, but it is still above 2. The low-flow intensifying cluster's steeper slope indicates that the streams are flashier, as there is a larger range between the few extreme flows and the moderately large events. The high-flow intensifying cluster's smaller slope indicates more consistent values between extreme and more common high flows, or less flashy behavior. The low-flow intensifying and high-flow enhancing clusters also show a significant difference ($\alpha = 0.05$) in this *high-flow duration index* and have an effect size of 1.5, but lack perfect sorting.

Further results also suggests that the low-flow intensifying cluster sites are predominately governed by a few very large flows and have more constant low-flows during the rest of the year when compared to the high-flow intensifying and high-flow enhancing cluster sites. Specifically, we found that *mean daily flow/median daily flow* index values have a structure similar to that of the *high-flow duration* index values, with the low-flow intensifying cluster sites having large values and the high-flow intensifying and high-flow enhancing cluster sites having smaller values. The difference is significant ($\alpha = 0.05$), but the effect size is only 1.

A few additional indices that describe flow variability show distinctions between the clusters, though not as strongly as duration curve indices, which describe long-term integrated watershed response. These daily flow variability indices are also shown in Fig. 4b and Table 4b. The *coefficient of variation (CV) of daily flow* has perfect sorting between the low-flow intensifying and high-flow intensifying clusters, with an effect size of 1.7. The low-flow intensifying cluster watersheds have a larger CV. Only one site in the high-flow enhancing cluster has a *CV of daily flow* value larger than any site in the low-flow intensifying cluster. The low-flow intensifying cluster sites also have a more *variable number*

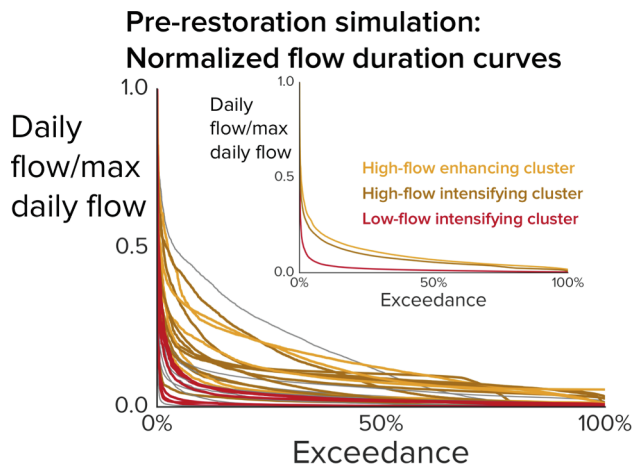

Fig. 5. Pre-restoration flow duration curves for watersheds in the low-flow intensifying cluster tend to be more concave (steeper near 0%, and less steep across larger exceedances) than those from the high-flow enhancing and high-flow intensifying clusters. The inset panel shows the median for each cluster. Lines are colored as in the rest of the paper, with the remaining watersheds shown as thin grey lines.

of flow reversals across years than the high-flow intensifying cluster sites, with effect size 1, though they are not distinguishable from the high-flow enhancing cluster sites due to one highly variable site in the latter cluster. The *CV in the rate of flow change during periods of increasing flow* is larger in the low-flow intensifying cluster watersheds than the high-flow intensifying cluster watersheds. Finally, the *median flow change for consecutive days of increasing flow* is significantly larger in the high-flow enhancing cluster than the high-flow intensifying cluster. This suggests that discharge increases more quickly in the watersheds in the high-flow enhancing cluster than those in the high-flow intensifying cluster, consistent with the former cluster's larger hydraulic conductivity.

The low-flow intensifying cluster has smaller and relatively more variable pre-restoration low-flow than the high-flow enhancing or high-flow intensifying clusters, as shown in Fig. 4c and Table 4c. The *baseflow fraction* index value is larger in both the high-flow intensifying and the high-flow enhancing clusters than in the low-flow intensifying cluster ($\alpha = 0.05$), with effect sizes 2.2 and 1.5, respectively. The high-flow intensifying cluster only has one out of nine values that are lower than any in the low-flow intensifying cluster. Associated with its lower *baseflow fraction*, the low-flow intensifying cluster has a larger *variance in the daily baseflow* estimated using a digital filter (Arnold and Allen, 1999) and a larger *CV in the date of the annual minimum flow*. Supplement SI-8 shows that there is a similar difference across clusters in a number of high flow index values calculated for the pre-restoration simulation.

Other indices did not show significant differences between clusters. These similarities may also be important for interpreting the similarities and differences in watershed response to land-cover change, and are discussed in Supplement SI-9. These indices include the *average flow change during periods of rising flow* and the *slope of the duration curve from 75% to 99% exceedance*.

4. Discussion

4.1. Indicators of watershed response

We employed 29 hydrologic site models to investigate the change in a variety of streamflow indices after simulated land-cover restoration. These results provide useful information about the types of watersheds where watershed service investments in land-cover restoration might provide valuable water resource improvements, and where they might instead disrupt local flow benefits. In particular, sites in the low-flow intensifying cluster display the behavior desired from watershed service investment for land-cover restoration: spreading stream discharge over time even if the penalty is a loss of annual water yield. Sites in the high-flow intensifying cluster instead see a decline in water yield, with a particularly notable decline during low-flow periods. Sites in the high-flow enhancing cluster show the same basic responses as those in the high-flow intensifying cluster, but less so, to the point that declines in water yield and particularly low-flow might not be noticed.

Our analysis is distinct from the spatial targeting of land-cover change inside a watershed, which has been the focus of other work (Guswa et al., 2014; Jackson et al., 2013; Vogl et al., 2016a,b). Instead we investigate the general conditions under which land-cover change might provide tangible watershed services.

4.1.1. Watershed characteristics

Soil depth, saturated water content, and hydraulic conductivity have significantly different values across the delineated clusters. This suggests that watersheds with relatively thin soils that cannot hold significant amounts of water may be more likely to transfer water from high flow to low-flow regimes upon land-cover restoration. This outcome could be beneficial even though it is likely to also produce a reduction in total water yield. Among sites with thicker soils and higher saturated water contents, those sites with relatively high hydraulic conductivities may be likely to see smaller reductions in low flow and smaller relative increases in high flow as the water can flow quickly through the soil profile and thus is affected less by vegetation. In contrast, similar sites but having low hydraulic conductivity soils may be likely to see a decrease in low flow as well as a relative rise in high flow, representing a significant water resource cost to land-cover restoration.

4.1.2. Flow duration curve:

The *average flow variability* and the *high-flow duration* indices based on the flow duration curve for the pre-restoration case also showed significant differences across clusters. This led us to further investigate the features of the flow duration curves for sites in each cluster. In Fig. 5 we plot the 29 duration curves where flow (y-axis) is normalized by the maximum flow for each site. Lines are color-coded by the clusters identified in Fig. 2. The inset panel provides a simple summary by displaying the median curves for all sites in each cluster. There are 8 sites that are not in the three primary clusters, which are colored light grey.

There is a clear difference in the nature of the duration curves for each cluster. The watersheds in the low-flow intensifying cluster tend to have much more concave *duration curves* than those in the high-flow intensifying or high-flow enhancing clusters, as seen in Fig. 5. This is consistent with the observations above, wherein the watersheds with more concave normalized pre-restoration *duration curves* shift more water from the high-flow period to low-flow periods after restoration. In contrast, the watersheds with less concave normalized pre-restoration *duration curves* transfer water to low-flow periods efficiently before restoration. After restoration, these watersheds experience increased evapotranspiration and consequently provide less discharge during low-flow periods. Overall, those watersheds with stronger variability in flow are more likely to have restoration of pre-development land cover increase low flows, while those with a smoother distribution of flows are not.

4.2. Field studies

Despite significant effort by many hydrologists, a clear understanding of the factors that drive the hydrologic response to land-cover change remains elusive. Field studies remain key constraints on such understanding, and our results are consistent with a number of field studies. However, many field studies have focused on land-cover changes that are not directly applicable to watershed service

investment for several reasons: 1) many investigate annual water yield, which is not the most relevant measure for most water resource considerations; 2) many focus on small-scale experimental catchments of 1–1000 ha with extreme land-cover change, often approaching 100%, to achieve statistical power, but which are not necessarily realistic for watershed investments that change smaller fractions of larger watersheds with areas in the 100s-1000s km^2; 3) others address differences in hydrologic response with land-cover across catchments, but do not provide clear guidance for the response to change; and 4) many studies have a strong bias towards temperate and cold climates where most data are available (Bosch and Hewlett, 1982; Brown et al., 2005, 2013; Carrillo et al., 2011; Hamel et al., 2017; Ogden et al., 2013; Price et al., 2011; Sahin and Hall, 1996; Sawicz et al., 2011; Scott and Lesch, 1997; Stednick, 1996; Wagener et al., 2007; Whitehead and Robinson, 1993; Yadav et al., 2007).

The broad paired catchment literature is consistent with our results, showing a reduction in annual water yield under afforestation (Brown et al., 2005; Farley et al., 2005; Filoso et al., 2017). Previous work has suggested that changes in water yield cannot be detected experimentally with changes of $<20\%$ of the watershed because of the natural variability (Bosch and Hewlett, 1982; Stednick, 1996). However, Brown et al. (2005) suggest it is still possible to predict the response of smaller land-cover changes. Such prediction is consistent with the small reductions in water yield for almost all sites, and in all three clusters, that are apparent because the watershed simulations hold all conditions identical except for land cover.

Studies of low-flow changes due to land-cover change find more variability than those of annual water yield, consistent with our results. Paired catchment studies have found that afforestation decreases low flow (Farley et al., 2005), particularly that forested catchments had 90% exceedance flows that were 10–90% lower than the 90% exceedance flows of non-forested catchments (Brown et al., 2013; Scott and Lesch, 1997). This result is similar to that from our high-flow intensifying and high-flow enhancing clusters, though stronger because they often represent the watershed response to much larger land-cover change than the 10% we simulated. Other work has found that forested watersheds had larger low-flows from 1 to 50% (Ogden et al., 2013) and 6–24% (Price et al., 2011), which could be consistent with our low-flow intensifying cluster. Note that these were not, however, experimental land-cover changes, but rather comparisons across paired watersheds with different amounts of forest from previous disturbance. Price et al. (2011) tested the effect of the percent area of colluvium on baseflow in the presence of different amounts of forest cover, and found that increased colluvium and forest cover were positively related to increases in baseflow metrics. If we assume that increased colluvium corresponds with decreased soil depth, this is consistent with our findings that sites with smaller soil depth were in the low-flow intensifying cluster. The variable low-flow effects of land cover are consistent with the distinct, and differing-direction responses seen in the watersheds in the low-flow intensifying versus high-flow intensifying clusters. However, it is not possible to draw robust links between these field studies and our models because of the lack of data on the field study conditions. None of the sites in this study have significant fog capture, so we do not see the potential increased low flow because of increased water input from forests in these uncommon circumstances (Bruijnzeel et al., 2011; Ellison et al., 2012).

4.3. Analysis caveats

As a moderate-complexity model, SWAT includes structural limitations that may affect the results of this work. Most importantly, SWAT's hydrologic response units (HRUs) are not fully spatially connected. Within subbasins the HRUs are spatially discontiguous units that connect to the stream, and subbasins are connected through the stream network. The lack of connectivity may be particularly important for land-cover change in riparian zones, where tree removal may have large effects on streamflow (Everson et al., 2007; Scott, 1999; Scott et al., 2004).

Our analysis assumed a uniform land-cover restoration across the watershed with shifts from a single land-cover to another. Sensitivity analyses assessed the effects of both localized restoration and more complex land-cover shifts, and showed that the differences were small (Dennedy-Frank and Gorelick, 2019, Appendices A.7 & A.8). However, the lack of full spatial connectivity does mean that riparian zones, in particular, may not be appropriately sensitive to these changes. There was no relationship between the streamflow changes after restoration and either the size or number of either HRUs or subbasins across sites.

5. Conclusions

We investigated the similarities and differences of *streamflow changes* after the restoration of pre-development land cover in 29 sites around the world. The sites represent a novel database of previously peer-reviewed hydrologic models simulated using SWAT, a tool used extensively by the land and water management communities. We find three primary clusters sharing common hydrologic responses to restoration: 1) low-flow intensifying cluster, 2) high-flow intensifying cluster, and 3) high-flow enhancing cluster.

Under restoration, the low-flow intensifying cluster sites show enhanced discharge during low flow periods and a decline in discharge during high-flow periods. Shifting discharge from high-flow to low-flow periods is typically a desired outcome for watershed services. In contrast, the high-flow intensifying cluster is characterized by sites in which there is a decrease in discharge during low-flow periods and an increased variability of discharge in the near-average flow regime. Similarly the high-flow enhancing cluster sites exhibit a decrease in discharge during the low-flow periods, though of smaller magnitude than the high-flow intensifying cluster.

Among watershed characteristics, soil depth and saturated soil water content show statistically significant differences across these clusters. Our results suggest that these soil properties may be a primary driver in watershed response to land-cover restoration. No non-soil watershed characteristics showed a statistically significant difference across the three clusters. In sites where the soils are relatively thin, restoration results in watershed responses that generate a more even temporal distribution of flow, which is often desired. In contrast, where soils are thick and hydraulic conductivity is low there may instead be a loss of watershed service because of the increased opportunity for evapotranspiration. Soil depth and hydraulic properties serve as aggregate indicators of effects such as land cover, climate and source material, and may serve as a particularly valuable indicator of coevolution state (Troch et al., 2015).

We show that several flow indices, particularly those describing attributes of the flow duration curve, have statistically significant differences across clusters. The *average flow variability* index and the *high-flow duration* index are most important. The flow duration curves for sites in the low-flow intensifying cluster are generally more concave than those in the high-flow intensifying and high-flow enhancing clusters, which indicates that under land-cover restoration the water in the high-flow periods of the low-flow intensifying cluster sites is shifted to low flow periods. In the high-flow intensifying and high-flow enhancing clusters, land-cover restoration does not transfer additional water to low flow, instead increasing water loss to evapotranspiration. These results provide a key step in the systematic understanding of how land-cover shifts change streamflow and watershed services, and inform better prediction of sites where restoration will add value as natural capital.

CRediT authorship contribution statement

P. James Dennedy-Frank: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Resources, Data curation,

Writing - original draft, Writing - review & editing, Visualization, Project administration. **Steven M. Gorelick:** Conceptualization, Methodology, Resources, Writing - review & editing, Visualization, Supervision, Funding acquisition.

Declaration of Competing Interest

The 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.

Acknowledgments

We gratefully acknowledge those who were willing to share their site models: Hau Zhang, Khoi Dao, Dulce Rodrigues, Donizete Pereira, Samira Akhavan, Hiro Somura, Danilo Fukunaga, Saeed Vaghefi, A. Pouyan Nejadhashemi, Bikesh Shrestha, Tram Lo, Emmanuel Obuobie, Pawel Marcinkowski, Mikolaj Piniewski, Johanna Richards, Matjaz Glavan, Milad Jajarmizadeh, Rewati Niraula, Sarah Praskievicz, Rui Jiang, Diana Pascual, Katrin Bieger, Partha Saha, and Deepthi Rajsekhar, along with all their collaborators who agreed to share the site model.

We especially thank Darren Ficklin, who shared several site models in the western US, Karim Abbaspour who assisted us in collecting several site models from Iran, and Aijin Zhang who shared two sites in China.

This work was supported in part by the National Science Foundation under grant ICER/EAR-1829999 to Stanford University as part of the Belmont Forum – Sustainable Urbanisation Global Initiative (SUGI)/ Food-Water-Energy Nexus theme. Any opinions, findings, conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. The authors also thank Stanford University's Department of Earth System Science for financial support.

Model input files are available upon request due to their provenance with many other researchers.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jhydrol.2020.125121.

References

Akhavan, S., Abedi-Koupai, J., Mousavi, S.-F., Afyuni, M., Eslamian, S.-S., Abbaspour, K.C., 2010. Application of SWAT model to investigate nitrate leaching in Hamadan-Bahar Watershed, Iran. Agric. Ecosyst. Environ. 139 (4), 675–688. https://doi.org/10.1016/j.agee.2010.10.015.

Arnold, J.G., Allen, P.M., 1999. Automated Methods for Estimating Baseflow and Ground Water Recharge from Streamflow Records. JAWRA J. Am. Water Resour. Assoc. 35 (2), 411–424.

Arnold, J.G., Srinivasan, R., Muttiah, R.S., Williams, J.R., 1998. Large area hydrologic modeling and assessment part I: model development. J. Am. Water Resour. Assoc. 34 (1), 73–89.

Arnold, J.G., Gassman, P.W., White, M.J., 2010. New Developments in the SWAT Ecohydrology Model. In: 21st Century Watershed Technology: Improving Water Quality and Environment Conference Proceedings. Universidad EARTH, Costa Rica: American Society of Agricultural and Biological Engineers. https://doi.org/10.13031/2013.29393.

Bennett, G., Ruef, F., 2016. Alliances for Green Infrastructure: State of Watershed Investment 2016 (Ecosystem Marketplace). Forest Trends, Washington, DC, pp. 76.

Bosch, J.M., Hewlett, J.D., 1982. A review of catchment experiments to determine the effect of vegetation changes on water yield and evapotranspiration. J. Hydrol. 55 (1/4), 3–23.

Brauman, K.A., 2015. Hydrologic ecosystem services: linking ecohydrologic processes to human well-being in water research and watershed management. Wiley Interdiscipl. Rev. Water 2 (4), 345–358. https://doi.org/10.1002/wat2.1081.

Bremer, L.L., Auerbach, D.A., Goldstein, J.H., Vogl, A.L., Shemie, D., Kroeger, T., Nelson, J.L., Benítez, S.P., Calvache, A., Guimarães, J., Herron, C., Higgins, J., Klemz, C., León, J., Lozano, J.S., Moreno, P.H., Nuñez, F., Veiga, F., Tiepolo, G., 2016. One size does not fit all: Natural infrastructure investments within the Latin American Water Funds Partnership. Ecosyst. Serv. 17, 217–236. https://doi.org/10.1016/j.ecoser.2015.12.006.

de Bressiani, D.S, Gassman, P.W., Fernandes, J.G., Garbossa, L.H.P., Srinivasan, R., Bonumá, N.B., Mendiondo, E.M., 2015. Review of soil and water assessment tool (SWAT) applications in Brazil: challenges and prospects. Int. J. Agric. Biol. Eng. 8 (3), 9–35. https://doi.org/10.3965/ijabe.v8i3.1765.

Brown, A.E., Zhang, L., McMahon, T.A., Western, A.W., Vertessy, R.A., 2005. A review of paired catchment studies for determining changes in water yield resulting from alterations in vegetation. J. Hydrol. 310 (1–4), 28–61. https://doi.org/10.1016/j.jhydrol.2004.12.010.

Brown, A.E., Western, A.W., McMahon, T.A., Zhang, L., 2013. Impact of forest cover changes on annual streamflow and flow duration curves. J. Hydrol. 483, 39–50. https://doi.org/10.1016/j.jhydrol.2012.12.031.

Bruijnzeel, L.A., Mulligan, M., Scatena, F.N., 2011. Hydrometeorology of tropical montane cloud forests: emerging patterns. Hydrol. Process. 25 (3), 465–498. https://doi.org/10.1002/hyp.7974.

Carrillo, G., Troch, P.A., Sivapalan, M., Wagener, T., Harman, C., Sawicz, K., 2011. Catchment classification: hydrological analysis of catchment behavior through process-based modeling along a climate gradient. Hydrol. Earth Syst. Sci. 15 (11), 3411–3430. https://doi.org/10.5194/hess-15-3411-2011.

Chaplin-Kramer, R., Sharp, R.P., Weil, C., Bennett, E.M., Pascual, U., Arkema, K.K., Brauman, K.A., Bryant, B.P., Guerry, A.D., Haddad, N.M., Hamann, M., Hamel, P., Johnson, J.A., Mandle, L., Pereira, H.M., Polasky, S., Ruckelshaus, M., Shaw, M.R., Silver, J.M., Vogl, A.L., Daily, G.C., 2019. Global modeling of nature's contributions to people. Science 366 (6462), 255–258. https://doi.org/10.1126/science.aaw3372.

Clausen, B., Biggs, B.J.F., 2000. Flow variables for ecological studies in temperate streams: groupings based on covariance. J. Hydrol. 237 (3–4), 184–197. https://doi.org/10.1016/S0022-1694(00)00306-1.

Condon, L.E., Maxwell, R.M., 2017. Systematic shifts in Budyko relationships caused by groundwater storage changes. Hydrol. Earth Syst. Sci. 21 (2), 1117–1135. https://doi.org/10.5194/hess-21-1117-2017.

Daily, G.C., Polasky, S., Goldstein, J., Kareiva, P.M., Mooney, H.A., Pejchar, L., Ricketts, T.H., Salzman, J., Shallenberger, R., 2009. Ecosystem services in decision making: time to deliver. Front. Ecol. Environ. 7 (1), 21–28. https://doi.org/10.1890/080025.

Damania, R., Desbureaux, S., Hyland, M., Islam, A., Moore, S., Rodella, A.-S., Russ, J., & Zaveri E. (2017). Uncharted Waters. World Bank, Washington, DC. Retrieved from doi:10.1596/978-1- 4648-1179-1.

Dennedy-Frank, P.J., Gorelick, S.M., 2019. Insights from watershed simulations around the world: Watershed service-based restoration does not significantly enhance streamflow. Global Environ. Change 58, 101938. https://doi.org/10.1016/j.gloenvcha.2019.101938.

Dennedy-Frank, P.J., Muenich, R.L., Chaubey, I., Ziv, G., 2016. Comparing two tools for ecosystem service assessments regarding water resources decisions. J. Environ. Manage. 177, 331–340. https://doi.org/10.1016/j.jenvman.2016.03.012.

Döll, P., Douville, H., Güntner, A., Schmied, H.M., Wada, Y., 2016. Modelling freshwater resources at the global scale: challenges and prospects. Surv. Geophys. 37 (2), 195–221. https://doi.org/10.1007/s10712-015-9343-1.

Duan, Q., Schaake, J., Andréassian, V., Franks, S., Goteti, G., Gupta, H.V., Gusev, Y.M., Habets, F., Hall, A., Hay, L., Hogue, T., Huang, M., Leavesley, G., Liang, X., Nasonova, O.N., Noilhan, J., Oudin, L., Sorooshian, S., Wagener, T., Wood, E.F., 2006. Model Parameter estimation experiment (MOPEX): an overview of science strategy and major results from the second and third workshops. J. Hydrol. 320 (1), 3–17. https://doi.org/10.1016/j.jhydrol.2005.07.031.

Ellison, D., Futter, N.M., Bishop, K., 2012. On the forest cover-water yield debate: from demand- to supply-side thinking. Glob. Change Biol. 18 (3), 806–820. https://doi.org/10.1111/j.1365-2486.2011.02589.x.

Everson, C. S., South Africa, & Water Research Commission. (2007). Effective management of the riparian zone vegetation to significantly reduce the cost of catchment management and enable greater productivity of land resources. Gezina, South Africa: Water Research Commission.

Farley, K.A., Jobbagy, E.G., Jackson, R.B., 2005. Effects of afforestation on water yield: a global synthesis with implications for policy. Glob. Change Biol. 11 (10), 1565–1576. https://doi.org/10.1111/j.1365-2486.2005.01011.x.

Ficklin, D.L., Stewart, I.T., Maurer, E.P., 2013. Effects of climate change on stream temperature, dissolved oxygen, and sediment concentration in the Sierra Nevada in California. Water Resour. Res. 49 (5), 2765–2782. https://doi.org/10.1002/wrcr.20248.

Filoso, S., Bezerra, M.O., Weiss, K.C.B., Palmer, M.A., 2017. Impacts of forest restoration on water yield: A systematic review. PLoS ONE 12 (8), e0183210. https://doi.org/10.1371/journal.pone.0183210.

Fisher, J.R.B., Acosta, E.A., Dennedy-Frank, P.J., Kroeger, T., Boucher, T.M., 2017. Impact of satellite imagery spatial resolution on land use classification accuracy and modeled water quality. Remote Sens. Ecol. Conserv. https://doi.org/10.1002/rse2.61.

Francesconi, W., Srinivasan, R., Pérez-Miñana, E., Willcock, S.P., Quintero, M., 2016. Using the soil and water assessment tool (SWAT) to model ecosystem services: a systematic review. J. Hydrol. 535, 625–636. https://doi.org/10.1016/j.jhydrol.2016.01.034.

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

Gassman, P.W., Reyes, M.R., Green, C.H., Arnold, J.G., 2007. The Soil and Water Assessment Tool: Historical development, applications, and future research directions Invited Review Series. Am. Soc. Agric. Biol. Eng. 50, 1211–1250.

Good, P.I., 2005. Permutation, Parametric and Bootstrap tests of Hypotheses, 3rd ed. Springer, New York.

Guswa, A.J., Brauman, K.A., Brown, C., Hamel, P., Keeler, B.L., Sayre, S.S., 2014.

Ecosystem services: challenges and opportunities for hydrologic modeling to support decision making. Water Resour. Res. 50 (5), 4535–4544. https://doi.org/10.1002/2014WR015497.
Hagberg, A., Schult, D., & Renieris, M. (2010). PyGraphViz. Retrieved from http://pygraphviz.github.io/.
Haines, A.T., Findlayson, B.L., McMahon, T.A., 1988. A global classification of river regimes. Appl. Geogr. 8 (4), 255–272.
Hamel, P., Guswa, A.J., Sahl, J., Zhang, L., 2017. Predicting dry-season flows with a monthly rainfall–runoff model: performance for gauged and ungauged catchments. Hydrol. Process. 31 (22), 3844–3858. https://doi.org/10.1002/hyp.11298.
Hartuv, E., Schmitt, A.O., Lange, J., Meier-Ewert, S., Lehrach, H., Shamir, R., 2000. An algorithm for clustering cDNA fingerprints. Genomics 66 (3), 249–256. https://doi.org/10.1006/geno.2000.6187.
Hortness, J. E. (2006). Estimating low-flow frequency statistics for unregulated streams in Idaho (Scientific Investigations Report No. 2006–5035). US Department of the Interior, US Geological Survey. Retrieved from http://pubs.usgs.gov/sir/2006/5035/.
Hughes, J.M.R., James, B., 1989. A hydrological regionalization of streams in Victoria, Australia, with implications for stream Ecology. Mar. Freshw. Res. 40 (3), 303–326. https://doi.org/10.1071/mf9890303.
Jackson, B., Pagella, T., Sinclair, F., Orellana, B., Henshaw, A., Reynolds, B., McIntyre, N., Wheater, H., Eycott, A., 2013. Polyscape: A GIS mapping framework providing efficient and spatially explicit landscape-scale valuation of multiple ecosystem services. Landsc. Urban Plann. 112, 74–88. https://doi.org/10.1016/j.landurbplan.2012.12.014.
Kamada, T., Kawai, S., 1989. An algorithm for drawing general undirected graphs. Inform. Process. Lett. 31 (1), 7–15. https://doi.org/10.1016/0020-0190(89)90102-6.
Kennen, J. G., Henriksen, J. A., & Nieswand, S. P. (2007). Development of the Hydroecological Integrity Assessment Process for Determining Environmental Flows for New Jersey Streams (Scientific Investigations Report No. 2007–5206) (p. 65). Reston, Virginia: US Department of the Interior, US Geological Survey. Retrieved from http://pubs.usgs.gov/sir/2007/5206/.
Kepner, W.G., Ramsey, M.M., Brown, E.S., Jarchow, M.E., Dickinson, K.J.M., Mark, A.F., 2012. Hydrologic futures: using scenario analysis to evaluate impacts of forecasted land use change on hydrologic services. Ecosphere 3 (7), 1–25. https://doi.org/10.1890/ES11-00367.1.
Khoi, D.N., Suetsugi, T., 2014a. Impact of climate and land-use changes on hydrological processes and sediment yield—a case study of the Be River catchment, Vietnam. Hydrol. Sci. J. 59 (5), 1095–1108. https://doi.org/10.1080/02626667.2013.819433.
Khoi, D.N., Suetsugi, T., 2014b. The responses of hydrological processes and sediment yield to land-use and climate change in the Be River Catchment Vietnam. Hydrol. Process. 28 (3), 640–652. https://doi.org/10.1002/hyp.9620.
Kim, I., Arnhold, S., Ahn, S., Le, Q.B., Kim, S.J., Park, S.J., Koellner, T., 2017. Land use change and ecosystem services in mountainous watersheds: predicting the consequences of environmental policies with cellular automata and hydrological modeling. Environ. Modell. Software. https://doi.org/10.1016/j.envsoft.2017.06.018.
Kollet, S.J., Maxwell, R.M., 2006. Integrated surface–groundwater flow modeling: a free-surface overland flow boundary condition in a parallel groundwater flow model. Adv. Water Resour. 29 (7), 945–958. https://doi.org/10.1016/j.advwatres.2005.08.006.
Kroeger, T., Klemz, C., Shemie, D., Boucher, T., Fisher, J. R. B., Acosta, E., Dennedy-Frank, P. J., Cavassani, A. T., Garbossa, L. H., Blainski, E., Santos R. C., Petry, P., Gilberti, S., & Dacol, K. (2017). Assessing the Return On Investment in Watershed Conservation: Best Practices Approach and Case Study for the Rio Camboriú PWS Program, Santa Catarina, Brazil. Arlington, VA: The Nature Conservancy.
Kroeger, T., Klemz, C., Boucher, T., Fisher, J.R.B., Acosta, E., Cavassani, A.T., Dennedy-Frank, P.J., Garbossa, L., Blainski, E., Santos, R.C., Gilberti, S., Petry, P., Shemie, D., Dacol, K., 2019. Returns on investment in watershed conservation: application of a best practices analytical framework to the Rio Camboriú Water Producer program, Santa Catarina, Brazil. Sci. Total Environ. 657, 1368–1381. https://doi.org/10.1016/j.scitotenv.2018.12.116.
Laaha, G., Demuth, S., Hisdal, H., Kroll, C. N., Van Lanen, H. A. J., Nester, T., Rogger, M., Sauquet, E., Tallaksen, L. M., Woods, R. A., Young, A., Blöschl, G., Sivapalan, M., Wagener, T., Viglione, A., & Savenjie, H. H. G. (2013). Prediction of low flows in ungauged basins. In G. Blöschl, M. Sivapalan, T. Wagener, A. Viglione, & H. H. G. Savenije (Eds.), Runoff Prediction in Ungauged Basins: Synthesis across Processes, Places, and Scales. Cambridge, UK: Cambridge University Press.
LACC, & TNC. (2013). Natural Infrastructure: An Opportunity for Water Security in 25 Cities in Latin America (p. 16). Mexico: Latin American Conservation Council, with TNC. Retrieved from http://waterfunds.org/sites/default/files/booklet_tnc_letter_ingles_non_methodology_baja.pdf.
Logsdon, R.A., Chaubey, I., 2013. A quantitative approach to evaluating ecosystem services. Ecol. Model. 257, 57–65. https://doi.org/10.1016/j.ecolmodel.2013.02.009.
Martin, G. R., Fowler, K. K., & Arihood, L. D. (2016). Estimating Selected Low-Flow Frequency Statistics and Harmonic-Mean Flows for Ungaged, Unregulated Streams in Indiana (Scientific Investigations Report No. 2016–5102). US Department of the Interior, US Geological Survey. Retrieved from http://dx.doi.org/10.3133/sir20165102.
Moffett, K.B., Gorelick, S.M., McLaren, R.G., Sudicky, E.A., 2012. Salt marsh ecohydrological zonation due to heterogeneous vegetation–groundwater–surface water interactions. Water Resour. Res. 48 (2), W02516. https://doi.org/10.1029/2011WR010874.
Moriasi, 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 (3), 885–900.
Naeem, S., Ingram, J.C., Varga, A., Agardy, T., Barten, P., Bennett, G., et al., 2015. Get the science right when paying for nature's services. Science 347 (6227), 1206–1207. https://doi.org/10.1126/science.aaa1403.
Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I — A discussion of principles. J. Hydrol. 10 (3), 282–290. https://doi.org/10.1016/0022-1694(70)90255-6.
Newman, A., Sampson, K., Clark, M. P., Bock, A., Viger, R. J., & Blodgett, D. (2014). A large-sample watershed-scale hydrometeorological dataset for the contiguous USA. (Version doi:10.5065/D6MW2F4D). Boulder, CO: UCAR/NCAR.
Newman, A.J., Clark, M.P., Sampson, K., Wood, A., Hay, L.E., Bock, A., Viger, R.J., Blodgett, D., Brekke, L., Arnold, J.R., Hopson, T., Duan, Q., 2015. Development of a large-sample watershed-scale hydrometeorological data set for the contiguous USA: data set characteristics and assessment of regional variability in hydrologic model performance. Hydrol. Earth Syst. Sci. 19 (1), 209–223. https://doi.org/10.5194/hess-19-209-2015.
Ogden, F.L., Crouch, T.D., Stallard, R.F., Hall, J.S., 2013. Effect of land cover and use on dry season river runoff, runoff efficiency, and peak storm runoff in the seasonal tropics of Central Panama. Water Resour. Res. 49 (12), 8443–8462. https://doi.org/10.1002/2013WR013956.
Olden, J.D., Poff, N.L., 2003. Redundancy and the choice of hydrologic indices for characterizing streamflow regimes. River Res. Appl. 19 (2), 101–121. https://doi.org/10.1002/rra.700.
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É., 2011. Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. 12, 2825–2830.
Peel, M.C., Finlayson, B.L., McMahon, T.A., 2007. Updated world map of the Köppen-Geiger climate classification. Hydrol. Earth Syst. Sci. 11 (5), 1633–1644. https://doi.org/10.5194/hess-11-1633-2007.
Peel, Murray C., Chiew, F. H., Western, A. W., & McMahon, T. A. (2000). Extension of unimpaired monthly streamflow data and regionalisation of parameter values to estimate streamflow in ungauged catchments. Australian Natural Resources Atlas Website.
Peel, Murray C., McMahon, T.A., Finlayson, B.L., 2010. Vegetation impact on mean annual evapotranspiration at a global catchment scale. Water Resour. Res. 46 (9), W09508. https://doi.org/10.1029/2009WR008233.
Piniewski, M., Kardel, I., Giełczewski, M., Marcinkowski, P., Okruszko, T., 2014. Climate change and agricultural development: adapting polish agriculture to reduce future nutrient loads in a coastal watershed. Ambio 43 (5), 644–660. https://doi.org/10.1007/s13280-013-0461-z.
Poff, N., 1996. A hydrogeography of unregulated streams in the United States and an examination of scale-dependence in some hydrological descriptors. Freshw. Biol. 36 (1), 71–79. https://doi.org/10.1046/j.1365-2427.1996.00073.x.
Price, K., Jackson, C.R., Parker, A.J., Reitan, T., Dowd, J., Cyterski, M., 2011. Effects of watershed land use and geomorphology on stream low flows during severe drought conditions in the southern Blue Ridge Mountains, Georgia and North Carolina United States. Water Resour. Res. 47 (2), W02516. https://doi.org/10.1029/2010WR009340.
Puckridge, J.T., Sheldon, F., Walker, K.F., Boulton, A.J., 1998. Flow variability and the ecology of large rivers. Mar. Freshw. Res. 49 (1), 55–72. https://doi.org/10.1071/mf94161.
Ramankutty, N., Foley, J.A., 1999. Estimating historical changes in global land cover: croplands from 1700 to 1992. Global Biogeochem. Cycles 13 (4), 997–1027. https://doi.org/10.1029/1999GB900046.
Richter, B., Baumgartner, J., Wigington, R., Braun, D., 1997. How much water does a river need? Freshw. Biol. 37 (1), 231–249. https://doi.org/10.1046/j.1365-2427.1997.00153.x.
Richter, B.D., Baumgartner, J.V., Powell, J., Braun, D.P., 1996. A method for assessing hydrologic alteration within ecosystems. Conserv. Biol. 10 (4), 1163–1174.
Risley, J. C., Stonewall, A., & Haluska, T. L. (2008). Estimating flow-duration and low-flow frequency statistics for unregulated streams in Oregon (Scientific Investigations Report No. 2008–5126). US Department of the Interior, US Geological Survey. Retrieved from http://pubs.usgs.gov/sir/2008/5126/.
Sahin, V., Hall, M.J., 1996. The effects of afforestation and deforestation on water yields. J. Hydrol. 178 (1), 293–309. https://doi.org/10.1016/0022-1694(95)02825-0.
Sawicz, K., Wagener, T., Sivapalan, M., Troch, P.A., Carrillo, G., 2011. Catchment classification: empirical analysis of hydrologic similarity based on catchment function in the eastern USA. Hydrol. Earth Syst. Sci. 15 (9), 2895–2911. https://doi.org/10.5194/hess-15-2895-2011.
van der Schans, M. (2015). factor_rotation: Python code for for factor rotation. Python. Retrieved from https://github.com/mvds314/factor_rotation.
Scott, David F., 1999. Managing riparian zone vegetation to sustain streamflow: results of paired catchment experiments in South Africa. Can. J. For. Res. 29 (7), 1149–1157. https://doi.org/10.1139/x99-042.
Scott, David F., Lesch, W., 1997. Streamflow responses to afforestation with Eucalyptus grandis and Pinus patula and to felling in the Mokobulaan experimental catchments, South Africa. J. Hydrol. 199 (3), 360–377. https://doi.org/10.1016/S0022-1694(96)03336-7.
Scott, D.F., Bruijnzeel, L. A., Vertessy, R. A., & Calder, I. R. (2004). HYDROLOGY | Impacts of Forest Plantations on Streamflow. In Encyclopedia of Forest Sciences (pp. 367–377). Elsevier. https://doi.org/10.1016/B0-12-145160-7/00272-6.
Smith, T., Marshall, L., McGlynn, B., 2018. Typecasting catchments: Classification, directionality, and the pursuit of universality. Adv. Water Resour. 112, 245–253. https://doi.org/10.1016/j.advwatres.2017.12.020.
Stark, P., Millman, J., Ottoboni, K., van der Walt, S., Brett, M. 2015. permute: Permutation tests and confidence sets for python. Python. Retrieved from https://pypi.org/project/permute/.
Stednick, J.D., 1996. Monitoring the effects of timber harvest on annual water yield. J.

Hydrol. 176 (1), 79–95. https://doi.org/10.1016/0022-1694(95)02780-7.
Strauch, M., Volk, M., 2013. SWAT plant growth modification for improved modeling of perennial vegetation in the tropics. Ecol. Model. 269, 98–112. https://doi.org/10.1016/j.ecolmodel.2013.08.013.
Suliman, A.H.A., Jajarmizadeh, M., Harun, S., Darus, I.Z.M., 2015. Comparison of semi-distributed, GIS-based hydrological models for the prediction of streamflow in a large catchment. Water Resour. Manage. 29 (9), 3095–3110. https://doi.org/10.1007/s11269-015-0984-0.
Texas A&M University; USDA ARS. (2017, March 2). SWAT Literature Database for Peer-Reviewed Journal Articles. Retrieved March 1, 2017, from https://www.card.iastate.edu/swat_articles/index.aspx.
Therrien, R., Sudicky, E.A., 1996. Three-dimensional analysis of variably-saturated flow and solute transport in discretely-fractured porous media. J. Contam. Hydrol. 23 (1), 1–44. https://doi.org/10.1016/0169-7722(95)00088-7.
Troch, P.A., Lahmers, T., Meira, A., Mukherjee, R., Pedersen, J.W., Roy, T., Valdés-Pineda, R., 2015. Catchment coevolution: a useful framework for improving predictions of hydrological change? Water Resour. Res. 51 (7), 4903–4922. https://doi.org/10.1002/2015WR017032.
Van Liew, M.W., Veith, T.L., Bosch, D.D., Arnold, J.G., 2007. Suitability of SWAT for the conservation effects assessment project: comparison on USDA agricultural research service watersheds. J. Hydrol. Eng. 12 (2), 173–189.
Vogl, A.L., Dennedy-Frank, P.J., Wolny, S., Johnson, J.A., Hamel, P., Narain, U., Vaidya, A., 2016a. Managing forest ecosystem services for hydropower production. Environ. Sci. Policy 61, 221–229. https://doi.org/10.1016/j.envsci.2016.04.014.
Vogl, A.L., Bryant, B.P., Hunink, J.E., Wolny, S., Apse, C., Droogers, P., 2016b. Valuing investments in sustainable land management in the Upper Tana River basin, Kenya. J. Environ. Manage. 195 (1), 78–91. https://doi.org/10.1016/j.jenvman.2016.10.013.
Vogl, A.L., Goldstein, J.H., Daily, G.C., Vira, B., Bremer, L., McDonald, R.I., Shemie, D., Tellman, B., Cassin, J., 2017. Mainstreaming investments in watershed services to enhance water security: Barriers and opportunities. Environ. Sci. Policy 75, 19–27. https://doi.org/10.1016/j.envsci.2017.05.007.
Wada, Y., van Beek, L.P.H., Bierkens, M.F.P., 2011. Modelling global water stress of the recent past: on the relative importance of trends in water demand and climate variability. Hydrol. Earth Syst. Sci. 15 (12), 3785–3808. https://doi.org/10.5194/hess-15-3785-2011.
Wagener, T., Sivapalan, M., Troch, P., Woods, R., 2007. Catchment classification and hydrologic similarity. Geogr. Compass 1 (4), 901–931. https://doi.org/10.1111/j.1749-8198.2007.00039.x.
Welkowitz, J., Cohen, B. H., & Lea, R. B. (2012). Introductory statistics for the behavioral sciences (7th ed.). Hoboken, NJ: Wiley.
White, E.D., Easton, Z.M., Fuka, D.R., Collick, A.S., Adgo, E., McCartney, M., Awulachew, S.B., Selassie, Y.G., Steenhuis, T.S., 2011. Development and application of a physically based landscape water balance in the SWAT model. Hydrol. Process. 25 (6), 915–925. https://doi.org/10.1002/hyp.7876.
Whitehead, P.G., Robinson, M., 1993. Experimental basin studies—an international and historical perspective of forest impacts. J. Hydrol. 145 (3), 217–230. https://doi.org/10.1016/0022-1694(93)90055-E.
Williams, J. R. (1995). The EPIC Model. In Computer Models of Watershed Hydrology (pp. 909–1000). Highlands Ranch, CO: Water Resources Publications.
Wood, P.J., Agnew, M.D., Petts, G.E., 2000. Flow variations and macroinvertebrate community responses in a small groundwater-dominated stream in south-east England. Hydrol. Process. 14 (16–17), 3133–3147. https://doi.org/10.1002/1099-1085(200011/12)14:16/17<3133::AID-HYP138>3.0.CO;2-J.
World Meteorological Organization. (2009). Manual on Low-flow Estimation and Prediction (Operational Hydrology Report No. 50). Geneva, Switzerland.
Yadav, M., Wagener, T., Gupta, H., 2007. Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins. Adv. Water Resour. 30 (8), 1756–1774. https://doi.org/10.1016/j.advwatres.2007.01.005.
Yoon, J.,, 2017. Evaluation of water security in Jordan using a multi-agent hydro-economic model (Doctoral dissertation). Stanford University, Stanford, CA. Retrieved from https://searchworks.stanford.edu.
Zaherpour, J., Gosling, S.N., Mount, N., Schmied, H.M., Veldkamp, T.I.E., Dankers, R., Eisner, S., Gerten, D., Gudmundsson, L., Haddeland, I., Hanaski, N., Kim, H., Leng, G., Liu, J., Masaki, Y., Oki, T., Pokhrel, Y., Satoh, Y., Schewe, J., Wada, Y., 2018. Worldwide evaluation of mean and extreme runoff from six global-scale hydrological models that account for human impacts. Environ. Res. Lett. 13 (6), 065015. https://doi.org/10.1088/1748-9326/aac547.
Zhang, A., Zhang, C., Chu, J., Fu, G., 2015. Human-induced runoff change in northeast china. J. Hydrol. Eng. 20 (5), 04014069. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001078.
Zhang, L., Dawes, W.R., Walker, G.R., 2001. Response of mean annual evapotranspiration to vegetation changes at catchment scale. Water Resour. Res. 37 (3), 701–708. https://doi.org/10.1029/2000WR900325.
Ziegeweid, J. R., Lorenz, D. L., Sanocki, C. A., & Czuba, C. R. (2015). Methods for Estimating Flow-Duration Curve and Low- Flow Frequency Statistics for Ungaged Locations on Small Streams in Minnesota (Scientific Investigations Report No. 2015–5170). US Department of the Interior, US Geological Survey. Retrieved from http://dx.doi.org/10.3133/sir20155170.