Received: 4 March 2017 | Accepted: 21 September 2017
DOI: 10.1002/hyp.11369

WILEY

SI URBAN HYDROLOGY

Hydrologic metrics for status-and-trends monitoring in urban and urbanizing watersheds

Derek B. Booth[1] | Christopher P. Konrad[2]

[1] Bren School of Environmental Science & Management, University of California Santa Barbara, Santa Barbara, CA 93106, USA
[2] United States Geological Survey, 934 Broadway, Tacoma, WA 98402, USA

Correspondence
Derek B. Booth, Bren School of Environmental Science & Management, University of California Santa Barbara, Santa Barbara, CA 93106, USA.
Email: dbooth@bren.ucsb.edu

Funding information
U.S. Army Corps of Engineers, Grant/Award Number: W912HZ-12-2-0016; Washington Department of Ecology, Grant/Award Number: 15-05

Abstract
Local governmental agencies are increasingly undertaking potentially costly “status-and-trends” monitoring to evaluate the effectiveness of stormwater control measures and land-use planning strategies or to satisfy regulatory requirements. Little guidance is presently available for such efforts, and so we have explored the application, interpretation, and temporal limitations of well-established hydrologic metrics of runoff changes from urbanization, making use of an unusually long-duration, high-quality data set from the Pacific Northwest (USA) with direct applicability to urban and urbanizing watersheds. Three metrics previously identified for their utility in identifying hydrologic conditions with biological importance that respond to watershed urbanization—T_{Qmean} (the fraction of time that flows exceed the mean annual discharge), the Richards-Baker Index (characterizing flashiness relative to the mean discharge), and the annual tally of wet-season day-to-day flow reversals (the total number of days that reverse the prior days' increasing or decreasing trend)—are all successful in stratifying watersheds across a range of urbanization, as measured by total contributing area of urban development. All metrics respond with statistical significance to multidecadal trends in urbanization, but none detect trends in watershed-scale urbanization over the course of a single decade. This suggests a minimum period over which dependable trends in hydrologic alteration (or improvement) can be detected with confidence. The metrics also prove less well suited to urbanizing watersheds in a semi-arid climate, with only flow reversals showing a response consistent with prior findings from more humid regions. We also explore the use of stage as a surrogate for discharge in calculating these metrics, recognizing potentially significant agency cost savings in data collection with minimal loss of information. This approach is feasible but cannot be implemented under current data-reporting practices, requiring measurement of water-depth values and preservation of the full precision of the original recorded data. With these caveats, however, hydrologic metrics based on stage should prove as or more useful, at least in the context of status-and-trends monitoring, as those based on subsequent calculations of discharge.

KEYWORDS
hydrologic metrics, impervious area, status-and-trends monitoring, stormwater control measures, urbanization

1 | INTRODUCTION

1.1 | Why urban stream monitoring?

Streamflow gaging of small streams has historically been limited, with most monitoring efforts expended on larger rivers that carry the majority of water from any given region. The value of data from these smaller systems, however, has become more widely appreciated because of their sensitivity to land cover and water usage, their dominance of the total length of stream networks, and their unique ecological roles in retention and processing of nutrients and organic material and in the life cycles of aquatic species (Meyer et al., 2007; Peterson et al., 2001). For stormwater management, continuous daily records of streamflow, in particular, can inform (a) the status and trends of hydrologic alteration in small streams receiving stormwater; (b) the conditions in nearby, ungaged streams; (c) the development of standards or performance targets from an understanding of baseline (unaltered) streamflow; and (d) the effectiveness of management actions intended to reduce stormwater impacts (Konrad & Voss, 2012). In the United States, regulatory requirements for status-and-trends monitoring of receiving waters affected by urban stormwater discharges have begun to be imposed under provisions of the Clean Water Act (e.g., Section S8 of Washington Department of Ecology, 2012), and the increasing investment of states and municipalities in stormwater management continues to raise questions of whether these efforts are having any discernible, beneficial effects on downstream receiving waters (e.g., Ahiablame, Engel, & Chaubey, 2012; Roy et al., 2014).

1.2 | Prior investigations

Land-cover changes resulting from urbanization have been long recognized to alter the hydrology of watersheds and the flow regime of streams, particularly small streams (e.g., Leopold, 1968). However, even when monitoring data have been available, there has been little consensus over the years about the "best" metrics to describe these alterations or even what criteria should be used to identify an ideal metric. The earliest studies tended to focus on the increased magnitude of floods of a particular recurrence interval (Sauer, Thomas, Strickler, & Wilson, 1983), of which the compilation by Hollis (1975), who reported urban-induced peak discharge increases of two to five for a given intensity and duration of rainfall, remains one of the more robust characterizations of this widely recognized phenomenon.

Subsequent work on characterizing urban-altered flow regimes, actively developed in the Pacific Northwest (USA) by King County's Basin Planning Program in the late 1980s (e.g., King County, 1990) and later embraced more broadly (MacRae, 1997; Vogel & Fennessey, 1995), focused on cumulative flow durations over multidecade records of daily discharge. These studies, typically using model simulation of continuous hydrographs, projected increases in flow durations following full watershed urbanization of one to two orders of magnitude over a range of sediment-transporting flows, with the greatest proportional increases for the largest discharges.

These findings are not unique to the Pacific Northwest, although they have a particularly long history of investigation there. Elsewhere, Hawley and Bledsoe (2011) explored the influence of urbanization on peak flows and flow durations in a distinctly different climatological setting, that of semi-arid southern California (southwest USA), making use of 43 long-term United States Geological Survey (USGS) gage records of which six had watershed imperviousness greater than 10%. They found the greatest influence of urbanization on peak flows at moderate flood discharges (return periods <5-year event), consistent with prior compilations but with even greater multipliers than previously reported for these relatively frequent events. Also distinct from the findings for humid regions, the influence of urbanization appeared to increase flow durations but proportionately *less* for progressively higher discharges. Hawley and Bledsoe (2011) speculated that the hydrology of semi-arid streams may be more sensitive to the effects of urbanization than those of more commonly studied humid regions; Booth, Roy, Smith, and Kapps (2016) also noted that in semi-arid regions with infrequent but large storms, flow recurrences will be more affected by the magnitude of decadal-scale storms than by the amount of urban land cover. However, neither publication explored alternative metrics of urban-induced hydrologic alteration.

Other indictors of hydrologic change resulting from watershed urbanization, particularly the increase in effective impervious area and the decrease in forest cover, were also explored during the 1980s and 1990s in the Pacific Northwest, including the frequency at which discharge exceed a chosen threshold of presumed streambed disturbance or significant streambank erosion (Booth, 1991). This metric was identified under the assumption that it could highlight changes of particular importance to biota, particularly benthic macroinvertebrates that depend on a relatively stable substrate (Hawley, Wooten, MacMannis, & Fet, 2016).

More recently, metrics seeking to measure various streamflow characteristics of ecological significance have been presented. Their application has emphasized the effects of dam regulation on rivers having long periods of record spanning preregulation and postregulation, or the assessment of hydrologic differences among streams related to both physiographic setting and land/water uses over large regions (e.g., indicators of hydrologic alteration, Kennen, Henriksen, & Nieswand, 2007; Richter, Baumgartner, Powell, & Braun, 2012). A suite of metrics focused more explicitly on urbanization-driven hydrologic changes was explored by Konrad and Booth (2002), who suggested that changes in flashiness, peak flow, and baseflow were a logical starting point for assessing stormwater impacts. Focusing on those attributes of the flow regime with likely biological effects, Konrad and Booth (2005) explored metrics that characterized the variability in high flows, low flows, daily flows, and the distribution of runoff between peak flows and baseflow. They tested these metrics on 13 small gaged watersheds across the USA—five that had undergone little land-cover change over an 80-year gage record and eight that had seen substantial urbanization over the same period. They found that no single metric reliably discriminated "urbanized" from "non-urbanized" watersheds; and no urbanized watershed showed a systematic change in every hydrologic metric. Thus, there is no "quick fix" for detecting and characterizing the effects of watershed-scale urbanization on flow regime, but many metrics show promise, and a diverse suite is most likely to provide the most robust indications of hydrologic conditions and change.

In the Pacific Northwest, DeGasperi et al. (2009) explored the relationship between a biological indicator (the benthic index of biotic integrity [B-IBI; Karr, 1998]), watershed imperviousness, and hydrology through the investigation of eight hydrologic metrics (Table 1). This work was continued in King County (2012), which made use of relationships developed between hydrologic metrics and B-IBI scores to evaluate the potential *biological* effectiveness of alternative stormwater management approaches to flow control. As with the results of Konrad and Booth (2005), all correlations between any given flow metric and B-IBI scores were imperfect, although the overall trends were as hydrologic theory and biological inference would anticipate—flashy, more "urban-dominated" flow regimes correlated with degraded biological conditions.

TABLE 1 Metrics used by DeGasperi et al. (2009) and King County (2012) to evaluate the relationship between flow alternation and biological condition in small urban and suburban streams

Metric name	Description
Low-pulse count	Number of times each calendar year that discrete low flow pulses occurred
Low-pulse duration	Annual average duration of low flow pulses during a calendar year
High-pulse count	Number of days each water year that discrete high flow pulses occur
High-pulse duration	Annual average duration of high flow pulses during a water year
High-pulse range	Range in days between the start of the first high flow pulse and the end of the last high flow pulse during a water year
Flow reversals	The number of times that the flow rate changed from an increase to a decrease or vice versa during a water year. Flow changes of less than 2% are not considered
T_{Qmean}	The fraction of time during a water year that the daily average flow rate is greater than the annual average flow rate of that year
R-B Index	Richards-Baker Index—A dimensionless index of flow oscillations relative to total flow, based on daily average discharge measured during a water year

1.3 | Approach

This study seeks a scientifically defensible methodology that resource agencies and municipalities can apply to satisfy the multiple needs for hydrologic monitoring of urban streams. Our approach is to test the utility and sensitivity of a subset of previously defined hydrologic metrics on a preexisting set of agency-collected flow data, with the assumption that this data set is a credible example of what could be considered "feasible" for similar, future efforts. We focus first on multiple sites within a limited geographical area of North America (western Washington State, USA, within the Pacific Northwest region; Figure 1) to minimize intrinsic between-site variability. The objectives are to identify the most efficient set of hydrologic metrics, collected over the least amount of time, that are capable of yielding meaningful results with minimal cost (thus maximizing feasibility that the monitoring may actually take place). We then explore the applicability of these findings from western Washington to a substantially different climatological regime (southern California, USA), acknowledging that conclusions from even two disparate regions may not fully apply to other parts of the North American continent (or beyond). Lastly, we evaluate the potential to improve the feasibility of implementation without compromising value by eliminating one of stream gaging's most costly elements, direct discharge measurements.

2 | METHODS

2.1 | Selection of flow metrics for hydrologic monitoring

For practical applications in monitoring programs, there is a trade-off between a large, comprehensive set of metrics that can fully characterize all aspects of the flow regime; and a more modest, more easily interpreted set that addresses the key elements of an urban-altered hydrologic regime. To reduce the list of prospective hydrologic metrics to a robust yet tractable number for broad application in monitoring programs, two criteria were applied to the metrics offered by DeGasperi et al. (2009; Table 1): strength of the biology–hydrology correlation and the potential for common stormwater-management approaches to influence the value of the metric over time. So, for example, "high pulse range" (the number of days between the first and last high flow of a water year) shows a good correlation with B-IBI, but urban stormwater management is not as likely to influence this metric as, for example, a measure of within-storm flashiness or peak discharge.

Applying this rationale (and the empirical results of biology–hydrology correlation presented in King County, 2012), three annual-scale metrics from Table 1 were selected for evaluation as prospective components of a hydrologic monitoring program for urban and urbanizing watersheds: T_{Qmean}, Richards-Baker Index (henceforth, RBI), and flow reversals.

T_{Qmean} is the aggregate fraction of time during a water year that a hydrograph lies above the mean discharge for that water year (Konrad & Booth, 2002). Thus, a stream whose hydrograph is primarily a slowly varying baseflow with only limited peak flows will spend most of the time close to its mean annual discharge, resulting in values of T_{Qmean} at or above 0.40. In contrast, a very flashy hydrograph will have peaks that greatly exceed the *magnitude* of baseflow, thus raising the overall annual mean discharge, but the *duration* of those excursions may be rather brief. T_{Qmean} values for such systems commonly fall to values around 0.20.

The RBI (Baker, Richards, Loftus, & Kramer, 2004) is calculated for each water year as the sum of all day-to-day discharge differences (i.e., the absolute value of the difference between today's flow and yesterday's flow) divided by the sum of daily discharges. Graphically, the numerator represents the length of the line in time-discharge space making up a continuous average hydrograph, whereas the denominator is the sum of all daily discharges over the same period. Although "flashiness" is a phenomenon that can be expressed at multiple time scales, the original definition of this index and all subsequent applications are based on daily average flows, a convention that we follow here as well.

Flow reversals are the simple tally of the number of days during the fall and winter seasons (specifically, 1 October to 30 April) when the flow has changed from a rising or a falling trend to its opposite from one day to the next. A minimum threshold of change is commonly applied to avoid counting minor fluctuations; following King County (2012), that threshold was set at 2% for the following analyses. Thus, for example, the daily sequence of discharges 90➔100➔95 would count as a reversal, but 99➔100➔99 would not.

For each of these metrics, their correlation with biological health (as measured by B-IBI) was found to be relatively strong and monotonic (DeGasperi et al., 2009; King County, 2012). In these aquatic systems, more uniform and less flashy flow regimes are associated with more diverse species assemblages with a greater proportion of intolerant species. Thus, biologically "better" conditions are associated with higher values of T_{Qmean} and with lower values of the RBI and the annual tally of fall/winter flow reversals. These relationships provide a clear basis to recognize the relative "status" of any given site

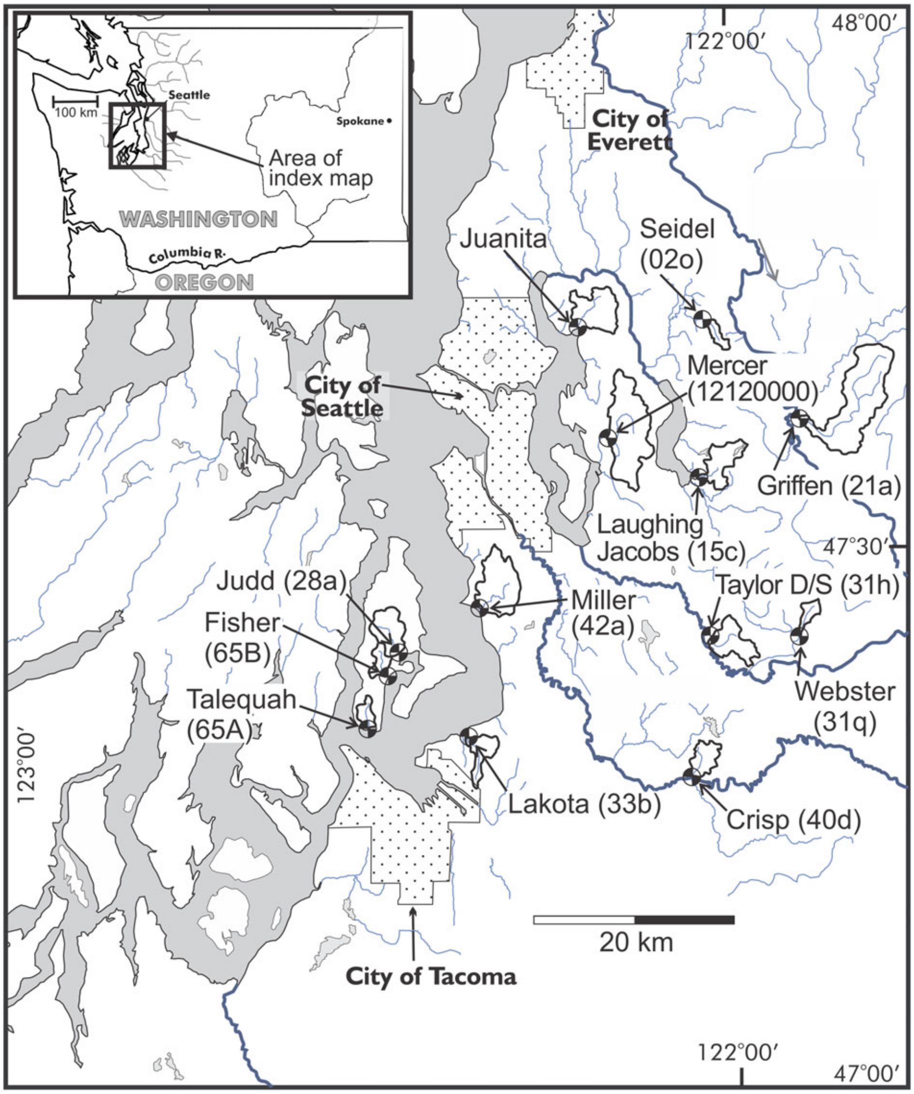

FIGURE 1 All sites considered in the hydrologic metric analysis for this section

TABLE 2 Site list. All data are from King County, except Mercer Creek (USGS gage 12120000). The watersheds fall into three natural groups based on their 2011 urban land-cover percentage and are so indicated by the shading. The three least urban watersheds (Webster, Griffen, and Fisher) serve as useful "control" sites insofar as they each have urban land cover less than 3%, forest cover greater than 60%, and essentially no discernable change in urbanization over the 10-year period covered by the 2001 and 2011 National Land Cover Databases

Gage	Gage #	Latitude	Longitude	Drainage area (km^2)	Start date Q (WY)	Stop data Q (WY)	Forest 2011 (%)	Urban 2011 (%)	Urban change 2001–2011 (%)
Webster	31q	47.4164	−121.9195	4.64	2010	2015	93.3	0	0.0
Griffen	21a	47.6163	−121.9070	44.54	2002	2015	62.9	0.3	0.0
Fisher	65B	47.3841	−122.4815	5.03	2005	2015	60.9	2.7	0.0
Tahlequah	65A	47.3345	−122.5089	3.98	2005	2015	81.4	4.5	0.0
Judd	28a	47.4034	−122.4688	12.12	2000	2015	62.2	4.7	0.1
Crisp	40d	47.2883	−122.0672	8.02	1995	2015	46.4	15.8	4.2
Seidel	02o	47.7117	−122.0519	3.75	2009	2015	53.7	16.9	15.5
Taylor D/S	31h	47.4207	−122.0412	13.17	1992	2015	40.0	22.4	0.9
L Jacobs	15c	47.5654	−122.0521	11.89	1992	2015	25.8	46.0	3.9
Lakota	33b	47.3288	−122.3726	8.96	1990	2009	9.9	71.6	3.1
Mercer	12120000	47.6031	−122.1797	32.30	1956	2015	12.0	71.7	1.2
Juanita	27a	47.7077	−122.2149	16.99	1993	2015	10.0	78.0	2.0
Miller	42a	47.4455	−122.3520	23.13	1989	2015	4.8	80.7	3.5

Note. WY = water year.

from the perspective of overall stream health (sensu Karr, 1998) on the basis of their flow metrics.

2.2 | Study region and sites

The three selected hydrologic metrics were tested across a range of streams in the Pacific Northwest where the primary difference was the degree of watershed urbanization, with all located in the Puget Sound lowlands of western Washington State (USA; Figure 1). The region is relatively homogenous, experiencing a cool maritime climate characterized by mild, wet winters and warm, drier summers. The City of Seattle, which lies in the approximate geographical centre of the region, receives an annual average of 97 cm of precipitation, most of which falls as rain in long-duration, low-intensity storms. The rainy season extends from October to April (and sometimes beyond), with over 80% of annual precipitation falling during that period. The region's postglacial landscape of broad plateaus and valleys supports an abundance of smaller streams, which generally drain watersheds up to a few tens of km^2 at moderate gradients (0.5–3% slope). These streams have historically provided anadromous salmonid habitat and consequently have become the focus of more intensive attention and monitoring in response to Endangered Species Act listing of several Puget Sound salmon runs.

2.3 | Hydrologic data selection and analysis

Every stream with hydrologic data available from the King County Hydrologic Information Center (http://green2.kingcounty.gov/hydrology/) was evaluated; streams were excluded from further analysis if they had less than 5 years of record, and their level of urban land cover was close to other site(s) with lengthier records or because they were tributary to a farther downstream gage. On this basis, data were available for 12 second-order and third-order streams in the central and east-central Puget Lowland that drain 3–50 km^2 with a relatively long record (at least 10 years for most) of daily average flow data; and with climate, elevation, and soils typical of the central Puget Sound lowlands (Table 2). Their contributing watersheds span a wide range of urbanization, from nearly undeveloped to more than 70% urban land cover. An additional non-King County gage site, Mercer Creek (USGS gage 12120000), was also included because it has the longest record (60 years) of any site within the study region, and the data are of equivalent quality and presentation.

Value for watershed urban land cover was derived from the 2001 and 2011 National Land Cover Database (Homer et al., 2015), summing up categories 22, 23, and 24 ("Developed, Low Intensity," "Developed, Medium Intensity," and "Developed, High Intensity," respectively) within the contributing watershed to each gage, expressing their sum as a percentage of the contributing watershed area as calculated in ArcMap 10.1. Although King County was an early adopter of stormwater regulations (the first such regulations were implemented in 1979), only in recent years have the mitigation measures shown any potential for effectiveness (e.g., Booth & Jackson, 1997) and in total are not judged to materially affect the conclusions of this work, namely, the suitability of hydrologic metrics to detect a long-term impact of urbanization.

Hydrologic data were available in 15-min, hourly, and daily time increments for the King County gages and daily (only) for Mercer Creek. To maintain homogeneity of the data set and consistency with most prior analyses of the hydrologic metrics, a daily average discharge was calculated (as needed) and used for all sites. Pairwise correlations between the selected hydrologic metrics and between these metrics and land cover were determined without assuming normal distribution using the Kendall Rank Correlation test (R Core Team, 2016). Correlation coefficients with p values <0.05 were deemed significant.

Although we acknowledge the strong preference for International System units among the scientific community, we developed these

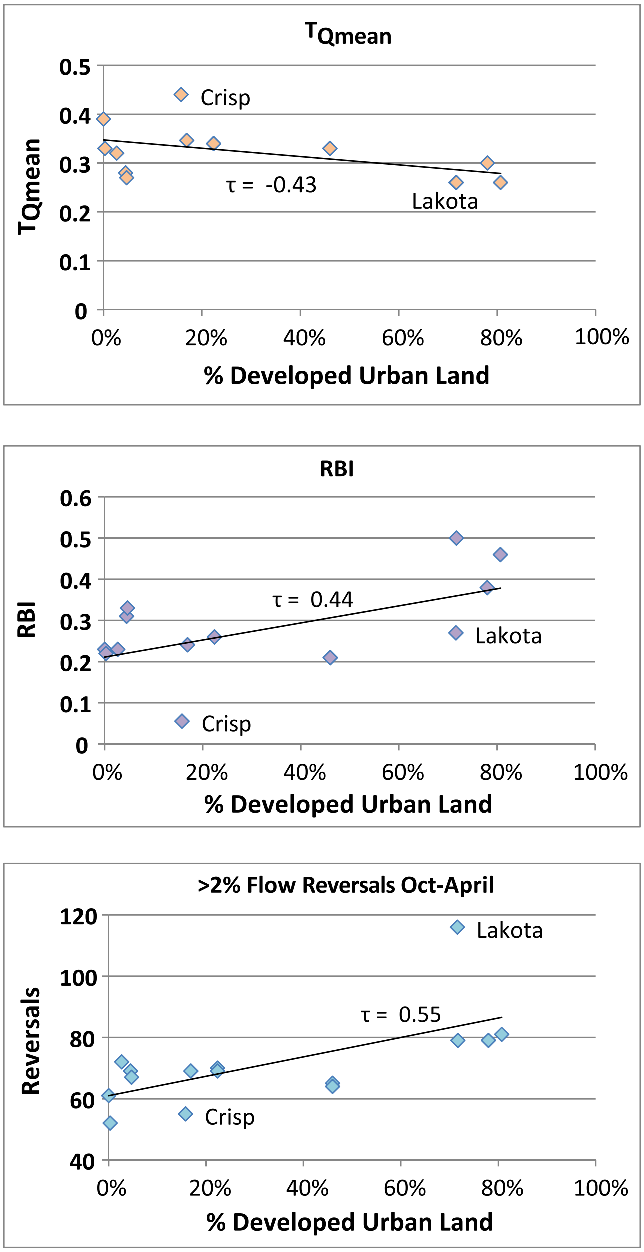

FIGURE 2 All sites with analysed hydrologic data, with the specified metric plotted against the 2011 watershed urban land cover (all p values <0.05). Each point is a single site, calculated for each individual water year independently, with all annual values then averaged over the full period of record. The two labelled sites (Crisp and Lakota) display somewhat anomalous relationships relative to the trends established by the others. RBI = Richards-Baker Index

analyses in the U.S. Customary System units of discharge (feet, cubic feet per second) to preserve the precision of the original data. We offer the conversion to International System units where appropriate.

2.4 | Climate stationarity

We offer no formal test of any potential influence of climate change of the periods of record covered by these gages, in part because inspection of long-term records suggests no obvious systematic changes over this period that outweigh those of urbanization, and in part because a prior investigation of this topic found little evidence that such changes could be reliably identified (Rosenberg et al., 2010). Looking ahead to infer how such changes might affect the utility of these metrics in the future is even more problematic: "… simulations generally agree that peak discharges will increase, although the range of predicted change (from a slight decrease to a near-doubling, depending on the selected recurrence interval, watershed, and underlying GCM simulation) is much too large to provide a basis for engineering design. The comparative simulation results are most confounding for the smallest watershed areas, wherein even the net direction of change (i.e., a future increase or a future decrease) is in part dependent on the choice of GCM" (Rosenberg et al., 2010, p. 346).

3 | RESULTS

3.1 | Land cover–hydrology relationships

The 13 selected gage sites, in aggregate, display the anticipated relationships between urban land cover and hydrology: with increasing levels of urbanization, the full-record average values of T_{Qmean} decreased, the RBI increased, and the tally of annual fall/winter flow reversals increased (Figure 2). However, the significant scatter in the graphs of all metrics reinforces the long-standing recognition that "urban land cover" is a good but not perfect surrogate for hydrologic alteration of a watershed, and that each metric responds differently within a given watershed setting.

Examples of local disparities within an overall urban-driven trend are readily identified. For example, the flow of Crisp Creek, with a moderate 15.8% watershed urban land cover, is supported by abundant deep groundwater flow (which is why a tribal fish hatchery has made use of its cold, reliable flow since 1987). This site is an outlier on the plots for all three metrics, because the steady groundwater flow dampens the expression of urban flashiness. In contrast, Lakota Creek (71.6% urban land cover) is a steep tributary to Puget Sound that drains a largely urban and suburban watershed. It is fully "on trend" with respect to T_{Qmean} relative to other watersheds of comparable urban land-cover percentages (e.g., that of Mercer Creek has an identical T_{Qmean} value of 0.26 with an urban land cover of 71.7%), but its RBI is below the regional trend (i.e., less flashy) whereas its flow reversals are well above the corresponding trend (i.e., more flashy).

3.2 | Comparisons between metrics

Differences between metrics at the same site can be assessed more systematically by comparing their pairwise behaviour to one another. Figure 3 shows these comparisons, which demonstrate the overall good but not perfect correspondence between metric pairs. As with the consistent relationships that are expressed between urban land cover and hydrologic metrics, other attributes of each watershed are also likely to impose similar responses on each of the metrics—not only urban land cover but also baseflow contribution, hillslope and channel gradients, and watershed size. Thus, the relatively good correlation between metrics (particularly between T_{Qmean} and RBI, two related measures of the magnitude of high flow peaks relative to more common, persistent flows) is not surprising. It also suggests that seeking yet additional metrics for evaluation may not result in a commensurate increase in understanding.

3.3 | Ability to detect trends

These data sets are well suited to evaluate the ability of these metrics to detect changes over time, given the decade to multidecade length for many of them and the parallel availability of land-cover data from both 2001 and 2011, a period fully covered by nine of these records. The aggregated results from this subset of nine sites, however, are not particularly encouraging, and they do not offer much promise for systematic detection of decadal-scale hydrologic trends, even for those watersheds with relatively rapid rates of change (Figure 4).

Although several watersheds showed changes in specific metrics beyond the range of variability defined by the near-"control" sites (i.e., those with land-cover change ≈ 0 in Figure 4), no site shows a consistent response in all three metrics. Reversals at the control sites define the widest range of weather-induced variability observed during the monitoring period, for which only Taylor and Crisp Creeks exceed; and for those two, the apparent trend of Crisp Creek suggests a *less* flashy regime, despite its relatively high rate of urban land-cover change, whereas the trend for reversals at Taylor Creek contradicts those for T_{Qmean} and RBI. Seidel Creek has a relatively short hydrologic record (spanning 7 years in total with data available only for water years 2009, 2010, 2011, 2012, and 2015) but the greatest change in urban land cover between 2001 and 2011 (15.5%, with an accompanying decrease in forest cover of 23.5%). Rapid suburban development of its watershed in the decade of the 2000s (Figure 5) resulted in significant hydrologic changes for Seidel Creek, but even here, the (relatively sparse) data paint a somewhat ambiguous picture of hydrologic changes (Figure 6).

4 | DISCUSSION

4.1 | Minimum duration of monitoring for detecting trends

Status-and-trends monitoring programs typically have a predetermined lifespan. The "status" obviously can be determined at the outset of such a program, but detecting any "trends" requires an assumed period over which change will be expressed. The potential limitations of an overly brief monitoring period are best explored in Mercer Creek, which benefits from a near-continuous 60-year hydrologic record that spans a period when urban development was only just beginning in this 32-km^2 watershed up to its current condition

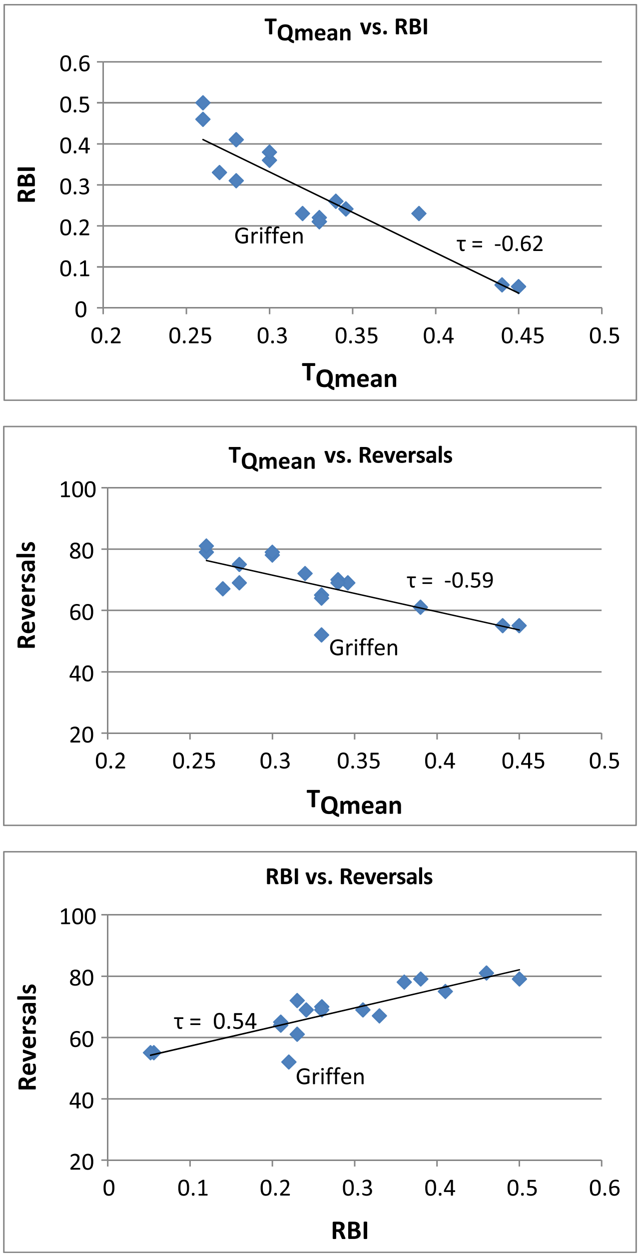

FIGURE 3 All sites with analysed hydrologic data, with the specified metrics plotted against one another using the values for each individual water year averaged over the full period of record to minimize the effects of interannual variations in precipitation on metric values (all *p* values <0.02). The only systematic outlier is Griffen Creek, which has many fewer flow reversals then either of its other two indictors might otherwise suggest. Possible explanations for this behaviour are that the watershed is the largest of this group and has one of the largest fractions of wetlands (>5% watershed area) of any site. RBI = Richards-Baker Index

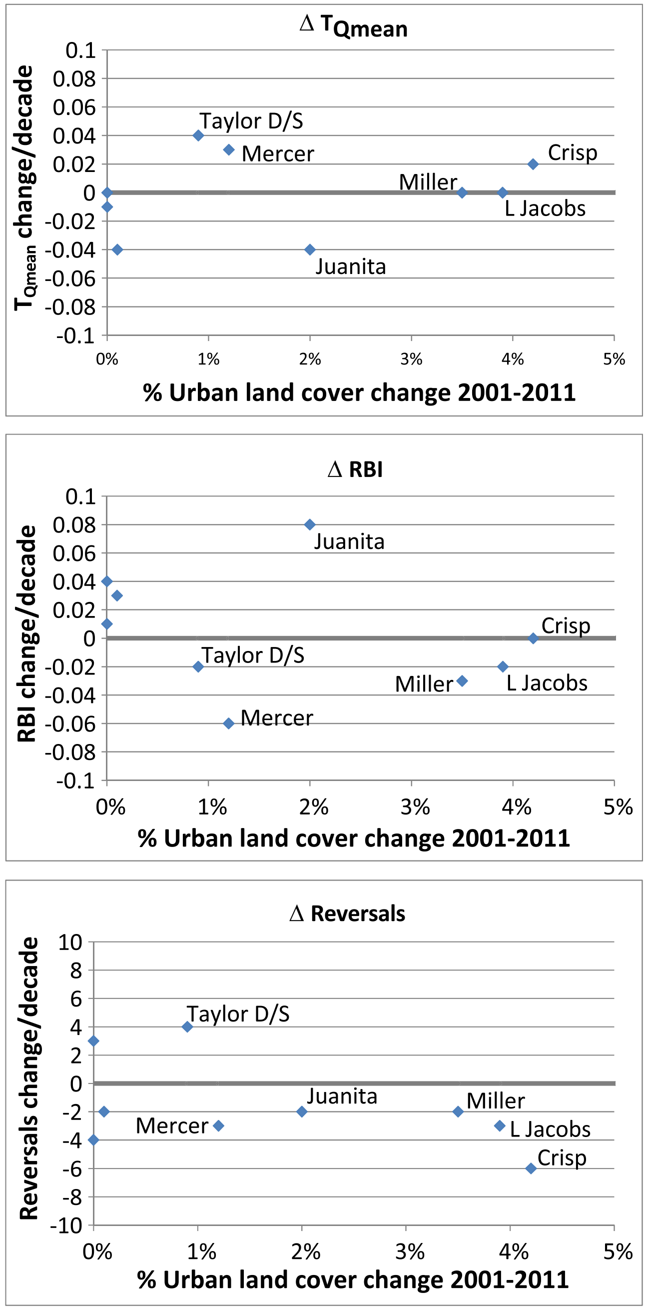

FIGURE 4 Rate of change in hydrologic metrics as a function of urban land-cover change. No site with active urbanization during the decade 2001–2011 shows a consistent pattern with respect to all three metrics, and none of these change relationships are statistically significant (all *p* values >0.05). The three sites with little or no land-cover change (plotting close or on the *y*-axis and all with <5% urban land cover as of 2011) suggest the range of natural variability for each of these metrics on a decadal time scale. RBI = Richards-Baker Index

of more than 70% urban land cover. The trends for all three hydrologic metrics here are significant and consistent over the full period of record (Figures 7 and 8), likely covering times when urban land cover would have been increasing as rapidly as any other site in this study over the last 10 years (i.e., >5%/decade). The metrics also all suggest a possible trend reversal over the last ~10 years or so, particularly well expressed by a reduction in the RBI but also displayed in T_{Qmean} (an increase) and in flow reversals (a less distinct reduction). These long-term records also suggest that the RBI has the lowest interannual variability and flow reversals the greatest.

However, even for the hydrologic metric with the lowest variability (RBI), at least 2 to 3 decades of record would have been necessary to identify a consistent trend. Land-cover changes over a single decade at any of the sites, overall, do not produce any systematic, statistically significant changes in the hydrologic metrics (Figure 4). Although widespread and highly effective stormwater management could

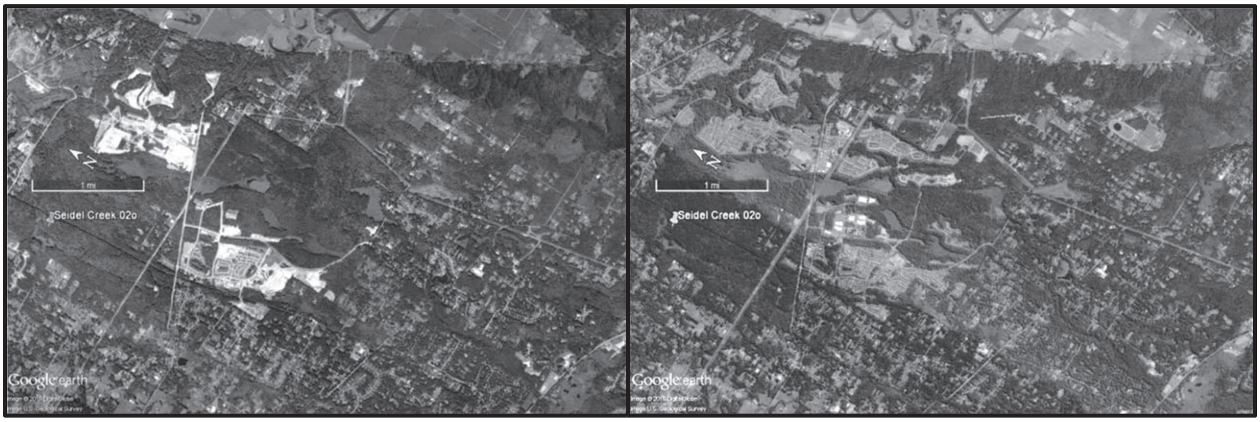

FIGURE 5 Aerial views of Seidel Creek watershed from 2002 (left) and 2014 (right). Imagery from Google Earth. Over one-half square kilometre of this 3.5 km^2 watershed converted to urban land cover during the decade 2001–2011

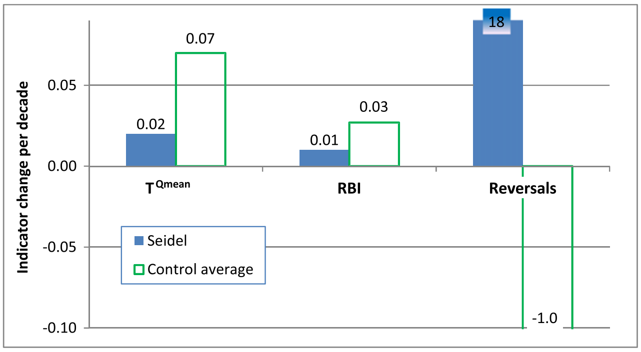

FIGURE 6 Decadal rate of change in hydrologic metric values for Seidel Creek, extrapolated from the period 2009–2015. Trends in this watershed imply *less* hydrologic changes than those expressed for the control sites using T_{Qmean} or Richards-Baker Index (RBI); however, the trend for reversals is dramatically more rapid (i.e., more urban) that for those same control sites

presumably produce more rapid hydrologic responses (a plausible but by no means demonstrable explanation for the recent metric trends for Mercer Creek), a multidecadal time frame is likely to be the minimum duration of monitoring that would be required to detect statistically meaningful trends in hydrologic metrics at a watershed scale (see also Konrad & Booth, 2002).

4.2 | Comparative performance of flow metrics in maritime and semi-arid regions

This evaluation of hydrologic metrics may lack universal applicability, because trends displayed by the metrics that respond to watershed urbanization may also have a strong dependency on the general patterns of rainfall and the relative magnitude of extreme events (Booth et al., 2016). Thus, the above analyses presently can be applied with confidence only to a region of low variation in flood magnitudes relative to the mean annual flood (e.g., Farquharson, Meigh, & Sutcliffe, 1992; Lewin, 1989). However, an opportunity to test the utility of these same flow metrics in a very different region, that of semi-arid watersheds in southern California, USA, has been provided by a recent compilation of gage records that span the mid-20th century urbanization of this region (Hawley & Bledsoe, 2011). Three USGS gages from this compilation provide useful case studies, displaying consistent

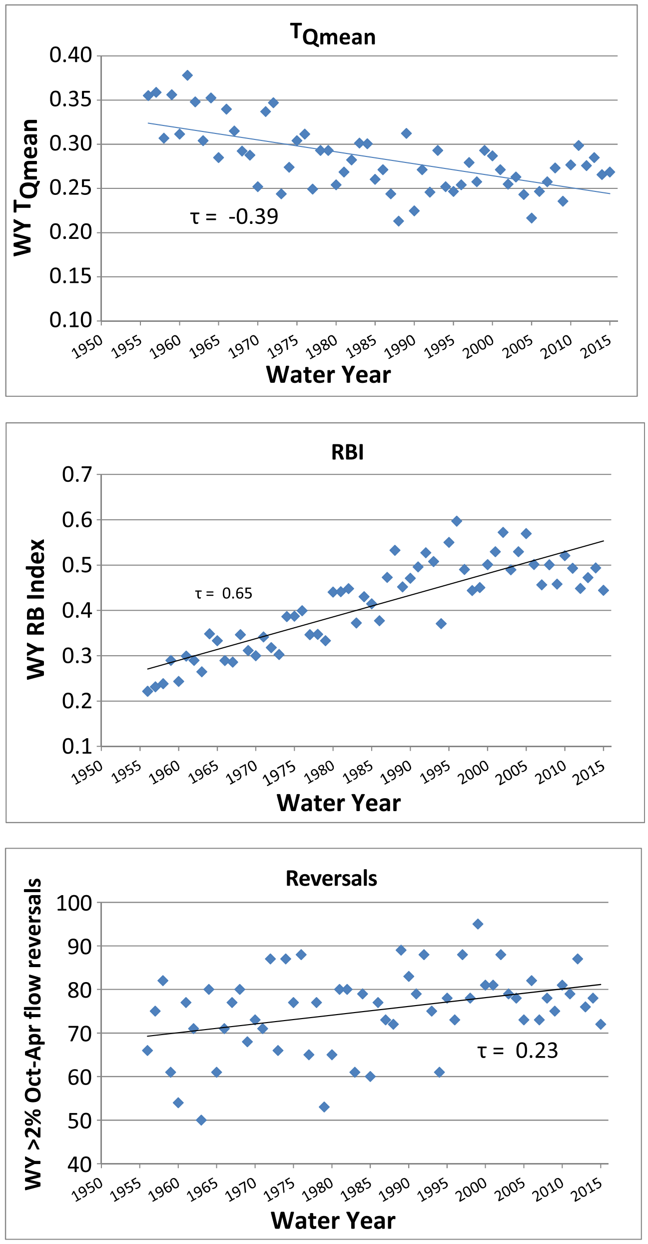

FIGURE 7 Water-year (WY) values of the three hydrologic metrics for Mercer Creek (USGS gage 12120000), the longest record in the data set (all *p* values <0.01). RBI = Richards-Baker Index

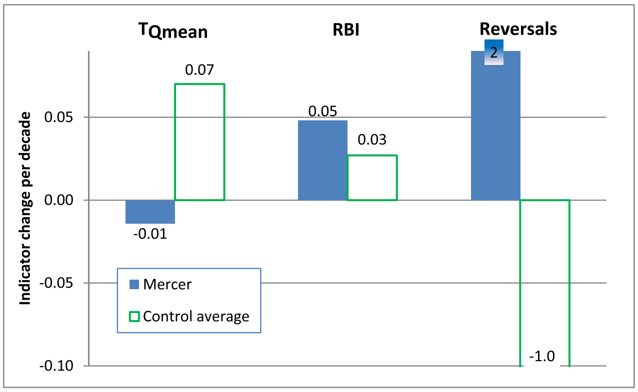

FIGURE 8 Average rate of change in hydrologic metrics for Mercer Creek over the full period of record (water year 1956–2016, with control-site averages from the last decade provided for reference). All Mercer Creek changes are consistent with increasing urbanization, supporting an inference of long-term increase in flashiness corresponding to the multidecadal period of urbanization in the watershed. RBI = Richards-Baker Index

metric responses to one another but which are significantly different from those of the Pacific Northwest.

The original analysis of these gage records by Hawley and Bledsoe (2011) emphasized changes in flood recurrence and flow durations. Arroyo Simi (USGS gage 11105850, 180 km^2 drainage area; 34°16′23″N, 118°47′13″W) was their primary case study for watershed urbanization. It displayed more than a tenfold increase in the 2-year flow and a threefold increase in the 25-year flow between preurban (1938–1958) and posturban (1959–1983) periods. In their nearby non-urban control watershed (Hopper Creek; USGS gage 11110500; 62 km^2 drainage area; 34°24′03″N, 118°49′32″W), these flood discharges differed by only 20% between these same two periods. Despite this consistency in peak-discharge response to urbanization, the annual trends in both T_{Qmean} and RBI for Arroyo Simi showed no systematic patterns over the period of mid-century urbanization—but that of flow reversals was dramatic and suggests a relatively abrupt initiation of urbanization in the early 1960s (Figure 9).

Additional results were obtained by applying the same three metrics to another southern California gage, that of Aliso Creek (USGS gage 11047500; 33°37′34″N, 117°41′03″W), selected from the Hawley and Bledsoe (2011) gage inventory as a third southern California example because it has one of the highest modern impervious-area percentages (20.3%) and drains one of the smallest total drainage areas (22.6 km^2, similar to those from the Pacific Northwest above). It too underwent significant urbanization in the early 1960s, as suggested by the abrupt increase in flow reversals around that time. Also as with Arroyo Simi, however, the changes in both T_{Qmean} and RBI over time do not match their anticipated response to mid-century urbanization (Figure 10). Thus, interregional application interpretation of hydrologic metrics must be approached with great caution.

4.3 | Suitability of stage as a surrogate for discharge

Accurate streamflow gaging can require significant levels of both expertise and time/cost, because it requires not only the continuous recording of water level (stage) but also relatively frequent site visits

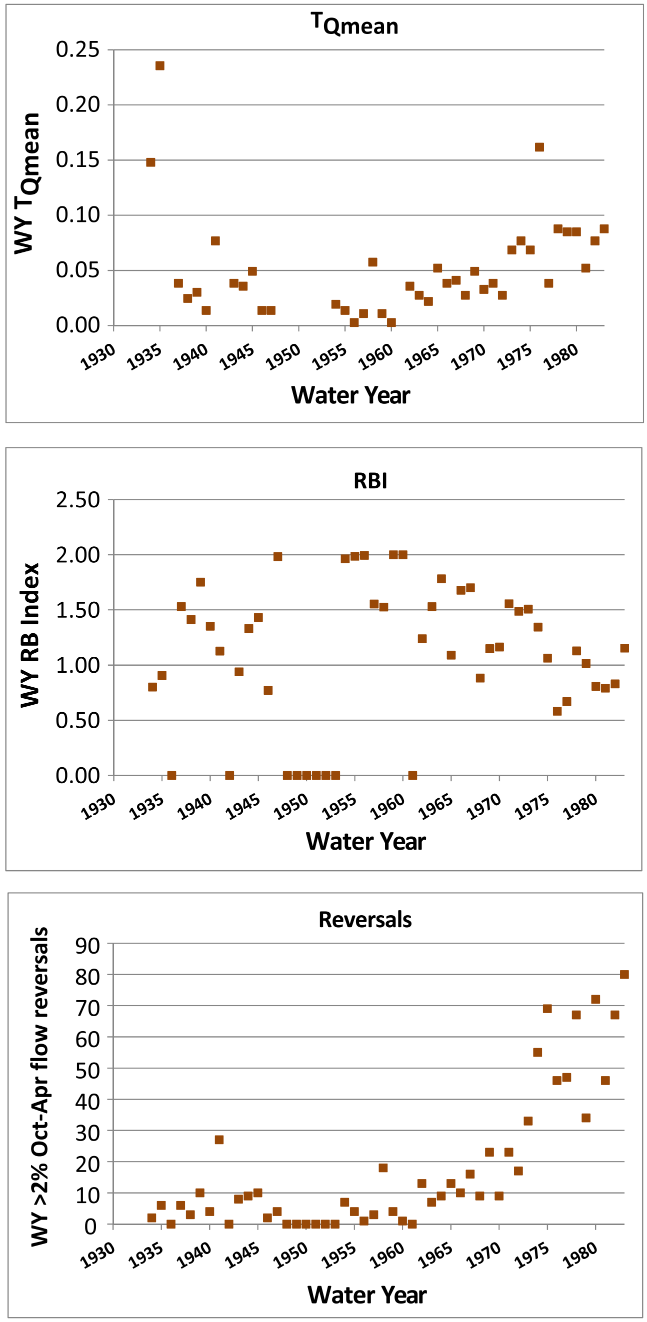

FIGURE 9 Water-year (WY) values of the three hydrologic metrics for Arroyo Simi (USGS gage 11105850). Only the pattern of flow reversals is consistent with mid-century advent of watershed urbanization (Hawley & Bledsoe, 2011); both T_{Qmean} and Richards-Baker Index (RBI), if naively interpreted, would imply an early record increase in flashiness followed by a marked decline. Missing or zero values, particularly in the early 1950s, are years of no recorded flow at all

to directly measure discharge. The resulting relationship between recorded stage and measured discharge (the rating curve) is generally considered accurate only within the range that discharges have been measured (i.e., it is reliable for interpolation but progressively less so for extrapolation), which requires site visits during times of high or peak flow. High flow measurements in small urban streams are particularly challenging because of rapidly changing streamflow over the time required for a streamflow measurement. Of course, this also will typically correspond to times when *every* such site is experiencing

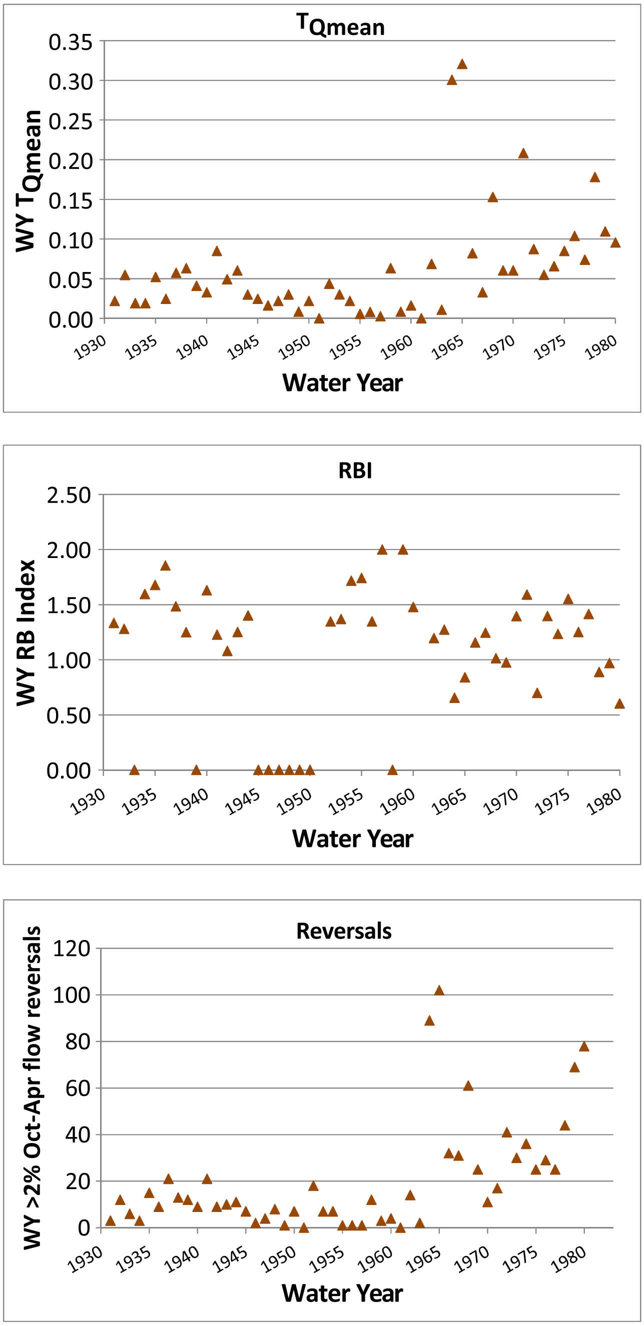

FIGURE 10 Water-year (WY) values of the three hydrologic metrics for Aliso Creek (USGS gage 11047500). As with Arroyo Simi (Figure 9), only the pattern of flow reversals is consistent with mid-century advent of watershed urbanization. The responses of T_{Qmean} and RBI indicate a dramatically more flashy system overall than those of the Pacific Northwest (compare ranges with those of Figure 2), but the responses of these two metrics to urbanization are opposite to those in the Pacific Northwest. RBI = Richards-Baker Index

such flows, making measurement logistics difficult for a limited number of trained crews. In addition, the underlying relationship between stage and discharge can change, most commonly as a result of erosion or sediment deposition at the gaging site, and so rating curves must be developed a new following significant (or potentially significant) channel-altering events. These requirements all increase the cost of collecting discharge records, a significant determinant in whether or not *any* hydrologic data are available for analysis.

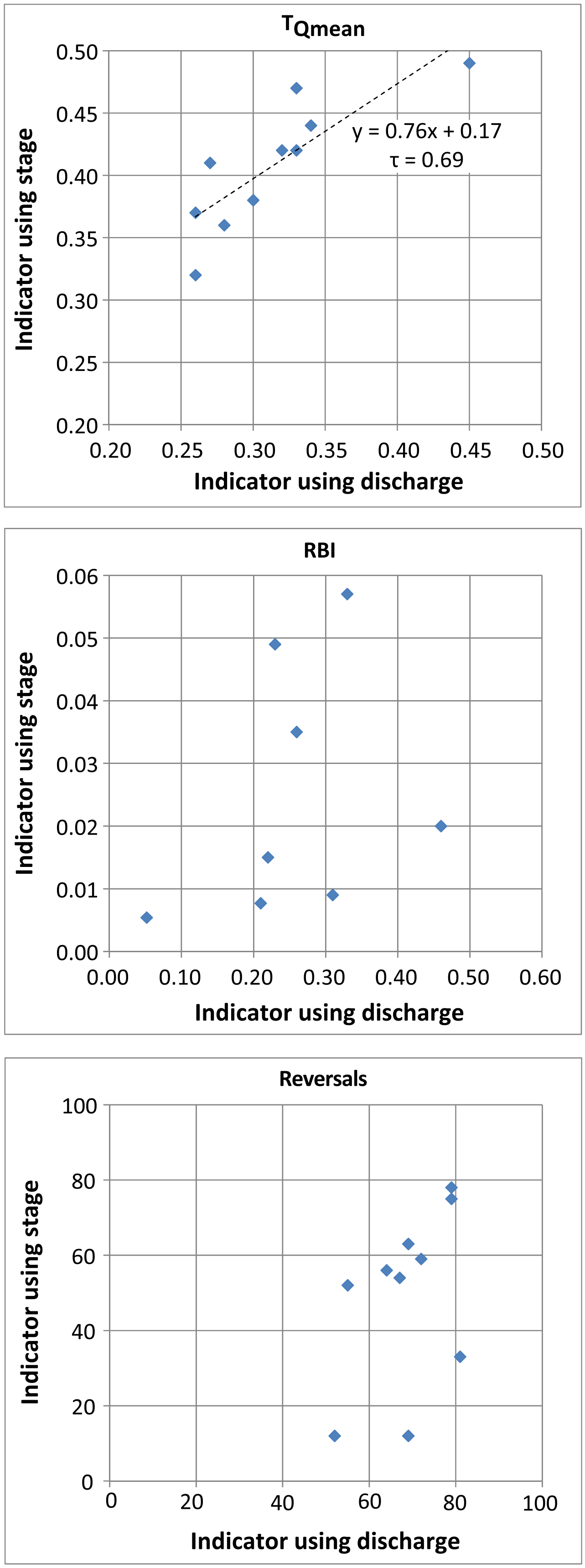

FIGURE 11 The three hydrologic metrics recommended for use, comparing the decadal-averaged values for each site calculated using the discharge record (*x*-axis) and the stage record (*y*-axis). Only T_{Qmean} shows as a useful relationship with the data as they are presently archived. RBI = Richards-Baker Index

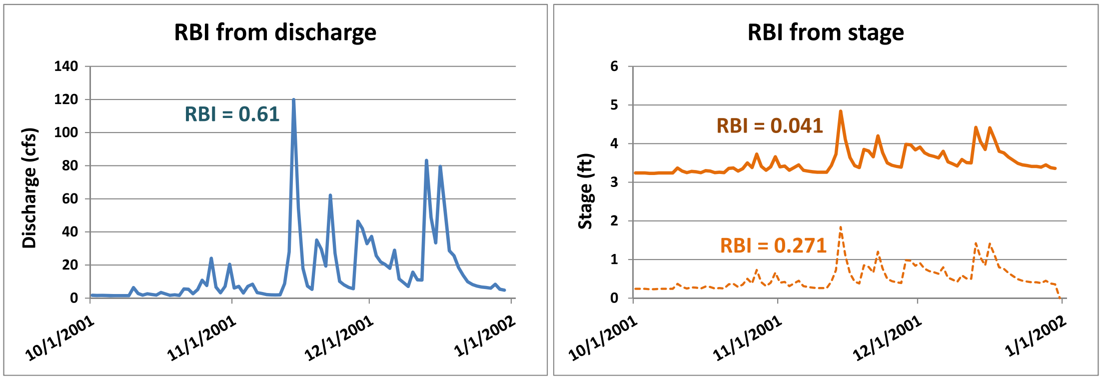

FIGURE 12 Comparison of discharge (left) and stage (right) records from Miller Creek (gage 42a) for the first 3 months of water year 2002. The stage record as reported is the upper curve on the right panel; the lower curve reflects an arbitrary lowering of the datum, as might occur after a scouring event in the channel or if the gage location were moved. Although no physical change exists between the two records, the calculated Richards-Baker Index (RBI) from the "shifted" record is more than five times larger. This metric shift would not occur if the data were of actual flow *depth*, rather than stage. cfs = cubic feet per second; ft = feet

However, for many applications, the conversion of stage to discharge may be unnecessary. Because discharge is normally a calculated value derived from stage, those parameters that depend on patterns or variations in discharge should be more accurately represented by direct evaluation of the raw data (i.e., stage, with the understanding that any temporal changes that alter the stage–depth relationship still need to be recognized). In addition, many of the issues associated with fluctuations in the flow, such as sediment transport or substrate disturbance, are only dependent on stage (because stage should function as a direct measure of flow depth, a key determinant of the tractive stress that mobilizes sediment); the absolute discharge is in fact irrelevant. Only for those applications that require a direct knowledge of the flow magnitude (e.g., culvert capacity, floodplain inundation, solute and suspended material loads) is the conversion to discharge mandatory.

For these reasons, exploring the use of stage data as a surrogate for discharge was identified as a complementary topic for hydrologic monitoring of urban streams. In general, hydrologic metrics have been developed and implemented solely on the basis of discharge, and so the purpose of this exploration was to determine the degree to which stage can be used effectively as a surrogate for discharge and to identify any potential pitfalls to the naïve substitution of one measurement (i.e., stage) for another (discharge). In contrast to the rich literature on discharge-based hydrologic metrics, evaluation of stage-based metrics is not readily identifiable in either agency reporting or formal publications. We therefore consider this a preliminary evaluation only, with the hope that it will spur further evaluation of whether the reduced cost of implementation improves the feasibility of collecting meaningful data (e.g., Fanelli, Prestegarrd, & Palmer, 2017).

The methods used for this comparison were analogous to those described above for the data selection and calculations based on discharge. The same set of gage records (Table 2) was mined for suitable data sets to evaluate the potential substitution of stage for discharge for calculating and interpreting hydrologic metrics. Although stage must have been recorded for all dates with reported discharge, the data are not readily available for all such entries. From the population of gage records used to evaluate the hydrologic metrics, 10 have at least 10 years of jointly reported daily stage–discharge data from which comparisons can be made. Evaluations of both individual years and record-averaged values and trends were made to determine the suitability and the limitations of using stage records without needing to invest the additional effort in developing and maintaining a rating curve. Statistical significance was evaluated as described for the discharge-based metric calculations.

Results from calculating the three metrics based on both the discharge record and the stage record are mixed (Figure 11). T_{Qmean} shows by far the most consistent relationship, although values of this metric calculated on stage are consistently higher than when calculated on discharge. This result is reasonable, given that stage (i.e., flow depth) is typically related to discharge as a power function with an exponent less than 1 (Leopold & Maddock, 1953). The variance of stage relative to its mean value will be lower than

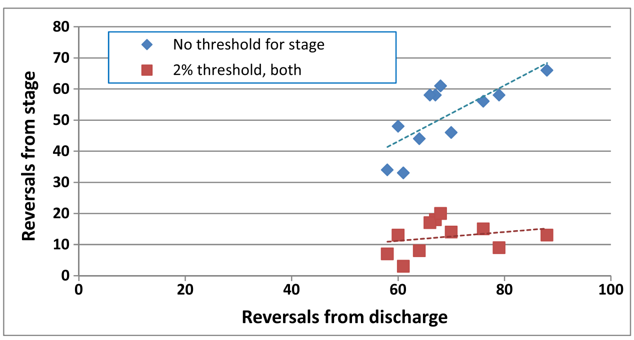

FIGURE 13 Comparison of alternative flow reversal records (from Tahlequah Creek [gage 65A], the site with the worst naïve correlation of reversal calculations between stage and discharge values). Using the 2% minimum day-to-day threshold for identifying flow reversals on both the (recorded) stage record and the (calculated) discharge record, there is essentially no correlation between the two metrics, with a fivefold (or greater) difference between them in any given year. Eliminating the threshold for identifying stage reversals improves the correlation, although neither achieves a p value <0.05. The remaining mismatch is likely a consequence of rounding the reported stage values (which span only a 2-ft range over the period of record and are reported to the nearest 0.01 ft [3 mm], whereas discharge spans an order of magnitude greater range of values but is also reported with a precision of 0.01 cfs [3×10^{-4} m^3/s])

discharge, so T_{Qmean} calculated from stage should be consistently closer to 0.5 than T_{Qmean} calculated from discharge. The overall close correspondence between these two alternative approaches, however, suggests that this metric could be calculated and interpreted using either data set with only minimal uncertainty associated with its use or integration with prior studies.

The other two metrics, however, have rather poor correlations between calculations using the two alternative data sets, and so these require further discussion.

The RBI, the quotient of summed day-to-day discharge differences divided by the sum of daily discharges, depends not only on the magnitude of interday fluctuations (an intuitive measure of flashiness, which is why the RBI is widely used) but also on the overall magnitude of the denominator. Using discharge data, this relationship is understandable: an interday fluctuation can be considered "large" only in the context of the overall magnitude of discharge. However, the "magnitude" of stage is entirely arbitrary, because the datum from which it is measured can be any value (and may well change from year to year or even within a single water year; Figure 12).

This result does not require that RBI be calculated only from discharge, but it does require that the actual flow *depth* (i.e., a physical measurement of the flow) be preserved from the original field measurements and pressure transducer record. This is not commonly done, and it would need to be incorporated into any procedure that sought to avoid the added time and expense of creating stage–discharge rating curves. Unlike stage, depth is not an arbitrary value, and fluctuations around an average depth are quite likely to have physical and biological importance. Without these data, however, extraction of a meaningful value of the RBI is not possible. And as with T_{Qmean}, the fractional power–law relationship between discharge and depth means that the range of RBI values (even if correctly calculated on actual flow depth) will be compressed relative to the discharge-calculated values, and so direct comparison between the two alternative approaches will not be meaningful.

Flow reversals, the tally of daily flow reversals during the fall and winter seasons that exceed a specified threshold to avoid "counting" even miniscule reversals in the annual total (here, 2%), should in principle be entirely unaffected by whether stage or discharge is the variable being used, because any discharge record is based on a monotonic function of stage (i.e., if stage increases, then calculated discharge increases and vice versa). The poor correlation between these two approaches (Figure 11, right) is therefore not an intrinsic shortcoming of the data but rather of its typical implementation.

A given change in discharge will likely reflect a somewhat smaller change in stage, but that attenuated response should not alter the fundamental relationship between flow reversals calculated from these two alternative parameters. Although the tally of discharge reversals here invariably exceeded stage reversals for every site, these two parameters showed no systematic pattern to one another using the 2% threshold for identifying a "qualifying" reversal for both (Figure 13, squares). This correlation is improved by imposing no threshold for identifying reversals in the stage record (Figure 13, diamonds), but it is not eliminated.

The source of these anomalous findings appears to reside in data management. In the gage records from King County and the USGS, discharge has been calculated with the full precision of the original stage data but the archived stage is only reported to three significant digits. Thus, identically reported day-to-day records of the stage may nonetheless be associated with day-to-day differences in calculated discharge (and for which many examples can be observed in the downloaded records). To eliminate this shortcoming, the full precision of the recorded stage data would need to be preserved throughout the archiving and calculating of stage-based reversals.

5 | CONCLUSIONS

Our results can provide useful guidance to local and regional agencies seeking to evaluate the hydrologic effects of watershed urbanization. Three metrics—T_{Qmean}, the RBI, and the annual tally of wet-season interday flow reversals—all successfully stratify watersheds across a range of urban development and display statistically significant multidecadal trends that quantify the long-accepted hydrologic responses of humid-region watersheds to urbanization. However, they cannot reliably detect trends in watershed-scale urbanization over the course of a single decade. They are also less well suited to urbanizing watersheds in a semi-arid climate, with only flow reversals showing a response consistent with prior findings from more humid regions. Using stage as a surrogate for discharge in the calculation of these metrics appears plausible and a potentially worthwhile savings in time and cost for implementing agencies, but this replacement cannot be implemented unless the original data for water depth are preserved (and with the full precision of the original recorded data). With these caveats, however, there is every reason to expect that hydrologic metrics based on stage will prove as or more useful (and less costly), at least in the context of status-and-trends monitoring of urban watersheds, as those based on subsequent calculations of discharge.

ACKNOWLEDGMENTS

We thank King County, and particularly Jeff Burkey with King County and the Hydrologic Information Center (http://green2.kingcounty.gov/hydrology/) for assistance in the acquisition and overall maintenance of the gage database that informed this work. The analyses were developed in part via contract 15-05 between Stillwater Sciences, Inc. and the Lower Columbia Fish Recovery Board, with funds provided by the Washington Department of Ecology; additional support to DBB was provided under the Ecosystem Management and Restoration Research Program through Cooperative Agreement W912HZ-12-2-0016 between the U.S. Army Corps of Engineers and UC Santa Barbara. We were most fortunate to receive three excellent anonymous journal reviews plus the comments of Associate Editor Ted Endreny, which have greatly improved the manuscript in its final form.

ORCID

Derek B. Booth http://orcid.org/0000-0002-5242-4089
Christopher P. Konrad http://orcid.org/0000-0002-7354-547X

REFERENCES

Ahiablame, L. M., Engel, B. A., & Chaubey, I. (2012). Effectiveness of low impact development practices: Literature review and suggestions for future research. *Water, Air, and Soil Pollution*, *223*, 4253–4273.

Baker, D. B., Richards, R. P., Loftus, T. T., & Kramer, J. W. (2004). A new flashiness index: Characteristics and applications to Midwestern rivers and streams. *Journal of the American Water Resources Association*, *40*, 503–522.

Booth, D. B. (1991). Urbanization and the natural drainage system—Impacts, solutions, and prognoses. *Northwest Environmental Journal*, *7*, 93–118.

Booth, D. B., & Jackson, C. R. (1997). Urbanization of aquatic systems—Degradation thresholds, stormwater detention, and limits of mitigation. *Journal of the American Water Resources Association*, *33*, 1077–1090.

Booth, D. B., Roy, A. H., Smith, B., & Kapps, K. A. (2016). Global perspectives on the urban stream syndrome. Freshwater. *Science*, *35*, 412–420.

DeGasperi, C. L., Berge, H. B., Whiting, K. R., Burkey, J. J., Cassin, J. L., & Fuerstenberg, R. R. (2009). Linking hydrologic alteration to biological impairment in urbanizing streams of the Puget Lowland, Washington, USA. *Journal of the American Water Resources Association*, *45*, 512–533.

Fanelli, R., Prestegaard, K., & Palmer, P. (2017). Evaluation of infiltration-based stormwater management to restore hydrological processes in urban headwater streams. *Hydrological Processes*, *31*, 3306–3319.

Farquharson, F. A. K., Meigh, J. R., & Sutcliffe, J. V. (1992). Regional flood frequency analysis in arid and semi-arid areas. *Journal of Hydrology*, *138*, 487–501.

Hawley, R. J., & Bledsoe, B. P. (2011). How do flow peaks and durations change in suburbanizing semi-arid watersheds? *Journal of Hydrology*, *405*, 69–82.

Hawley, R. J., Wooten, M. S., MacMannis, K. R., & Fet, E. V. (2016). When do macroinvertebrate communities of reference streams resemble urban streams? The biological relevance of $Q_{critical}$. *Freshwater Science*, *35*, 778–794.

Hollis, G. E. (1975). The effect of urbanization on floods of different recurrence interval. *Water Resources Research*, *11*, 431–435.

Homer, C. G., Dewitz, J. A., Yang, L., Jin, S., Danielson, P., Xian, G., … Megown, K. (2015). Completion of the 2011 National Land Cover Database for the conterminous United States—Representing a decade of land cover change information. *Photogrammetric Engineering and Remote Sensing*, *81*, 345–354.

Karr, J. R. (1998). Rivers as sentinels: Using the biology of rivers to guide landscape management. In R. J. Naiman, & R. E. Bilby (Eds.), *River ecology and management: Lessons from the Pacific coastal ecosystems* (pp. 502–528). New York: Springer, New York.

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. U.S. *Geological Survey Scientific Investigations Report*, 2007–5206.

King County (1990). *Soos Creek Basin plan and final environmental impact statement* (pp. 330). Seattle: Department of Public Works, Surface Water Management Division.

King County (2012). *Stormwater retrofit analysis for Juanita Creek Basin in the Lake Washington watershed*. Prepared by Burkey. J., Wilgus, M., Berge, H. Water and Land Resources Division, Seattle, Washington: King County Department of Natural Resources and Parks.

Konrad, C. P., & Booth, D. B. (2002). Hydrologic trends resulting from urban development in western Washington streams. US Geological Survey Water-Resources Investigation Report 02-4040, pp. 40.

Konrad, C. P., & Booth, D. B. (2005). Hydrologic changes in urban streams and their ecological significance. *American Fisheries Society Symposium*, *47*, 157–177.

Konrad, C. P., & Voss, F. D. (2012). Analysis of streamflow-gaging network for monitoring stormwater in small streams in the Puget Sound Basin, Washington. US Geological Survey Scientific Investigations Report 2012-5020, pp. 16.

Leopold, L. B. (1968). Hydrology for urban planning: A guidebook on the hydrologic effects of urban land use. US Geological Survey. *Circular*, *554*, 18.

Leopold, L. B., & Maddock, T. (1953). The hydraulic geometry of stream channels and some physiographic implications. US Geological Survey Professional Paper 252, 57 pp.

Lewin, J. (1989). Floods in fluvial geomorphology. In K. Beven, & P. Carling (Eds.), *Floods: Hydrological, sedimentological and geomorphological implications* (pp. 265–284). Chichester, UK: John Wiley and Sons.

MacRae, C. R. (1997). Experience from morphological research on Canadian streams: Is control of the two-year frequency runoff event the best basis for stream channel protection? pp. 144–162. In L. A. Roesner (Ed.), *Effects of watershed development and management on aquatic ecosystems. Proceedings of an ASCE Engineering Foundation Conference, Snowbird, Utah, August 4–9, 1996*. Reston, Virginia, USA: American Society of Civil Engineers.

Meyer, J. L., Strayer, D. L., Wallace, J. B., Eggert, S. L., Helfman, G. S., & Leonard, N. E. (2007). The contribution of headwater streams to biodiversity in river networks. *Journal of the American Water Resources Association*, *43*, 86–103.

Peterson, B. J., Wollheim, W. M., Mulholland, P. J., Webster, J. R., Meyer, J. L., Tank, J. L., … McDowell, W. H. (2001). Control of nitrogen export from watersheds by headwater streams. *Science*, *292*, 86–90.

R Core Team (2016). *R: A language and environment for statistical computing, version 3.3.1 for Mac OS X*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.r-project.org/

Richter, B. D., Baumgartner, J. V., Powell, J., & Braun, D. P. (2012). A method for assessing hydrologic alteration within ecosystems. *Conservation Biology*, *10*, 1163–1174.

Rosenberg, E. A., Keys, P. W., Booth, D. B., Hartley, D., Burkey, J., Steinemann, A. C., & Lettenmaier, D. P. (2010). Precipitation extremes and the impacts of climate change on stormwater infrastructure in Washington state. *Climatic Change*, *102*, 319–349.

Roy, A. H., Rhea, L. K., Mayer, A. L., Shuster, W. D., Beaulieu, J. J., Hopton, M. E., … St Amand, A. (2014). How much is enough? Minimal responses of water quality and stream biota to partial retrofit stormwater management in a suburban neighborhood. *PLoS One*, *9*, e85011.

Sauer, V. B., Thomas, W.O. Jr, Strickler, V. A., & Wilson, K. V. (1983). Flood characteristics of urban watersheds in the United States. US Geological Survey Water Supply Paper 2207, 63 pp.

Vogel, R. M., & Fennessey, N. M. (1995). Flow duration curves II: A review of applications in water resources planning. *Journal of the American Water Resources Association*, *31*, 1029–1039.

Washington Department of Ecology (2012). National Pollutant Discharge Elimination System and state waste discharge general permit for discharges from large and medium municipal separate storm sewer systems. State of Washington, Department of Ecology, Olympia, Washington 98504-7600, Issue Date August 1, 2012, 75 pp. Retrieved from http://www.ecy.wa.gov/PROgrams/wq/stormwater/municipal/phaseIpermit/phipermit.html (accessed 2/19/2017).

How to cite this article: Booth DB, Konrad CP. Hydrologic metrics for status-and-trends monitoring in urban and urbanizing watersheds. *Hydrological Processes*. 2017;1–13. https://doi.org/10.1002/hyp.11369