Journal of Hydrology 564 (2018) 758–772

Contents lists available at ScienceDirect

Journal of Hydrology

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

Research papers

Evaluation of the impacts of hydrologic model calibration methods on predictability of ecologically-relevant hydrologic indices

J. Sebastian Hernandez-Suarez[a], A. Pouyan Nejadhashemi[a,*], Ian M. Kropp[a], Mohammad Abouali[a], Zhen Zhang[b], Kalyanmoy Deb[c]

[a] Department of Biosystems and Agricultural Engineering, Michigan State University, East Lansing, MI 48824, USA
[b] Physical Sciences Division, Department of Statistics, University of Chicago, Chicago, IL 60637, USA
[c] Department of Computer Science and Engineering, Michigan State University, East Lansing, MI 48824, USA

Abbreviations: AIC, Akaike information criterion; ALPHA_BF, baseflow alpha factor (days^{-1}); AR (1), autoregressive model with lag 1; BIOMIX, biological mixing efficiency; CANMX, maximum canopy storage (mm H2O); CAR (1), continuous autoregressive model with lag 1; CH_K (2), effective hydraulic conductivity in main channel alluvium (mm h^{-1}); CH_N (2), Manning's "n" value for the main channel; CN2, initial SCS runoff number for moisture condition II; DH15, High flow pulse duration with a threshold equal to the 75th percentile of the entire flow record (days); COIN, computational optimization and innovation laboratory at Michigan State University; DH17, high flow duration with a threshold equal to the median flow (days); DH19, high flow duration with a threshold equal to 7 times the median flow (days); DH20, high flow duration with a threshold equal to the 75th percentile value for the median annual flows (days); DH21, high flow duration with a threshold equal to the 25th percentile value for the median annual flows (days); DH23, flood duration with a threshold equal to the flow equivalent for a flood recurrence of 1.67 years (days); DL1, annual minimum daily flow (m^3 s^{-1}); DL2, annual minimum of 3-day moving average flow (m^3 s^{-1}); DL3, annual minima of 7-day means of daily discharge (m^3 s^{-1}); DL4, annual minima of 30-day means of daily discharge (m^3 s^{-1}); DL5, annual minima of 90-day means of daily discharge (m^3 s^{-1}); DL6, variability of annual minimum daily average flow; DL7, variability of annual minimum of 3-day moving average flow; DL8, variability of annual minimum of 7-day moving average flow; DL11, mean of 1-day minima of daily discharge; DL12, mean of 3-day minima of daily discharge; DL16, low flow pulse duration (days); EPA, environmental protection agency; EPCO, plant uptake compensation factor; ESCO, soil evaporation compensation factor; FDC, flow duration curve; FH1, high flood pulse count with a threshold equal to the 25th percentile of the entire flow record (year^{-1}); FH4, high flood pulse count with a threshold equal to 7 times median daily flow (year^{-1}); FH5, flood frequency with a threshold equal to the median flow (year^{-1}); FH8, flood frequency with a threshold equal to the 25th percentile of the entire flow record (year^{-1}); FH9, flood frequency with a threshold equal to the 75th percentile of the entire flow record (year^{-1}); FL1, low flood pulse count (year^{-1}); GLS, generalized least-square; GW_DELAY, groundwater delay time (days); GWQMN, threshold depth of water in the shallow aquifer required for return flow to occur (mm H_2O); GW_REVAP, groundwater "revap" coefficient; HRU, hydrologic response units; MA12, mean monthly flow for January (m^3 s^{-1}); MA13, mean monthly flow for February (m^3 s^{-1}); MA14, mean monthly flow for March (m^3 s^{-1}); MA15, mean monthly flow for April (m^3 s^{-1}); MA16, mean monthly flow for May (m^3 s^{-1}); MA17, mean monthly flow for June (m^3 s^{-1}); MA18, mean monthly flow for July (m^3 s^{-1}); MA19, mean monthly flow for August (m^3 s^{-1}); MA20, mean monthly flow for September (m^3 s^{-1}); MA21, mean monthly flow for October (m^3 s^{-1}); MA22, mean monthly flow for November (m^3 s^{-1}); MA23, mean monthly flow for December (m^3 s^{-1}); MA29, variability in June flows; MA30, variability in July flows; MA31, variability in August flows; MA32, variability in September flows; MA33, variability in October flows; MA34, variability in November flows; MA42, variability across annual flows; MA44, variability across annual flows; MA45, skewness in annual flows; MH6, mean maximum June monthly flow (m^3 s^{-1}); MH7, mean maximum July monthly flow (m^3 s^{-1}); MH10, mean maximum October monthly flow (m^3 s^{-1}); MH11, mean maximum November monthly flow (m^3 s^{-1}); MH21, high flow volume (days); MH22, high flow volume (days); MH23, high flow volume (days); MHIT, MATLAB hydrological index tool; ML7, mean minimum July monthly flow (m^3 s^{-1}); ML8, mean minimum August monthly flow (m^3 s^{-1}); ML9, mean minimum September monthly flow (m^3 s^{-1}); ML14, mean of annual minimum flows; ML15, low flow index; ML16, median of annual minimum flows; ML17, baseflow index based on the seven-day minimum flow; ML19, baseflow index based on the lowest annual daily flow; ML21, variability across annual minimum flows; ML22, specific mean annual minimum flows (m^3 s^{-1} km^{-2}); MOEA, multi-objective evolutionary algorithm; NRCS, natural resources conservation service; NASS, national agricultural statistics service; NCDC, national climatic data center; NED, national elevation dataset; NOAA, national oceanic and atmospheric administration; NSE, Nash-Sutcliffe efficiency; NSErel, relative NSE; NSEsqrt, root-squared-transformed NSE; NSGA-II, non-dominated sorted genetic algorithm II; NSGA-III, nondominated sorted genetic algorithm III; OF, objective function; Q25, flow exceeded 25% of the time; Q75, flow exceeded 75% of the time; RA1, rise rate (m^3 s^{-1} d^{-1}); RA3, fall rate (m^3 s^{-1} d^{-1}); RA6, change of flow – increasing (m^3 s^{-1}); RA7, change of flow – decreasing (m^3 s^{-1}); RA8, reversals (year^{-1}); REVAPMN, threshold depth of water in the shallow aquifer for "revap" or percolation to the deep aquifer to occur (mm H_2O); RCHRG_DP, deep aquifer percolation fraction; RMSE, root-mean-square error; SBX, simulated binary crossover; SOL_AWC, available water capacity of the soil layer (mm H_2O mm^{-1} soil); SSURGO, soil survey geographic database; SURLAG, surface runoff lag coefficient; SWAT, soil and water assessment tool; TH1, Julian date of annual maximum; TL1, Julian date of annual minimum; TL4, seasonal predictability of non-low flow; USDA, United States department of agriculture; USGS, United States geological survey

* Corresponding author at: Department of Biosystems & Agricultural Engineering, Farrall Agriculture Engineering Hall, 524 S. Shaw Lane, Room 225, Michigan State University, East Lansing, MI 48824, USA.
E-mail address: pouyan@msu.edu (A.P. Nejadhashemi).

https://doi.org/10.1016/j.jhydrol.2018.07.056
Received 9 January 2018; Received in revised form 19 July 2018; Accepted 21 July 2018
Available online 24 July 2018

A R T I C L E I N F O

This manuscript was handled by G. Syme, Editor-in-Chief, with the assistance of Xiaodong Zhang, Associate Editor

Keywords:
Hydrologic indicators
Environmental flows
Multi-objective calibration
Watershed modeling
NSGA-III

A B S T R A C T

Due to the significant costs associated with stream health monitoring, hydrological modeling is widely used to calculate ecologically-relevant hydrologic indices to better understand the overall condition of streams within large and diverse watersheds. However, hydrologic modeling's ability to replicate these indices is limited, especially when calibrating models by optimizing a single objective function or when selecting a single optimal solution. Hence, this study evaluates the performance of multi-objective model calibration in representing 167 hydrologic indices of ecological interest using the median values of different Pareto-optimal solution sets. For this purpose, two strategies based on three Nash-Sutcliffe Efficiencies (NSE), and root-mean-squared error (RMSE) with explicit hydrograph partitioning, were implemented. Additionally, the *k*-means clustering technique was employed to define subsets of Pareto-optimal solutions representing high, medium, and low flow conditions. The Soil and Water Assessment Tool (SWAT) was set up for the Honeyoey Creek-Pine Creek Watershed, located in Michigan, USA, and was calibrated using a single streamflow gauging station and the recently proposed multi-objective optimization algorithm – Non-dominated Sorting Genetic Algorithm III (NSGA-III). The results show that the NSE- and RMSE-based calibration strategies can represent 128 and 123 indices within a range of ± 30% relative errors, respectively. In addition, the results of this study demonstrated that using different set of solutions, instead of a single optimal solution, introduces more flexibility in the predictability of different hydrologic indices of ecological interest. Moreover, it is suggested that a multi-objective calibration approach allows the systematic identification of groups of poorly represented and closely related indices. Knowing these poorly represented indices would facilitate the formulation of additional objective functions intended to improve model performance or to detect structural inadequacies.

1. Introduction

Alteration of rivers (e.g. regime, temperature and water quality) driven by human interventions and changing environmental conditions are threatening water security and freshwater biodiversity around the world (Bunn and Arthington, 2002; Carpenter et al., 2011; Dudgeon et al., 2006; Hipsey et al., 2015; Vörösmarty et al., 2010). Traditionally, stream condition evaluation has used chemical and microbiological constituents as criteria (Karr and Yoder, 2004). However, the lack of holistic approaches resulted in further degradation of aquatic ecosystems (Hering et al., 2010; Jelks et al., 2008). To overcome this issue, biological assessments were introduced to provide additional insight into the overall ecological integrity of streams (US EPA, 2011; Woznicki et al., 2016a), and therefore, can be used for environmental management and decision-making.

Biological assessments measure the biota (e.g. fish, benthic macroinvertebrates, periphyton, amphibians) within a stream to obtain information regarding its biological integrity (US EPA, 2011). In this context, biological integrity is the capacity to support and maintain a "balanced, integrated, and adaptive" biological system within the expected structure and function of the natural habitat of a particular region (Karr, 1996; Karr and Dudley, 1981). Stream health integrates the physical, chemical and biological integrity of a stream, which supports living systems that are necessary for human well-being (Karr, 1999; Maddock, 1999). Stream health indices are generally based on individual characteristics comprised of species abundance and condition, species richness and composition, or trophic composition (Herman and Nejadhashemi, 2015). However, due to limited economic resources, it is not possible to measure these characteristics for all streams within a watershed. Therefore, stream health evaluation based on field data is very limited. To address this difficulty, several modeling approaches have been introduced to extend the available information to ungauged locations (Woznicki et al., 2015).

Streamflow regime has been recognized as a key determinant for sustaining biodiversity and ecological integrity. Thus, ecologically-relevant hydrologic indices are often used as predictors in stream health models besides landscape factors and water quality indicators (Poff and Zimmerman, 2010; Woznicki et al., 2016a). Prediction of ecologically-relevant hydrologic indices include the use of regional statistic approaches (Carlisle et al., 2009; Dhungel et al., 2016; Knight et al., 2012; Patrick and Yuan, 2017; Sanborn and Bledsoe, 2006; Yang et al., 2016), and hydrological modeling (Caldwell et al., 2015; Kennen et al., 2008; Kiesel et al., 2017; Olsen et al., 2013; Vis et al., 2015; Wenger et al., 2010; You et al., 2014). The use of hydrological models is especially preferred when it is necessary to evaluate the change of stream health driven by modifications in land use, environmental conditions or management practices (Poff et al., 2010; Shrestha et al., 2016; Woznicki et al., 2016b). However, hydrologic models' ability to replicate ecologically-relevant indices is limited (Vigiak et al., 2018). Some studies have shown that the use of typical calibration approaches (i.e. single-objective based on widely used performance metrics) produces poor representations of some streamflow regime characteristics (Murphy et al., 2013; Vis et al., 2015). For instance, while average conditions are generally well-predicted, low- and high-flow indices are frequently over or under predicted (Wenger et al., 2010). Moreover, no model has been found to provide all selected ecologically-relevant hydrologic indices within ± 30% of the observed values (Caldwell et al., 2015; Vis et al., 2015). Therefore, several studies have proposed to explicitly include ecologically-relevant hydrologic indices into the objective functions for model calibration to improve the overall performance of streamflow regime simulations (Murphy et al., 2013; Shrestha et al., 2014; Vis et al., 2015). For instance, Kiesel et al. (2017) and Zhang et al. (2016) used multi-metric (i.e. aggregated) objective functions based on a reduced number of ecologically-relevant hydrological indices (12 and 16 indices, respectively). They found that it is possible to obtain better overall representations of streamflow regime compared to objective functions based only on conventional performance metrics (e.g. coefficient of efficiency, mean squared errors, correlation coefficient). However, optimal solutions were still unable to effectively represent all hydrological indices individually, especially when they are not explicitly included in the objective function formulation (Kiesel et al., 2017). In addition, optimal solutions are sensitive to the weights assigned to each ecological-relevant hydrological index in the multi-metric objective function (Zhang et al., 2016). Furthermore, different authors have suggested the use of typical performance metrics with transformations (e.g. square root, reciprocal, or normalization of streamflow values) (Garcia et al., 2017; Oudin et al., 2006; Pushpalatha et al., 2012) or explicit hydrograph partitioning (Pfannerstill et al., 2014) for model calibration. However, these approaches have mainly shown improvements in the representation of target flow conditions rather than the overall streamflow regime, or have been evaluated using traditional performance metrics instead of ecologically-relevant hydrological indices.

Furthermore, the aforementioned studies have mainly implemented single-objective algorithms for model calibration. Therefore, these previous studies present no integrated perspectives on relationships between different performance metrics or hydrological indices involved in the model calibration process. Meanwhile, multi-objective optimization algorithms are very useful for evaluating the tradeoffs between different metrics and objective functions involved in hydrologic model calibration (Price et al., 2012), and can provide sets of solutions able to represent different flow conditions (Efstratiadis and Koutsoyiannis, 2010; Reed et al., 2013). However, Pareto-optimal solutions are not usually evaluated employing ecologically-relevant hydrological indices, but instead pure hydrological signatures based on, for example, Flow Duration Curve (FDC) segments or runoff ratios (Shafii and Tolson, 2015; van Werkhoven et al., 2009). Moreover, many studies have been more concerned about selecting a best single solution than analyzing the whole set of Pareto-optimal solutions, which could provide better results for overall streamflow regime representation. For instance, Vis et al. (2015) attributed high model efficiencies when using the median of several optimum solutions as a “more robust prediction” for ecologically-relevant hydrologic indicators. Therefore, the objective of this study is to identify an objective function best suited for stream health model applications using a multi-objective optimization algorithm for model calibration. Typical performance metrics that represent different flow conditions and explicit hydrograph partitioning are considered in this study. For this purpose, the Soil and Water Assessment Tool (SWAT) watershed model and the NSGA-III multi-objective optimization algorithm are jointly implemented. Then, a set of 167 ecologically-relevant hydrologic indices is used for evaluating the ability of the resulting Pareto-optimal solutions in representing the overall streamflow regime.

2. Materials and methods

Two strategies based on a multi-objective optimization technique for model calibration were compared to evaluate their abilities to predict ecologically-relevant hydrologic indices (Fig. 1). The first multi-objective strategy calibrates the model based on three different forms of Nash-Sutcliffe Efficiency (NSE) described by Krause et al. (2005) and Pushpalatha et al. (2012) that are suitable for evaluating high, medium, and low flows. In the second strategy, observed daily flow time series were divided into three categories (high, medium, and low flows) using explicit thresholds for low and high flow (flows exceeded 75% and 25% of the time, respectively). For each category, an objective function based on the root-mean-square error (RMSE) was computed. Pareto-optimal solutions for calibrated model parameters were obtained for each multi-objective strategy employing the NSGA-III algorithm (Deb and Jain, 2014; Jain and Deb, 2014).

For both strategies, the Soil and Water Assessment Tool (SWAT) (Arnold et al., 2012; Neitsch et al., 2011) was used to simulate daily streamflow discharge time series for every stream segment, and the MATLAB Hydrological Index Tool (Abouali et al., 2016) was employed to calculate 167 hydrologic indices intended to characterize streamflow regime (Olden and Poff, 2003). Hydrologic indices were computed for each Pareto-optimal point obtained from both multi-objective strategies and for the observed flow dataset. Then, model outputs and indices were evaluated with respect to the observed values using statistical analysis. For this purpose, Pareto-optimal points for each multi-objective strategy were clustered into three groups using the *k*-means method (Arthur and Vassilvitskii, 2007), while Generalized Least-Square (GLS) estimation was used to analyze the differences between the simulated and observed mean streamflow values. The median errors between the hydrologic indices based on Pareto-optimal solutions and observed time series were evaluated with respect to the ± 30% uncertainty bound for the observed values, as reported by previous studies (Caldwell et al., 2015; Kennard et al., 2009; Vis et al., 2015). Results were compared with the optimal solutions from a single-objective approach in which the objective functions are based on NSE as recommended by Moriasi

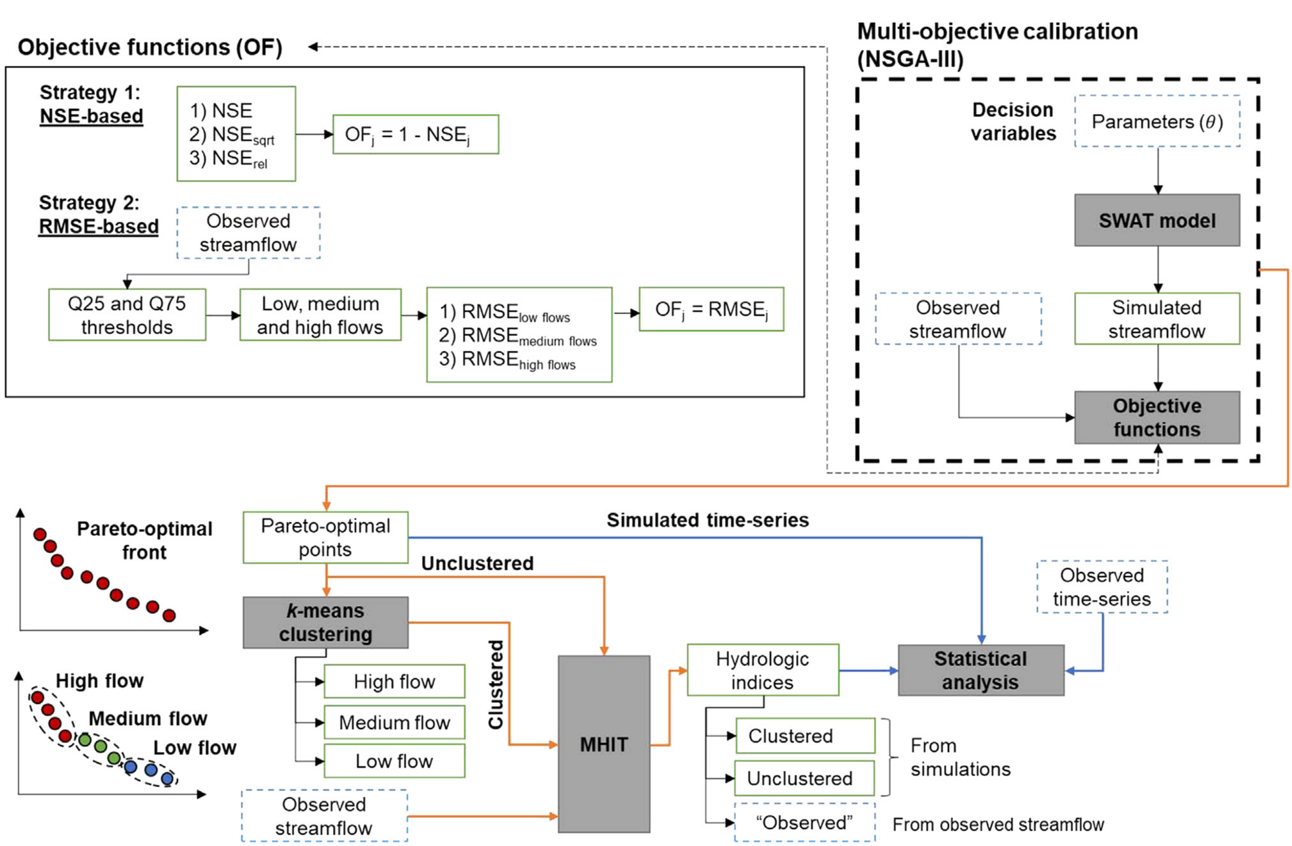

Fig. 1. A schematic diagram presenting the overall multi-objective model calibration and evaluation process. Q25 and Q75 are the flows exceeded 25% and 75% of the time, respectively, NSE is the standard Nash-Sutcliffe Efficiency, NSE$_{sqrt}$ is the root-squared-transformed NSE, NSE$_{rel}$ is the relative NSE, RMSE is the Root-Mean-Square Error, and MHIT is the MATLAB Hydrological Index Tool.

et al. (2015) for evaluation of hydrological models. Moreover, given that both NSE and RMSE are computed using the sum of the square differences between observed and simulated values, and NSE is non-dimensional, only NSE was considered for the single-objective model calibration.

2.1. Study area

In order to perform environmental flow analysis, it is desirable to identify areas where urbanization is limited, streamflow is not regulated or its alteration is negligible, and observed discharge records are available for almost all the studied period (Olden and Poff, 2003). The Honeyoey Creek–Pine Creek Watershed, with a drainage area of 1010 km^2, meets all the criteria, because urbanization is less than 4%, streamflow regulation is limited, and observed streamflow data for the period is complete. The study area is located in the Saginaw Bay Watershed in east-central Michigan (Fig. 2), which is the largest watershed in the state and is identified as an area of concern by the US Environmental Protection Agency due to pollution from sediment and nutrients that resulted in eutrophication, fish habitat degradation, and loss of recreational values, among others (EPA, 2015). The watershed has an average slope of 1.9% ranging from 12–39% in the headwaters to 0–1.4% in the lowlands (NED, 2015). The region has a temperate climate with distinct seasons (Andresen and Winkler, 2009). The average annual rainfall is about 840 mm for the period 1981–2010 (NOAA, 2017). However, the precipitation regime is bimodal, with maxima in May and September, and minima in February and July. Mean annual air temperature in the watershed is 9 °C, with a minimum monthly temperature of −9 °C in January and a maximum monthly temperature of 28 °C in July. The dominant land use is agriculture, covering about 50% of the watershed, followed by forests (24%), wetlands (16%) and pasturelands (7%) (NASS, 2012). Over 60% of the river network's riparian vegetation has not been altered by human activities. The dominant soil textures are loamy sand, sandy loam, loam/clay loam, and sand, which cover about 30, 26, 20 and 11% of the study area, respectively (NRCS, 2014). The average flow is about 10.3 m^3/s at the outlet of the watershed. High flows occur between March and May as a result of snow melting and high precipitation, while low flows occur between July and October, during summer and fall seasons (USGS, 2017). High flows, considered in this study as those that are exceeded at most 25% of the time (Q25), have magnitudes above 11.3 m^3/s while low flows are defined as those with values below 3.9 m^3/s, which is the discharge exceeded 75% of the time (Q75).

2.2. Data collection

Datasets required for the hydrologic modeling comprise topography, land use, soil properties, climate, and observed streamflow discharge. The National Elevation Dataset from the US Geological Survey (USGS) with 30 m spatial resolution was used to represent the topography of the watershed (NED, 2015). The land use characteristics were obtained from the Cropland Data Layer developed by the National Agricultural Statistics Service of the US Department of Agriculture (USDA-NASS) with 30 m spatial resolution (NASS, 2012). The soil properties were compiled from the Natural Resources Conservation Service's (NRCS) Soil Survey Geographic (SSURGO) Database, at a scale of 1: 250,000 (NRCS, 2014). Daily time series for precipitation and temperature from 2001 through 2014 were obtained from two weather stations that belong to the National Climatic Data Center (NOAA, 2017). Relative humidity, solar radiation and wind speed time series for the same time span were provided by the stochastic weather generator WXGEN (Neitsch et al., 2011) included in SWAT. Daily streamflow discharges between 2003 and 2014 were obtained from the Pine River Near Midland gauging station (ID 04155500) (USGS, 2017).

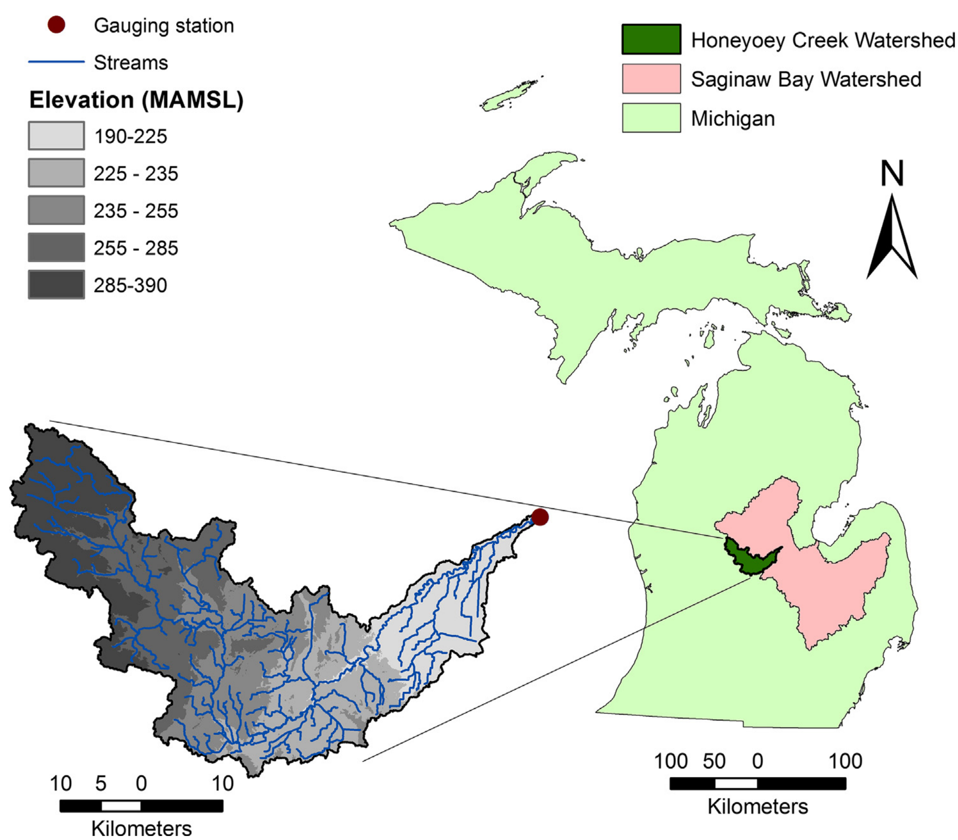

Fig. 2. Location and topography of the study area.

2.3. SWAT model description

The Soil and Water Assessment Tool (SWAT version 2012, rev. 614) is a semi-distributed, continuous-time, process-based hydrological model, which simulates water flow, sediment transport, and water quality processes in watersheds (Arnold et al., 1998). SWAT divides a watershed into subwatersheds that are further discretized into multiple units with homogeneous land use, slope, and soil characteristics known as hydrologic response units (HRU). The main processes in SWAT include snow accumulation and melting, evapotranspiration, infiltration, percolation losses, surface runoff, channel routing, and groundwater flows (Neitsch et al., 2011).

SWAT is used in this study for daily flow simulation between 2003 and 2014 for all 749 defined stream segments in the Honeyoey Creek–Pine Creek watershed. Fifteen parameters were selected for model calibration whose description and ranges of variation are presented in Table 1. The calibration period was defined between 2003 and 2008, while the validation period spans between 2009 and 2014. Meanwhile, two years of warm-up period (2001–2002) were considered to stabilize initial conditions of soil water (Cibin et al., 2010).

2.4. Hydrologic indices

The 171 hydrologic indices reported by Olden and Poff (2003) were evaluated for all Pareto-optimal solutions after completing each multi-objective model calibration. For this study, we discarded four indices related to the frequency and duration of zero-flow days and low flow spells, which were all equal to zero in this case. Then, the remaining 167 hydrologic indices were compared with the respective indices for the observed dataset, including the calibration and validation periods. The evaluated hydrologic indices characterize the streamflow regime in terms of magnitude, frequency, duration, timing and rate of change of flows (Poff et al., 1997; Richter et al., 1996) for a given daily time-series. These indices were classified into eleven groups: magnitude for low (ML), average (MA), and high (MH) flow conditions; frequency for low (FL), and high (FH) flow conditions; duration for low (DL), and high (DH) flow conditions; timing for low (TL), average (TA), and high (TH) flow conditions; and rate of change for average (RA) flow conditions. The hydrologic indices were computed using the MATLAB Hydrological Index Tool (MHIT), which has shown better computing performances in comparison to other available packages when handling high number of datasets (Abouali et al., 2016).

2.5. Objective functions

In this study, we contrast the ability of two commonly used procedures in representing a wide number of the streamflow metrics related to environmental flows indicated in Section 2.4. Each procedure refers to a specific three-dimensional objective space (Fig. 3). Further details are presented next.

2.5.1. Nash-Sutcliffe efficiency-based objective functions

In this strategy, NSE-based objective functions are formulated to represent different parts of the observed hydrograph. Krause et al. (2005) and Pushpalatha et al. (2012) indicated that standard NSE, Eq. (1) (Nash and Sutcliffe, 1970) is very sensitive to high flows on continuous simulations, given that the differences between simulated and observed values are squared. Meanwhile, NSE calculated on root squared transformed flows (Eq. (2), NSE_{sqrt}) has been found to provide a more balanced performance because the errors are more equally distributed on high and low flow portions of the hydrograph (Oudin et al., 2006; Pushpalatha et al., 2012). Additionally, relative NSE (Eq. (3), NSE_{rel}), described by Krause et al. (2005), suppresses the influence of peak flows on the efficiency computation, making it more sensitive to low flows. Pushpalatha et al. (2012) showed that NSE computed on the reciprocal of flow values (inverse transformed flows) is better suited for low flow conditions, focusing on the 20% lowest flows on average. However, in this study we decided to use the NSE_{rel} given that some of the ecologically-relevant hydrologic indices (e.g. low flow index, base flow indices, indices based on moving averages) computed by MHIT for low flows are based also on overall flow values. Therefore, NSE, NSE_{sqrt}, and NSE_{rel} were used to represent high, medium, and low flows, respectively.

$$NSE = 1 - \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\sum_{i=1}^{n}(O_i - \bar{O})^2} \tag{1}$$

Table 1
Calibrated ranges obtained with Pareto-optimal solutions. Values without brackets correspond to NSE-based strategy results while values within brackets correspond to RMSE-based strategy results.

Parameter**	Initial range	Calibrated ranges per cluster			
		All solutions	High Flow	Medium Flow	Low flow
BIOMIX	0–1	0–0.97 [0.02–0.99]	0–0.97 [0.02–0.96]	0–0.63 [0.03–0.99]	0.04–0.87 [0.07–0.99]
CN2*	(−0.25)–0.25	(−0.25)–0.25 [(−0.25)–0.25]	(−0.25)–(−0.22) [(−0.25)–0.23]	(−0.25)–(−0.21) [(−0.25)–0.25]	0.246–0.249 [(−0.24)–0.25]
CANMX	0–100	10–100 [1.4–91]	10–68 [4.7–78]	11–95 [20–91]	36–100 [1.4–70]
ESCO	0.01–1	0.74–1 [0.6–1]	0.85–0.96 [0.6–1]	0.74–0.91 [0.9–1]	0.91–1 [0.9–1]
EPCO	0.01–1	0.01–0.9 [0.01–0.9]	0.01–0.79 [0.01–0.8]	0.01–0.9 [0.01–0.9]	0.09–0.47 [0.1–0.9]
GW_DELAY	0–500	0–309 [0–499]	0–0.1 [0.01–0.34]	0–0 [0–415]	237–309 [141–499]
ALPHA_BF	0–1	0.05–0.29 [0.05–0.33]	0.23–0.29 [0.13–0.32]	0.19–0.28 [0.1–0.33]	0.05–0.11 [0.05–0.24]
GWQMN	0–5000	0–4861 [38–2018]	2–154 [38–636]	0–614 [479–2018]	1221–4861 [331–649]
GW_REVAP	0.02–0.2	0.02–0.2 [0.03–0.2]	0.02–0.2 [0.03–0.2]	0.03–0.19 [0.1–0.16]	0.11–0.2 [0.12–0.17]
REVAPMN	0–1000	48–959 [0.15–840]	50–953 [0.15–840]	48–959 [0.88–550]	60–378 [16.41–450]
RCHRG_DP	0–1	0.28–0.63 [0.28–0.75]	0.52–0.63 [0.28–0.64]	0.43–0.55 [0.3–0.45]	0.28–0.41 [0.3–0.75]
CH_N (2)	0.001–0.3	0.03–0.23 [0.02–0.3]	0.03–0.04 [0.02–0.04]	0.03–0.04 [0.02–0.08]	0.13–0.23 [0.05–0.3]
CH_K (2)	0–500	10–34 [12–57]	23–31 [21–51]	22–34 [28–57]	10–21 [12–52]
SOL_AWC*	(−0.25)–0.25	(−0.25)–0.25 [(−0.25)–0.23]	0.02–0.25 [(−0.19)–0.23]	0.11–0.25 [(−0.25)–0.16]	(−0.25)–(−0.13) [(−0.22)–0.23]
SURLAG	1–24	1–1.4 [1–19.2]	1–1.1 [1–1.2]	1–1.2 [1–13.4]	1.1–1.4 [1–19.2]

* These parameters are treated as global multiplying factors that modify the assigned values for each HRU depending on soil type and land use.

** ALPHA_BF, Baseflow alpha factor ($days^{-1}$); BIOMIX, Biological mixing efficiency; CANMX, Maximum canopy storage (mm H_2O); CH_K (2), Effective hydraulic conductivity in main channel alluvium ($mm\,h^{-1}$); CH_N (2), Manning's "n" value for the main channel; CN2, Initial SCS runoff number for moisture condition II; EPCO, Plant uptake compensation factor; ESCO, Soil evaporation compensation facor; GW_DELAY, Groundwater delay time (days); GWQMN, Threshold depth of water in the shallow aquifer required for return flow to occur (mm H_2O); GW_REVAP, Groundwater "revap" coefficient; REVAPMN, Threshold depth of water in the shallow aquifer for "revap" or percolation to the deep aquifer to occur (mm H_2O); RCHRG_DP, Deep aquifer percolation fraction; SOL_AWC, Available water capacity of the soil layer (mm H_2O mm^{-1} soil); SURLAG, Surface runoff lag coefficient.

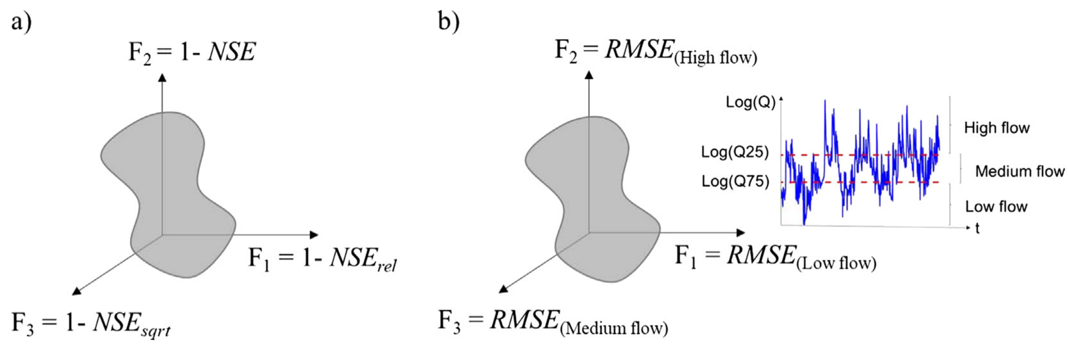

Fig. 3. Objective spaces for the SWAT model calibration: a) using different forms of Nash-Sutcliffe efficiency; b) after hydrograph partitioning using Q25 and Q75 thresholds for calculation of RMSE.

$$\mathrm{NSE_{sqrt}} = 1 - \frac{\sum_{i=1}^{n} (\sqrt{O_i} - \sqrt{P_i})^2}{\sum_{i=1}^{n} (\sqrt{O_i} - \sqrt{\bar{O}})^2} \tag{2}$$

$$\mathrm{NSE_{rel}} = 1 - \frac{\sum_{i=1}^{n} \left(\frac{O_i - P_i}{O_i}\right)^2}{\sum_{i=1}^{n} \left(\frac{O_i - \bar{O}}{\bar{O}}\right)^2} \tag{3}$$

where O and P are the observed and predicted values, respectively. For all NSE-based criteria, the objective functions (OF) were minimized by computing 1 – NSE, which have a range between zero and infinity. Values for any form of NSE range from minus infinity to one, while the corresponding OFs range from zero to infinity. A perfect fit between simulated and observed values is achieved when all NSE are equal to one and corresponding OFs are equal to zero.

2.5.2. Root-Mean-Square Error-based objective functions

In this strategy, RMSE-based objective functions are formulated for streamflow calibration. The time series for the entire study period (2003–2014) was divided into three categories representing high, medium, and low flows using the Q25 and Q75 thresholds obtained from the observed data. To have the same amount of observed and simulated points in each category, the simulated time series are divided following the observed time series partitioning. Then, the RMSE is computed for each flow category:

$$\mathrm{RMSE_j} = \sqrt{\frac{1}{n_j} \sum_{i=1}^{n_j} (O_i - P_i)^2} \tag{4}$$

where j refers to the flow category and n_j is the number of observations for each category. Each minimization OF is equal to the computed RMSE for each category. A perfect fit between observed and simulated values yield an RMSE equal to zero.

2.6. Multi-objective optimization algorithm

Multi-objective evolutionary algorithms (MOEAs) are population-based heuristic search methods that use randomly generated points that move towards an optimal Pareto front using evolutionary operators (Coello Coello et al., 2007). MOEAs have been widely implemented during the last two decades for water resources applications (Efstratiadis and Koutsoyiannis, 2010; Maier et al., 2014; Reed et al., 2013). For instance, the Non-dominated Sorting Genetic Algorithm II (NSGA-II) (Deb et al., 2002) has been widely-used for hydrologic model calibration (e.g. Bekele and Nicklow, 2007; Confesor and Whittaker, 2007; Lu et al., 2014; Shafii and De Smedt, 2009). The popularity of NSGA-II is mainly given by its simplicity, modularity, parameter-less property, and good performance for difficult two-objective problems (Deb, 2008). However, without any extensions or combinations with other approaches, the NSGA-II by itself has shown shortcomings for solving problems with three or more objectives (Deb and Jain, 2014; Reed et al., 2013; Sindhya et al., 2013).

NSGA-III, which is based on the NSGA-II framework, is the evolutionary multi-objective optimization algorithm used to implement the two multi-objective calibration strategies (Deb and Jain, 2014; Jain and Deb, 2014). NSGA-III is an elitist reference-point-based procedure that uses non-domination sorting to solve problems with four or more objectives. This procedure has also shown good performance solving cases with three objectives (Seada and Deb, 2016). The main difference between NSGA-II and NSGA-III is the niching method which is a procedure to maintain diversity among solutions (Deb, 2001). NSGA-II uses crowding distances, while NSGA-III is reference-directions-based (Deb and Jain, 2014; Jain and Deb, 2014). A reference direction is a line that crosses both the origin and a supplied reference point in the objective space. A detailed explanation of the algorithm can be found in Deb and Jain (2014) and Seada and Deb (2016). To the best of our knowledge, this is the first time that the NSGA-III algorithm is used for hydrologic model calibration.

NSGA-III's parameters are the population size (equal to the number of reference points, assigned here equal to 100), the number of generations (assigned here as 500), crossover and mutation probabilities (assigned as 0.9 and the reciprocal of the number of calibration parameters, respectively), and distribution indices for each genetic operation (i.e. Simulated Binary Crossover – SBX, and polynomial mutation, assigned as 10 and 20, respectively). The NSGA-III implementation used in this study was programmed in Java and was provided by the Computational Optimization and Innovation (COIN) Laboratory at Michigan State University. The Java code was adapted for this study to have a connection with SWAT to perform the automatic calibration process.

The convergence to the optimal Pareto front was evaluated using the hypervolume indicator, which is a measure of the volume enclosed by a Pareto front with respect to a specified reference point (Auger et al., 2009). The optimal Pareto front was selected when a steady behavior of the hypervolume indicator was observed across the preceding generations. In this study, the reference points for each strategy were selected with NSE = $\mathrm{NSE_{sqrt}}$ = $\mathrm{NSE_{rel}}$ = 0, and $\mathrm{RMSE_H}$ = $\mathrm{RMSE_M}$ = 30 m^3/s and $\mathrm{RMSE_L}$ = 10 m^3/s, which approximately correspond to the extreme objective function values visited by the optimization algorithm. The hypervolume indicator was computed for the Pareto front obtained at the end of each generation using the Walking Fish Group (WFG) algorithm (While et al., 2012, 2016), and the resulting values for each strategy were normalized to range between 0 and 1.

2.7. Model evaluation

In order to perform the statistical analysis for model evaluation, the Pareto-optimal points are clustered into three groups representing high, medium, and low flow conditions. Hence, the k-mean clustering method is employed for the two calibration strategies in order to identify separate sets of solutions that show better performances for each flow condition. The three clusters are identified for each calibration method, using the corresponding objective functions presented in Section 2.5. Thus, Pareto-optimal solutions with the highest NSE and NSE_{rel} values are going to be collected in the high and low flow clusters, respectively. Solutions with balanced NSE and NSE_{rel} values will comprise the medium flow cluster. Then, the simulated streamflow time series from each group are compared with the observed dataset, to evaluate whether the estimated mean differences between the simulations and observations are significant. To this end, the difference of each simulation with respect to the observed dataset is used as response to fit a simple regression with intercept using GLS estimation with Autoregressive model with lag 1, or AR (1), to account for the serial correlation for the time series. The differences are considered significant when the reported p-value is less than 0.05 (i.e. 95% confidence interval for the estimated mean of the distribution does not span zero), with positive (or negative) values for the difference indicating over/under-estimation of the actual observation. The process was repeated comparing high, medium, and low streamflow categories using the Q25 and Q75 thresholds defined for the RMSE-based calibration strategy. However, because the extracted time series for each flow category are irregularly spaced temporally, we modeled the difference from the corresponding samples between the observed series and each simulated series. Three different methods were used: Normal distribution, Student's t-distribution, and GLS with a Continuous Autoregressive model with lag 1, or CAR (1). Then, we determined the most appropriated test based on the smallest Akaike Information Criterion (AIC) value.

Finally, the hydrologic indices obtained for each Pareto-optimal solution were also grouped following the same three clusters determined above (high, medium, and low flow conditions). For each cluster and hydrologic index, the difference between the simulated and observed values were computed and divided by the respective observed values to obtain the relative error. Then, it was determined whether the median relative errors for each group were within or outside the ± 30% uncertainty bound. The procedure comprises of calculating all indices for each solution in the Pareto-optimal and then taking the median of these values. The comparison described above was also performed with no clustering of the Pareto-optimal solutions, in order to evaluate the effect of accounting for all solutions in the median values of the predicted hydrologic indices.

3. Results and discussion

3.1. Convergence and spread of optimal Pareto fronts obtained with multi-objective calibration strategies

The behavior of the normalized hypervolume indicator with respect to the number of generations is presented in Fig. 4. Results show a faster convergence for the NSE-based strategy compared to the RMSE-based strategy since the value for the normalized hypervolume indicator decreases faster with a decreasing generation number for the NSE-based strategy. In addition, Fig. 5 shows the final Pareto fronts obtained after 349 and 484 generations (i.e. where the steady behavior of the hypervolume indicator starts to be observed) using the NSE- and RMSE-based calibration strategies, respectively.

The solutions along the NSE-based optimal Pareto front range from 0.22 to 0.76 for NSE, from 0.37 to 0.73 for NSE_{sqrt}, and from 0.57 to 0.81 for NSE_{rel}. The optimal Pareto front is characterized by two distinct regions with significant tradeoffs. One of those regions, the low flow cluster, shows NSE_{rel} above 0.77 (Fig. 5a) with lower values for NSE and NSE_{sqrt} spanning from 0.22 to 0.35, and from 0.37 to 0.45, respectively. The other region, the high and medium flow clusters, shows acceptable NSE and NSE_{sqrt} values (i.e. from 0.73 to 0.76 and from 0.58 to 0.71, respectively) while NSE_{rel} ranges from 0.57 to 0.71. These results indicate that solutions with a very good representation of low discharges provide a poor representation of peak and medium flows (Guo et al., 2014; Shafii and De Smedt, 2009). However, the results also suggest that low discharges still exhibit acceptable efficiencies for the best representations of high and medium flows. Additionally, a strong linear correlation (R^2 = 0.91) between the NSE and NSE_{sqrt} objective functions is observed, though it is weaker (R^2 = 0.002) for smaller values for the NSE_{rel} objective function (i.e. low flow cluster). Hence, it is likely that NSE and NSE_{sqrt} are providing similar information for model calibration and therefore similar results can be achieved discarding one of these two objective functions. With respect to the RMSE-based optimal Pareto front, the performance measures range from 8.3 to 20.4 m^3/s for $RMSE_H$, from 2.1 to 4.9 m^3/s for $RMSE_M$, and from 0.81 to 3.6 m^3/s for $RMSE_L$ (Fig. 5b). These values indicate that, in general, the RMSE-based values range from 30% to 130% of the observed average discharges for each flow category (high, medium, and low) defined using the observed Q25 and Q75 thresholds. Moreover, the RMSE-based optimal Pareto clusters are layered along the $RMSE_H$ direction. This behavior suggests that the high flow cluster can represent some medium and low discharges that are well represented by medium and low flow clusters.

3.2. Reduction of initial parameter ranges by multi-objective calibration strategies

The Pareto-optimal calibrated SWAT parameter ranges varied according to the multi-objective calibration strategy (see Table 1). In general, results suggest that the NSE-based calibration strategy was able to provide narrower calibrated ranges for model parameters than the RMSE-based strategy. Moreover, for some parameters, the implementation of the k-means method to the optimal Pareto fronts allowed the identification of different ranges depending on the role of the objective functions in each cluster (e.g. importance of NSE and $RMSE_H$ for high flow, and importance of NSE_{rel} and $RMSE_L$ for low flow). In order to compare the results for calibration parameters, we considered whether or not they showed a significant reduction in their ranges after the calibration process (i.e. narrower calibration range with respect to initial calibration range), and whether or not they showed similar final ranges in each multi-objective calibration strategy. Regarding the significant reduction in calibration ranges, we found that a group of parameters describing mainly HRU and groundwater components showed very similar initial and final ranges for both calibration strategies. These parameters were BIOMIX (biological mixing), CANMX (max. canopy storage), EPCO (plant uptake), GW_REVAP (groundwater "revap" coefficient) and REVAPMN (groundwater threshold depth for "revap"). In contrast, the groundwater parameter GWQMN (threshold depth for flow return) reduced in range for both multi-objective strategies. However, the ranges varied depending on the calibration strategy. Likewise, some parameters mainly related to groundwater and

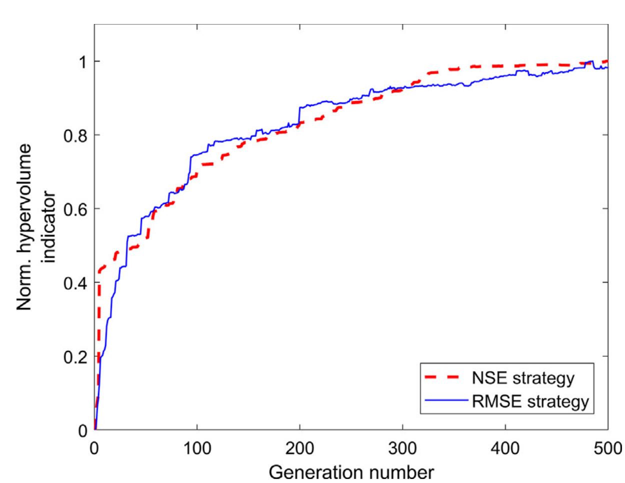

Fig. 4. Normalized hypervolume indicator behavior over the NSGA-III search process for each calibration strategy.

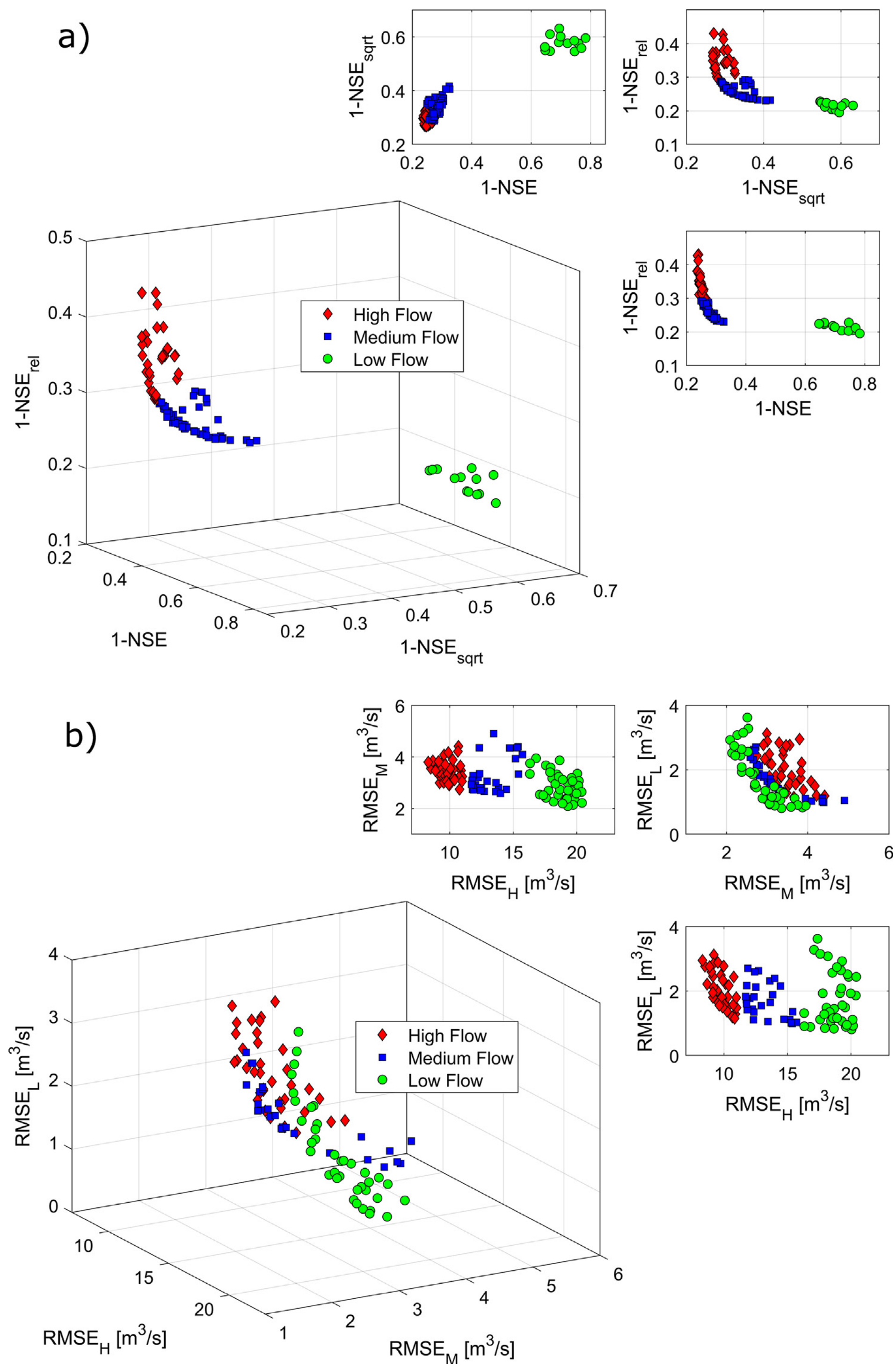

Fig. 5. Clustered Pareto-optimal solutions obtained by employing the NSGA-III algorithm and *k*-means clustering method for a) NSE-based and b) RMSE-based multi-objective calibration strategies. For each strategy (NSE- or RMSE-based), 100 Pareto-optimal solutions were clustered as follows: NSE-based: 32 high flow, 55 medium flow, and 13 low flow; RMSE-based: 36 high flow, 27 medium flow, and 37 low flow.

routing components (such as ESCO (soil evaporation), GW_DELAY (groundwater delay time), ALPHA_BF (baseflow factor), RCHRG_DP (percolation factor), CH_N (2) (Manning coefficient), and CH_K (2) (alluvium hydraulic conductivity)) had a smaller range after calibration regardless of calibration strategy. It is worth noting that GW_DELAY, ALPHA_BF and CH_N (2) were within different ranges depending on contrasting flow conditions (i.e. high and low flow clusters). For example, GW_DELAY resulted in a range of 0–0.1 days for high flow conditions, and a range of 237–309 days for low flow conditions using the NSE-based strategy. Meanwhile, CN2 (curve number for moisture condition II) and SOL_AWC (soil water capacity) showed contrasting ranges in high and low flows for the NSE-based strategy. For instance, low flow conditions favored positive multiplicative factors for CN2, increasing runoff potential, while providing negative multiplicative factors for SOL_AWC, reducing the available water capacity of soils. For high flow conditions, CN2 and SOL_AWC results were the opposite. However, the RMSE-based strategy did not provide reduced calibrated ranges for these two parameters. Finally, SURLAG (surface runoff lag) showed very similar reduced ranges for all flow conditions in the NSE-based strategy, while providing a reduced range only for high flow condition in the RMSE-based strategy.

3.3. Flow duration curves and streamflow time series representation

Fig. 6 presents FDC and hydrographs for the simulation period. Visual inspection of the simulated and observed curves reveals that the NSE-based strategy provides less variability than the RMSE-based strategy, represented by the width of light gray bound of solutions. For instance, Q25 and Q75 for the NSE-based strategy ranged from 7.4 to 11.8 m^3/s and from 2.6 to 4.3 m^3/s, respectively (Fig. 6a–c). Meanwhile, Q25 and Q75 for the RMSE-based strategy ranged from 4.6 to 12.2 m^3/s and from 2.6 to 5.9 m^3/s, respectively (Fig. 6d–f). This observation suggests that the higher variability for streamflow simulations given by the RMSE-based calibration strategy is consistent with the wide CN2 ((−0.25)–0.23 for high flow and (−0.24)–0.25 for low flow) and SOL_AWC ((−0.19)–0.23 for high flow and (−0.22)–0.23 for low flow) calibrated ranges as presented in Table 1. This was supported by previous studies in which, these parameters were identified as highly sensitive for streamflow calibration (van Griensven et al., 2006). In contrast, in the case of the NSE-based strategy, CN2 and SOL_AWC, which were treated as global multipliers (see Table 1), resulted in narrow calibrated ranges ((−0.25)–(−0.22) for high flow and

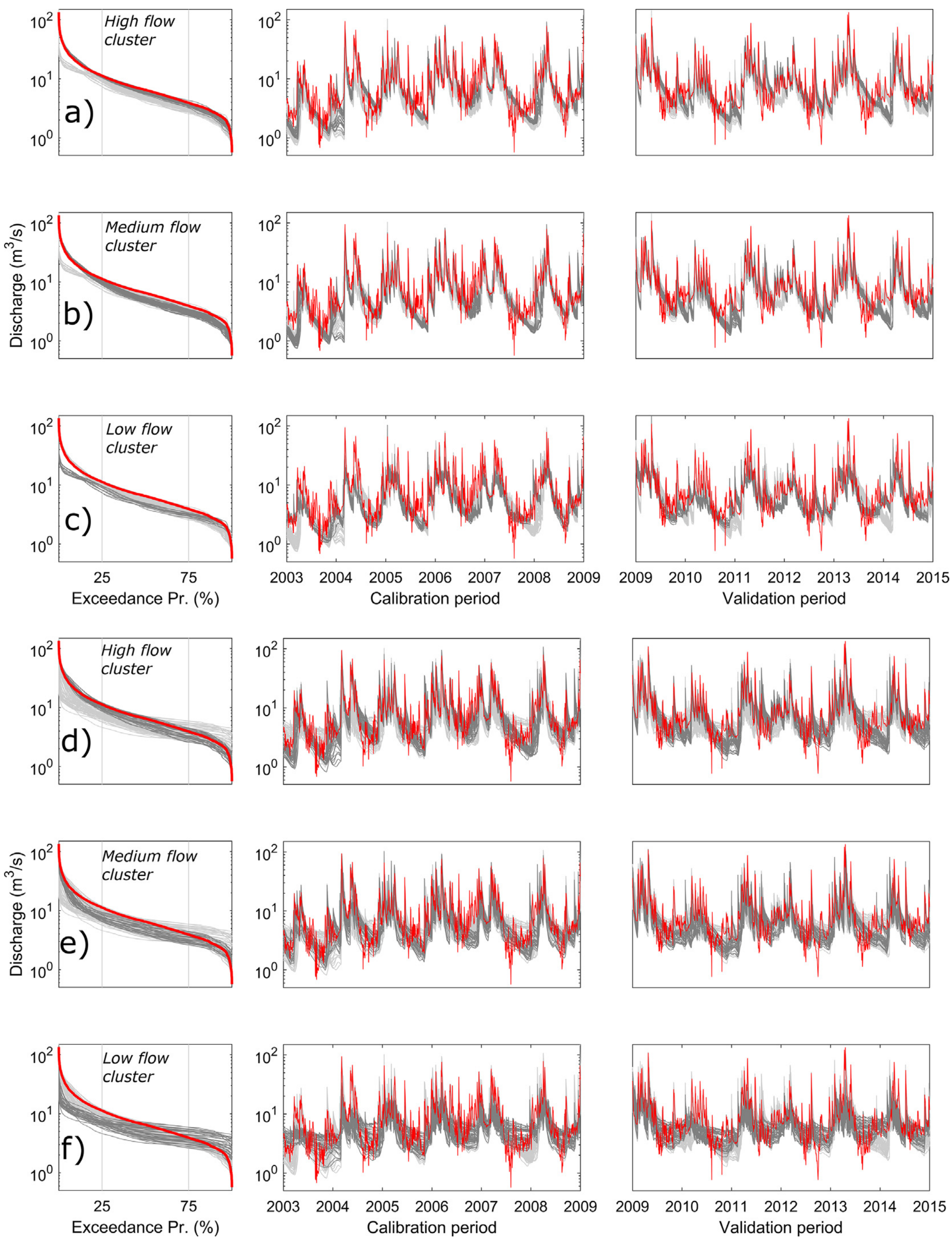

Fig. 6. Flow duration curves and time series obtained from Pareto-optimal solutions (light gray) and clustered (high, medium, and low flow) solutions (dark gray) for NSE-based (a, b, and c), and RMSE-based (d, e, and f) multi-objective calibration strategies. Red lines correspond to observed streamflow values. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

0.246–0.249 for low flow and 0.02–0.25 for high flow and (−0.25)–(−0.13) for low flow, respectively). Note that some extreme discharges, especially low flow events, lie outside all the simulation bounds provided by optimal Pareto solutions considered here. Also, some descending limbs and subsequent low flow pulses are poorly simulated in both calibration and validation periods. Therefore, limitations in the representation of extreme low and high flow related indices are expected. However, different sources of error may play a role here including input data uncertainties and structural inadequacies (Price et al., 2012). In both calibration strategies, high flow cluster bounds include most of the observed FDC, while shrinking the dispersion of simulated FDCs. It is worth noting that the group of simulated FDCs obtained from the NSE-based calibration strategy splits into two branches at the portion representing discharges exceeded 25% of the time (see the left section of the FDC subplots in Fig. 6a–c). For the aforementioned calibration strategy, only the low flow cluster does not have any simulated FDC representing the corresponding branch for high discharges (Fig. 6c). Additionally, medium flow cluster shows the largest variability in the NSE-based calibration strategy (Fig. 6b), while lower flow cluster does in the RMSE-based strategy (Fig. 6f).

3.4. Statistical analysis for predicted streamflow time-series

We performed the statistical analysis of the mean difference between observed and simulated streamflow time series for the simulation period comprised from 2003 to 2014. Most of the results showed that GLS with CAR (1) correlation outperforms other methods in estimating the mean streamflow difference, since other methods do not take into account the serial correlation in time series, as indicated by smaller AICs. In a few rare cases, the t-student model is better than GLS-CAR (1), meaning it is even more important to model the heavy-tail distribution rather than the serial correlation. The percentage of Pareto-optimal solutions in each cluster without enough evidence of significant difference with a confidence level of 95% (Table 2), only account for the results obtained with GLS with AR (1) or CAR (1) correlations. Results for the statistical analysis indicated that, in general, the NSE-based strategy provides many more Pareto-optimal solutions without significant differences, compared to the RMSE-based strategy. Table 2 also shows that most solutions that belong to the high flow cluster of the Pareto front yield good mean representations of overall, high, and medium streamflow time series values in both multi-objective calibration strategies. However, the percentages for the high flow category do not surpass 47%, because of the recurrent under-estimation of the mean of time series comprising high discharges. Surprisingly, the medium flow cluster resulted in more solutions with good mean representation of discharges below Q25 threshold (i.e. low flow values) in both calibration strategies. Therefore, it is possible to infer that solutions with simultaneous good performances based on both NSE and NSE_{rel} (Pareto front region conformed by high and medium flow clusters, see Fig. 5) lead to sound representation of overall streamflow time series and specific flow conditions, which is consistent with the graphical results obtained for FDCs and hydrographs presented in Fig. 6.

3.5. The level of predictability of ecologically-relevant hydrologic indices using multi and single-objective strategies

3.5.1. Multi-objective calibration strategies

For the period from 2003 to 2014, we computed the relative errors between 171 ecologically-relevant hydrologic indices obtained from the Pareto-optimal solutions and those obtained from the observed hydrograph. Results were organized according to the eleven hydrologic index groups described in Section 2.4. As described in Section 2.7, for each group and multi-objective calibration strategy, the analysis was conducted using the following subsets of Pareto-optimal solutions: complete Pareto front (i.e. all points) and high, medium, and low flow clusters obtained with the *k*-means method. Indices whose median relative errors were outside the ± 30% bound for all different subsets of Pareto-optimal solutions, are reported in Table 3 and described in Table S1. Hence, we considered that these indices were not well represented by the calibration strategies and the model structure employed in this study. The results indicated that the NSE-based calibration strategy was able to provide acceptable representation of 128 indices, while the RMSE-based calibration strategy did the same for 123 indices. Mostly, the RMSE-based strategy provided more dispersion for indices values than the NSE-based strategy (see Table 4).

In general, the calibrated model has high probability to misrepresent the ecologically-relevant hydrologic indices under extreme conditions (low and high flows), annual hydrograph recession period (June to October), and the flood pulses that happen during the low flow period (Fig. 6). As apparent in the figure, the overlap between the predicted versus the observed data are minimal during the low flow conditions (August–October) and high flow conditions (May–July). This fact is further supported by the high relative errors obtained for the average maximum monthly flows for June, July and October (i.e. MH6, MH7 and MH10) and for indices related to perennial streams (Olden and Poff, 2003) such as ML14, ML16, ML21, MH10, FL1, DL11, DL16, DH15, RA6 and RA7 (Table 3). This implies that the calibrated model is not able to accurately estimate key indices for perennial streams. Additionally, it is important to note that extreme high flows (May–July) are also being under-predicted, as indicated by the statistical analysis performed in Section 3.4. As a result, the level of error for estimating extreme high flow indices is large. This phenomenon has been observed in previous studies while being a subject of current research in hydrological science (Garcia et al., 2017; Murphy et al., 2013; Pfannerstill et al., 2014; Shrestha et al., 2014).

Regarding the magnitude of flow events, both multi-objective calibration strategies provided acceptable results for 76 out of 94 indices (81%). These strategies were not able to fully represent the variability of flows across months and years, and the magnitude (mean and median) of annual extreme flows. For instance, under medium flow conditions, results in Table 3 indicate a poor representation of the variability of some summer and fall monthly flows (i.e. MA29-MA33, which are expressed in terms of the coefficient of variation) and the skewness in annual flows (i.e. MA45, represented in terms of the difference between the mean and median annual flows,). Additionally, the NSE-based calibration strategy generated high relative errors for the variability across annual flows, expressed in terms of the range or 90th–10th percentiles (i.e. MA42 and MA44, which include extreme flow values). For low flow conditions, the mean and median of annual minimum flows were not well replicated (i.e. ML14 and ML16,

Table 2
Percentage of Pareto-optimal solutions without evidence of significant mean difference (α = 0.05) between simulated and observed time-series, considering different time series categories and clusters for both calibration strategies. Cluster with highest percentage for each flow category are in bold.

Category	Cluster	Percentage	
		NSE-based	RMSE-based
Complete time series	**High flow**	**88%**	50%
	Low flow	0%	0%
	Medium flow	27%	0%
High flow extracted time series	**High flow**	**47%**	28%
	Low flow	0%	0%
	Medium flow	13%	0%
Medium flow extracted time series	**High flow**	**94%**	56%
	Low flow	0%	5%
	Medium flow	22%	11%
Low flow extracted time series	High flow	0%	17%
	Low flow	62%	11%
	Medium flow	**75%**	26%

Table 3
List of ecologically-relevant hydrologic indices with all, high, medium, and low flow Pareto-optimal solutions having median relative errors outside the ± 30% bound, for each multi-objective calibration strategy.

Hydrologic index group	No. of indicators	Median values outside ± 30% relative error**		
		Both NSE- and RMSE-based	Only NSE-based	Only RMSE-based
Magnitude of flow events				
Average flow conditions	45	MA29, MA30, MA31, MA32, MA33, MA45	MA42, MA44	MA34
Low flow conditions	22	ML7, ML8, ML14, ML15, ML16, ML19, ML21, ML22		ML9, ML17
High flow conditions	27	MH6, MH7, MH10, MH21	MH22	MH11, MH23
Frequency of flow events				
Low flow conditions	3	FL1		
High flow conditions	11	FH5, FH9	FH4	FH1, FH8
Duration of flow events				
Low flow conditions	20	DL1, DL6, DL7, DL8, DL11, DL16		DL2, DL12
High flow conditions	24	DH15, DH17, DH21	DH19	DH20, DH23
Timing of flow events				
Average flow conditions	3			
Low flow conditions	4		TL4	
High flow conditions	3			
Rate of change in flow events				
Average flow conditions	9	RA3, RA6, RA7		

** DH15, High flow pulse duration with a threshold equal to the 75th percentile of the entire flow record (days); DH17, High flow duration with a threshold equal to the median flow (days); DH19, High flow duration with a threshold equal to 7 times the median flow (days); DH20, High flow duration with a threshold equal to the 75th percentile value for the median annual flows (days); DH21, High flow duration with a threshold equal to the 25th percentile value for the median annual flows (days); DH23, Flood duration with a threshold equal to the flow equivalent for a flood recurrence of 1.67 years (days); DL1, Annual minimum daily flow ($m^3 s^{-1}$); DL2, Annual minimum of 3-day moving average flow ($m^3 s^{-1}$); DL6, Variability of annual minimum daily average flow; DL7, Variability of annual minimum of 3-day moving average flow; DL8, Variability of annual minimum of 7-day moving average flow; DL11, Mean of 1-day minima of daily discharge; DL12, Mean of 3-day minima of daily discharge; DL16, Low flow pulse duration (days); FH1, High flood pulse count with a threshold equal to the 25th percentile of the entire flow record ($year^{-1}$); FH4, High flood pulse count with a threshold equal to 7 times median daily flow ($year^{-1}$); FH5, Flood frequency with a threshold equal to the median flow ($year^{-1}$); FH8, Flood frequency with a threshold equal to the 25th percentile of the entire flow record ($year^{-1}$); FH9, Flood frequency with a threshold equal to the 75th percentile of the entire flow record ($year^{-1}$); FL1, Low flood pulse count ($year^{-1}$); MA29, Variability in June flows; MA30, Variability in July flows; MA31, Variability in August flows; MA32, Variability in September flows; MA33, Variability in October flows; MA34, Variability in November flows; MA42, Variability across annual flows; MA44, Variability across annual flows; MA45, Skewness in annual flows; MH6, Mean maximum June monthly flow ($m^3 s^{-1}$); MH7, Mean maximum July monthly flow ($m^3 s^{-1}$); MH10, Mean maximum October monthly flow ($m^3 s^{-1}$); MH11, Mean maximum November monthly flow ($m^3 s^{-1}$); MH21, High flow volume (days); MH22, High flow volume (days); MH23, High flow volume (days); ML7, Mean minimum July monthly flow ($m^3 s^{-1}$); ML8, Mean minimum August monthly flow ($m^3 s^{-1}$); ML9, Mean minimum September monthly flow ($m^3 s^{-1}$); ML14, Mean of annual minimum flows; ML15, Low flow index; ML16, Median of annual minimum flows; ML17, Baseflow index based on the seven-day minimum flow; ML19, Baseflow index based on the lowest annual daily flow; ML21, Variability across annual minimum flows; ML22, Specific mean annual minimum flows ($m^3 s^{-1} km^{-2}$); RA3, Fall rate ($m^3 s^{-1} d^{-1}$); RA6, Change of flow- increasing ($m^3 s^{-1}$); RA7, Change of flow – decreasing ($m^3 s^{-1}$); TL4, Seasonal predictability of non-low flow.

respectively), also affecting the results for some indices depending on these values (e.g. low flow index, ML15; baseflow index, ML19; and variability across annual minimum flows, ML21). For high flow conditions, both calibration strategies had limitations in representing high flow volumes (e.g. MH21) and the mean maximum monthly flows for some summer and fall months (e.g. MH6, June; MH7, July; MH10, October).

Regarding the frequency of flow events, both multi-objective calibration strategies generated acceptable median values for 10 out of 13 indices (77%). The indices that were not well represented include the low flow pulse count (i.e. FL1) and some flood frequency indices that use the median and 75th percentile of flows as upper thresholds (i.e. FH5 and FH9). Moreover, the NSE-based strategy yielded poor representations of a high flood pulse count index based on a very high upper threshold (i.e. FH4, which uses 7 times median flow), while, the RMSE-based strategy produced limited results for high flood pulse count (i.e. FH1) and flood frequency using percentile 25th as threshold (i.e. FH8).

For the duration of flow events, both calibration strategies resulted in acceptable values for 32 out of 41 indices (78%). The results for this group of hydrologic indices were consistent with the poor representation of some indices describing the magnitude and frequency of flow events. For instance, duration indices related to magnitude and variability of daily and annual minima (i.e. DL1 and DL11, and DL6-DL8, respectively) yielded elevated relative errors for both strategies. Similarly, high flow indices with the median and 75th percentile of flows used as thresholds (i.e. DH17 and DH15, respectively), produced high relative errors too. Likewise, the NSE-based strategy presented difficulties representing high flow duration using seven times the median as an upper threshold (i.e. DH19). Moreover, the RMSE-based strategy yielded large deviations for the annual minima of 3-day means of daily discharge (i.e. DL2), the mean annual 3-day minimum of daily discharge (i.e. DL12), and indices related to flood duration (i.e. DH20, DH23) because of poor results in the identification of the number of extreme high flow events.

With respect to the timing of flow events, all the four indices were well reproduced using both multi-objective calibration strategies. However, the NSE-based strategy was not able to produce median relative errors within ± 30% for the seasonal predictability of non-low flow (i.e. TL4). Finally, regarding the rate of change in flow events, both multi-objective calibration strategies reproduced 6 out of 9 indices (67%) with median relative errors outside ± 30% for the fall rate (i.e. RA3) and change of flow for increasing and decreasing discharges (i.e. RA6 and RA7, respectively).

In comparison to previous studies where hydrological modeling was employed to predict ecological-relevant hydrologic indices (Caldwell et al., 2015; Kiesel et al., 2017; Murphy et al., 2013; Shrestha et al., 2014; Vis et al., 2015), the use of the median of different optimal-Pareto subsets improved the representation of some indicators (e.g. Julian day of annual minimum, TL1; high flood pulse count, FH1; rise rate, RA1; reversals, RA8). However, indices related to the frequency and duration of high and low flow pulses (e.g. FL1, DL16, and DH15, see below) were

Table 4
The lowest median relative error and corresponding interquartile range (IQR) and flow cluster for each multi-objective calibration strategy for the Indicators of Hydrologic Alteration (IHA). Values that exceed ± 30% bound of relative error are highlighted.

IHA group	Hydrologic index**	NSE-based			RMSE-based		
		Relative error	IQR	Cluster	Relative error	IQR	Cluster
Magnitude of monthly water conditions	MA12	-3.4%	10.5%	High	-6.9%	11.4%	High
	MA13	0.2%	6.1%	Low	-7.9%	9.8%	High
	MA14	12.9%	7.3%	High	9.3%	10.4%	High
	MA15	-7.8%	2.5%	High	-5.9%	6.2%	High
	MA16	1.2%	3.9%	High	-3.0%	8.5%	High
	MA17	-0.5%	7.4%	High	-4.5%	14.9%	High
	MA18	-1.0%	24.2%	All	-1.8%	32.6%	All
	MA19	7.5%	10.2%	Medium	-1.0%	30.6%	Medium
	MA20	1.8%	7.8%	High	0.7%	22.9%	High
	MA21	-21.9%	6.5%	High	-23.7%	38.6%	Low
	MA22	-17.9%	13.7%	High	-25.0%	20.7%	High
	MA23	-7.1%	8.1%	High	-13.9%	13.7%	High
Magnitude and duration of annual extreme water conditions (mean daily flow)	DL1	**36.8%**	10.5%	Medium	**71.8%**	48.2%	High
	DL2	10.3%	8.5%	Medium	**38.5%**	38.7%	High
	DL3	2.1%	16.8%	All	19.8%	33.4%	High
	DL4	-8.7%	6.8%	High	2.2%	30.6%	All
	DL5	-17.6%	6.1%	High	-14.8%	39.2%	Low
	DH1	-17.1%	3.7%	High	-14.6%	11.9%	High
	DH2	-8.9%	3.9%	Medium	-6.5%	10.3%	High
	DH3	-2.2%	3.8%	High	0.5%	9.8%	High
	DH4	-0.3%	4.2%	All	1.5%	10.1%	High
	DH5	-0.4%	5.1%	All	0.2%	6.3%	High
	ML17	13.1%	4.1%	Medium	**32.4%**	25.1%	High
Timing of annual extreme water conditions	TL1	6.7%	4.0%	Low	12.2%	12.8%	Low
	TH1	-0.4%	10.2%	Medium	0.4%	32.8%	All
Frequency and duration of high and low pulses	FL1	**-65.5%**	25.7%	Low	**-75.7%**	16.2%	Low
	DL16	**213.5%**	167.2%	Low	**144.9%**	181.2%	Low
	FH1	-28.4%	3.4%	High	**-44.9%**	16.1%	High
	DH15	**60.0%**	34.4%	Low	**100.0%**	53.4%	High
Rate and frequency of water condition changes	RA1	18.9%	11.6%	Medium	-16.9%	27.3%	Medium
	RA3	**-50.0%**	2.8%	High	**-53.0%**	13.2%	High
	RA8	5.5%	7.3%	Low	0.9%	25.2%	Low

** DH1, Annual maxima of 1-day means of daily discharge ($m^3 s^{-1}$) DH2, Annual maxima of 3-day means of daily discharge ($m^3 s^{-1}$); DH3, Annual maxima of 7-day means of daily discharge ($m^3 s^{-1}$); DH4, Annual maxima of 30-day means of daily discharge ($m^3 s^{-1}$); DH5, Annual maxima of 90-day means of daily discharge ($m^3 s^{-1}$); DH15, High flow pulse duration with a threshold equal to the 75th percentile of the entire flow record (days); DL1, Annual minimum daily flow ($m^3 s^{-1}$); DL2, Annual minimum of 3-day moving average flow ($m^3 s^{-1}$); DL3, Annual minima of 7-day means of daily discharge ($m^3 s^{-1}$); DL4, Annual minima of 30-day means of daily discharge ($m^3 s^{-1}$); DL5, Annual minima of 90-day means of daily discharge ($m^3 s^{-1}$); DL16, Low flow pulse duration (days); FH1, High flood pulse count with a threshold equal to the 25th percentile of the entire flow record ($year^{-1}$); FL1, Low flood pulse count ($year^{-1}$); MA12, Mean monthly flow for January ($m^3 s^{-1}$); MA13, Mean monthly flow for February ($m^3 s^{-1}$); MA14, Mean monthly flow for March ($m^3 s^{-1}$); MA15, Mean monthly flow for April ($m^3 s^{-1}$); MA16, Mean monthly flow for May ($m^3 s^{-1}$); MA17, Mean monthly flow for June ($m^3 s^{-1}$); MA18, Mean monthly flow for July ($m^3 s^{-1}$); MA19, Mean monthly flow for August ($m^3 s^{-1}$); MA20, Mean monthly flow for September ($m^3 s^{-1}$); MA21, Mean monthly flow for October ($m^3 s^{-1}$); MA22, Mean monthly flow for November ($m^3 s^{-1}$); MA23, Mean monthly flow for December ($m^3 s^{-1}$); ML17, Baseflow index based on the seven-day minimum flow; RA1, Rise rate ($m^3 s^{-1} d^{-1}$); RA3, Fall rate ($m^3 s^{-1} d^{-1}$); RA8, Reversals ($year^{-1}$); TH1, Julian date of annual maximum; TL1, Julian date of annual minimum.

poorly simulated. For instance, Table 4 presents the lowest relative error obtained for a suite of 32 indices included in the software Indicators of Hydrologic Alteration (IHA) (The Nature Conservancy, 2009) that were evaluated by Shrestha et al. (2014). For this group of indices, while calibrated solutions can properly reproduce the magnitude, duration and timing of different flow conditions (with difficulties for DL1, the annual minima of daily flows), the frequency and duration of low flood pulses (i.e. FL1 and DL16, respectively), the duration of high flow pulses (i.e. DH15), and the fall rate (i.e. RA3) still showed high deviances. It is important to note that most of the indices in Table 4 were well represented by high flow clusters, which are dominated by good performances for NSE or $RMSE_H$. However, the misrepresented low flow indices FL1 and DL16 obtained the lowest relative errors using optimal-Pareto solutions from low flow clusters. These clusters have a better description of low flow discharges, particularly in the seasonal transition from summer to fall.

3.5.2. Single-objective calibration

We obtained the individual model simulations that minimized each NSE-based objective function from the optimized Pareto front obtained after the NSGA-III algorithm implementation, with their corresponding results for the ecologically-relevant hydrologic indices. Maximum attained values for NSE, NSE_{sqrt} and NSE_{rel} were 0.76, 0.73 and 0.81, respectively. Some of the indices reported in Table 3 were represented within the satisfactory threshold of ± 30% using some of the single-objective calibrated models. As expected, the optimal NSE model provided acceptable results for high flow related indices: mean maximum monthly flows for June and July (MH6 and MH7, respectively), high flow volume using as threshold three times the median annual flow (MH22), high flow duration with seven times the median flow as the upper threshold (DH19) and the seasonal predictability of non-low flows (TL4). Similarly, NSE_{sqrt} produced acceptable outcomes for high flood pulse count with seven times the median daily flow as the upper threshold (FH4), in addition to MH22 (high flow volume), DH19 (high flow duration) and TL4 (seasonal predictability), also given by the optimal NSE model. Additionally, the optimal NSE_{sqrt} model provided acceptable results for the mean maximum October flow (MH10) which is a key index for perennial streams (related to low flows during the fall season). Meanwhile, the optimal NSE_{rel} model, which is insensitive to peak flows and biased towards low flows, surprisingly improved the representation of the high flow pulse duration (DH15), a key indicator for perennial streams. This occurs because NSE_{rel} significantly reduces the influence of absolute differences during high flow events (Krause et al., 2005). Therefore, the NSE_{rel} objective function has the property of benefiting simulations that better describe the overall shape of the hydrograph, which can be graphically evinced in Fig. 6c for the NSE-based low flow cluster simulations.

Indices presented in Table 4 that were poorly represented by the single-objective calibration approach using the standard NSE include MA19 (mean monthly flow for August), DL2 (annual minimum of 3-day average flow), TL1 (Julian date of annual minimum) and RA8 (reversals). Likewise, optimal NSE_{sqrt} model provided poor representation for MA19 (August mean flow), DL2 (3-day annual minimum), RA1 (rise rate) and R8 (reversals). In addition, *indices poorly represented by the* NSE_{rel} optimal model include MA14-MA17 and MA21-MA22 (mean monthly flows for March-June and October-November, respectively), ML17 (seven-day minimum flow divided by mean annual daily flows averaged across all years), DL5 (seasonal magnitude of minimum annual flow), and DH2-DH5 (magnitude of maximum annual flow from 3-day duration to seasonal), which are mainly related with high flow events. All the aforementioned indices were well-represented by the NSE-based multi-objective calibration strategy (see Table 4).

3.5.3. Comparison of single and multi-objective calibration approaches

In general, the two multi-objective performed better than the three single-objective approaches in calculating the hydrologic indices, within a satisfactory range of ± 30% relative error. Regarding the single-objective approaches, the optimal NSE and NSE_{sqrt} models were able to satisfactorily represent 71% of the indices, while the optimal NSE_{rel} model was able to satisfactorily represent 47%. Meanwhile, the multi-objectives (NSE- and RMSE-based) approaches were able to represent 77% and 74% of the indices, respectively. Results revealed that no calibration approach can be considered the best for all indices, while they can be complementary to each other. However, the main advantage of multi-objective calibration over the single-objective approaches is that they provide a range of solutions for the streamflow and hydrologic indices. This range is derived from the Pareto-optimal front, which can be used for uncertainty quantification purposes (Efstratiadis and Koutsoyiannis, 2010).

4. Conclusions

This study evaluated the predictability of 167 ecologically-relevant hydrologic indices using different approaches for model calibration. For this purpose, we compared the performance of two multi-objective and three single-objective formulations employing the NSGA-III algorithm and the SWAT model structure.

Among the multi-objective formulations, the NSE-based strategy was found to be the preferred multi-objective model calibration approach since: 1) it provided the fastest convergence, the highest number of well-predicted indices, and the smallest dispersion in the indices' values over the different subsets of Pareto-optimal simulations; 2) the high and medium flow clusters of the Pareto front region contained the highest percentages of solutions with no evidence of significant mean difference between simulated and observed time series; 3) the same Pareto front region contained the highest percentages of solutions with the lowest median relative error and dispersion measure for hydrological indices; and 4) the calibration ranges were narrower than the RMSE-based strategy.

Among the single-objective approaches, NSE and NSE_{sqrt} presented similar results while both outperformed the multi-objective calibration approaches for estimating high flow indices. Meanwhile in this study, the representation of regime components such as the frequency and duration of low flows and the overall timing and rate of change were in general better using the multi-objective calibration approaches. However, it is recommended to use the NSE-based single and multi-objective calibration approaches together since they provide more reliable information for a larger number of indices. In addition, we suggest omitting either the standard NSE- or NSE_{sqrt}-based objective function because they provide similar information.

In this study, we proposed the use of the NSE_{rel} performance measure in a multi-objective framework in order to improve the representation of low and extreme low flow events while maintaining a good overall representation of other flow conditions. However, we showed that an NSE_{rel} objective function improves low flows while highly sacrificing the representation of other flow conditions, as opposed to the standard NSE for high flows which improves high flows while maintaining acceptable representation of other flow conditions. Therefore, during the calibration process we observed that NSE_{rel} affects the overall central tendency values (e.g. median and mean flows) for different temporal scales (e.g. monthly, annual), which negatively impacts the representation of low flow indicators (e.g. baseflow index).

We also demonstrated that the use of different sets of solutions, instead of a single optimal solution, introduces more flexibility in the predictability of different hydrologic indices of ecological interest. Moreover, we were able to identify a reduced group of poorly represented indices, particularly in the transition from summer to fall seasons. Since this study was performed in a watershed with a single streamflow gauging station, it is recommended that future studies should focus on replicating this work in other regions with more streamflow gauging stations and different climatological and physiographical characteristics in which flow patterns are widely different.

Acknowledgments

This work is supported by the USDA National Institute of Food and Agriculture, Hatch project MICL02359. J. Sebastian Hernandez–Suarez is grateful to Fulbright, COLCIENCIAS (Colombian Administrative Department of Science, Technology and Innovation) and the Office of International Students and Scholars at Michigan State University for the financial support Granted during 2016/17.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.jhydrol.2018.07.056.

References

Abouali, M., Daneshvar, F., Nejadhashemi, A.P., 2016. MATLAB Hydrological Index Tool (MHIT): a high performance library to calculate 171 ecologically relevant hydrological indices. Ecol. Inf. 33, 17–23. https://doi.org/10.1016/j.ecoinf.2016.03.004.

Andresen, J., Winkler, J., 2009. Weather and climate. In: Schaetzl, R., Darden, J., Brandt, D. (Eds.), Michigan Geography and Geology. Pearson Custom Publishing, Boston, MA.

Arnold, J.G., Moriasi, D.N., Gassman, P.W., Abbaspour, K.C., White, M.J., Srinivasan, R., Santhi, C., Harmel, R.D., Griensven, A.Van, VanLiew, M.W., Kannan, N., Jha, M.K., 2012. Swat: model use, calibration, and validation. Trans. ASABE 55, 1491–1508 ISSN 2151-0032.

Arnold, J.G., Srinivasan, R., Muttiah, R.S., Williams, J.R., 1998. Large area hydrologic modeling and assesment Part I: Model development. JAWRA J. Am. Water Resour. Assoc. 34, 73–89. https://doi.org/10.1111/j.1752-1688.1998.tb05961.x.

Arthur, D., Vassilvitskii, S., 2007. K-Means + +: the Advantages of Careful Seeding. In: Proc. eighteenth Annu. ACM-SIAM Symp. Discret. algorithms 8, pp. 1027–1025. doi:10.1145/1283383.1283494.

Auger, A., Bader, J., Brockhoff, D., Zitzler, E., 2009. Theory of the hypervolume indicator. In: Proc. tenth ACM SIGEVO Work. Found. Genet. algorithms – FOGA '09 87. doi:10.1145/1527125.1527138.

Bekele, E.G., Nicklow, J.W., 2007. Multi-objective automatic calibration of SWAT using NSGA-II. J. Hydrol. 341, 165–176. https://doi.org/10.1016/j.jhydrol.2007.05.014.

Bunn, S.E., Arthington, A.H., 2002. Basic principles and ecological consequences of altered flow regimes for aquatic biodiversity. Environ. Manage. 30, 492–507. https://doi.org/10.1007/s00267-002-2737-0.

Caldwell, P.V., Kennen, J.G., Sun, G., Kiang, J.E., Butcher, J.B., Eddy, M.C., Hay, L.E., Lafontaine, J.H., Hain, E.F., Nelson, S.A.C., Mcnulty, S.G., 2015. A comparison of hydrologic models for ecological flows and water availability. Ecohydrology 8, 1525–1546. https://doi.org/10.1002/eco.1602.

Carlisle, D.M., Falcone, J., Wolock, D.M., Meador, M.R., Norris, R.H., 2009. Predicting the natural flow regime: models for assessing hydrological alteration in streams. River Res. Appl. 30, n/a-n/a. https://doi.org/10.1002/rra.1247.

Carpenter, S.R., Stanley, E.H., Vander Zanden, M.J., 2011. State of the World's freshwater ecosystems: physical, chemical, and biological changes. Annu. Rev. Environ. Resour. 36, 75–99. https://doi.org/10.1146/annurev-environ-021810-094524.

Cibin, R., Sudheer, K.P., Chaubey, I., 2010. Sensitivity and identifiability of stream flow generation parameters of the SWAT model. Hydrol. Process. 24, 1133–1148. https://doi.org/10.1002/hyp.7568.

Coello Coello, C.A., Lamont, G.B., Veldhuizen, D.A.Van, 2007. Evolutionary algorithms for solving multi-objective problems. Genetic and Evolutionary Computation Series, 2nd ed. Springer US, Boston, MA.

Confesor, R.B., Whittaker, G.W., 2007. Automatic calibration of hydrologic models with multi-objective evolutionary algorithm and pareto optimization. JAWRA J. Am. Water Resour. Assoc. 43, 981–989. https://doi.org/10.1111/j.1752-1688.2007.00080.x.

Deb, K., 2008. A robust evolutionary framework for multi-objective optimization. In: Proceedings of the 10th Annual Conference on Genetic and Evolutionary Computation – GECCO '08. ACM Press, New York, New York, USA, pp. 633 doi:10.1145/1389095.1389223.

Deb, K., 2001. Multi-objective Optimization using Evolutionary Algorithms. John Wiley & Sons.

Deb, K., Jain, H., 2014. An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, part I: solving problems with box constraints. IEEE Trans. Evol. Comput. 18, 577–601. https://doi.org/10.1109/TEVC.2013.2281535.

Deb, K., Pratap, A., Agarwal, S., Meyarivan, T., 2002. A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Trans. Evol. Comput. 6, 182–197. https://doi.org/10.1109/4235.996017.

Dhungel, S., Tarboton, D.G., Jin, J., Hawkins, C.P., 2016. Potential effects of climate change on ecologically relevant streamflow regimes. River Res. Appl. 32, 1827–1840. https://doi.org/10.1002/rra.3029.

Dudgeon, D., Arthington, A.H., Gessner, M.O., Kawabata, Z.-I., Knowler, D.J., Lévêque, C., Naiman, R.J., Prieur-Richard, A.-H., Soto, D., Stiassny, M.L.J., Sullivan, C.A., 2006. Freshwater biodiversity: importance, threats, status and conservation challenges. Biol. Rev. Camb. Philos. Soc. 81, 163–182. https://doi.org/10.1017/S1464793105006950.

Efstratiadis, A., Koutsoyiannis, D., 2010. One decade of multi-objective calibration approaches in hydrological modelling: a review. Hydrol. Sci. J. 55, 58–78. https://doi.org/10.1080/02626660903526292.

EPA, 2015. Saginaw River and Bay Area of Concern [WWW Document]. https://www.epa.gov/saginaw-river-bay-aoc (accessed 7.12.17).

Garcia, F., Folton, N., Oudin, L., 2017. Which objective function to calibrate rainfall–runoff models for low-flow index simulations? Hydrol. Sci. J. 62. https://doi.org/10.1080/02626667.2017.1308511.

Guo, J., Zhou, J., Lu, J., Zou, Q., Zhang, H., Bi, S., 2014. Multi-objective optimization of empirical hydrological model for streamflow prediction. J. Hydrol. 511, 242–253. https://doi.org/10.1016/j.jhydrol.2014.01.047.

Hering, D., Borja, A., Carstensen, J., Carvalho, L., Elliott, M., Feld, C.K., Heiskanen, A.S., Johnson, R.K., Moe, J., Pont, D., Solheim, A.L., de Bund, W.van, 2010. The European Water Framework Directive at the age of 10: a critical review of the achievements with recommendations for the future. Sci. Total Environ. 408, 4007–4019. https://doi.org/10.1016/j.scitotenv.2010.05.031.

Herman, M.R., Nejadhashemi, A.P., 2015. A review of macroinvertebrate- and fish-based stream health indices. Ecohydrol. Hydrobiol. 15, 53–67. https://doi.org/10.1016/j.ecohyd.2015.04.001.

Hipsey, M.R., Hamilton, D.P., Hanson, P.C., Carey, C.C., Coletti, J.Z., Read, J.S., Ibelings, B.W., Valesini, F.J., Brookes, J.D., 2015. Predicting the resilience and recovery of aquatic systems: a framework for model evolution within environmental observatories. Water Resour. Res. 51, 7023–7043. https://doi.org/10.1002/2015WR017175.

Jain, H., Deb, K., 2014. An evolutionary many-objective optimization algorithm using reference-point based nondominated sorting approach, Part II: handling constraints and extending to an adaptive approach. IEEE Trans. Evol. Comput. 18, 602–622. https://doi.org/10.1109/TEVC.2013.2281534.

Jelks, H.L., Walsh, S.J., Burkhead, N.M., Contreras-Balderas, S., Diaz-Pardo, E., Hendrickson, D.A., Lyons, J., Mandrak, N.E., McCormick, F., Nelson, J.S., Platania, S.P., Porter, B.A., Renaud, C.B., Schmitter-Soto, J.J., Taylor, E.B., Warren, M.L., 2008. Conservation status of imperiled North American freshwater and diadromous fishes. Fisheries 33, 372–407. https://doi.org/10.1577/1548-8446-33.8.372.

Karr, J., 1996. Ecological integrity and ecological health are not the same. Eng. within Ecol. constraints.

Karr, J.R., 1999. Defining and measuring river health. Freshwater Biol. 41, 221–234. https://doi.org/10.1046/j.1365-2427.1999.00427.x.

Karr, J.R., Dudley, D.R., 1981. Ecological perspective on water quality goals. Environ. Manage. 5, 55–68. https://doi.org/10.1007/BF01866609.

Karr, J.R., Yoder, C.O., 2004. Biological assessment and criteria improve total maximum daily load decision making. J. Environ. Eng. 594–604.

Kennard, M.J., Mackay, S.J., Pusey, B.J., Olden, J.D., Marsh, N., 2009. Quantifying uncertainty in estimation of hydrologic metrics for ecohydrological studies. River Res. Appl. 30, n/a-n/a. https://doi.org/10.1002/rra.1249.

Kennen, J.G., Kauffman, L.J., Ayers, M.A., Wolock, D.M., Colarullo, S.J., 2008. Use of an integrated flow model to estimate ecologically relevant hydrologic characteristics at stream biomonitoring sites. Ecol. Modell. 211, 57–76. https://doi.org/10.1016/j.ecolmodel.2007.08.014.

Kiesel, J., Guse, B., Pfannerstill, M., Kakouei, K., Jähnig, S.C., Fohrer, N., 2017. Improving hydrological model optimization for riverine species. Ecol. Indic. 80, 376–385. https://doi.org/10.1016/j.ecolind.2017.04.032.

Knight, R.R., Gain, W.S., Wolfe, W.J., 2012. Modelling ecological flow regime: an example from the Tennessee and Cumberland River basins. Ecohydrology 5, 613–627. https://doi.org/10.1002/eco.246.

Krause, P., Boyle, D.P., Base, F., 2005. Comparison of different efficiency criteria for hydrological model assessment. Adv. Geosci. 89–97.

Lu, S., Kayastha, N., Thodsen, H., Griensven, Van, a, Andersen, H.E.,, 2014. Multiobjective calibration for comparing channel sediment routing models in the soil and water assessment tool. J. Environ. Qual. 43, 110–120. https://doi.org/10.2134/jeq2011.0364.

Maddock, I., 1999. The importance of physical habitat assessment for evaluating river health. Freshwater Biol. 41, 373–391. https://doi.org/10.1046/j.1365-2427.1999.00437.x.

Maier, H.R., Kapelan, Z., Kasprzyk, J., Kollat, J., Matott, L.S., Cunha, M.C., Dandy, G.C., Gibbs, M.S., Keedwell, E., Marchi, A., Ostfeld, A., Savic, D., Solomatine, D.P., Vrugt, J.A., Zecchin, A.C., Minsker, B.S., Barbour, E.J., Kuczera, G., Pasha, F., Castelletti, A., Giuliani, M., Reed, P.M., 2014. Evolutionary algorithms and other metaheuristics in water resources: Current status, research challenges and future directions. Environ. Model. Softw. 62, 271–299. https://doi.org/10.1016/j.envsoft.2014.09.013.

Moriasi, D.N., Gitau, M.W., Pai, N., Daggupati, P., 2015. Hydrologic and water quality models: performance measures and evaluation criteria. Trans. ASABE 58, 1763–1785. https://doi.org/10.13031/trans.58.10715.

Murphy, J.C., Knight, R.R., Wolfe, W.J., Gain, S.W., 2013. Predicting ecological flow regime at ungaged sites: a comparison of methods. River Res. Appl. 29, 660–669. https://doi.org/10.1002/rra.2570.

Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I — a discussion of principles. J. Hydrol. 10, 282–290. https://doi.org/10.1016/0022-1694(70)90255-6.

NASS, 2012. CropScape - NASS CDL Program [WWW Document]. https://nassgeodata.gmu.edu/CropScape/ (accessed 7.12.17).

NED, 2015. The National Map: Elevation [WWW Document]. https://nationalmap.gov/elevation.html (accessed 7.12.17).

Neitsch, S.L., Arnold, J.G., Kiniry, J.R., Williams, J.R., 2011. Soil and Water Assessment Tool Theoretical Documentation Version 2009. Texas Water Resources Institute, Temple, Texas.

NOAA, 2017. Data Tools | Climate Data Online (CDO) | National Climatic Data Center (NCDC) [WWW Document]. https://www.ncdc.noaa.gov/cdo-web/datatools/ (accessed 7.12.17).

NRCS, 2014. Web Soil Survey [WWW Document]. https://websoilsurvey.sc.egov.usda.gov/App/WebSoilSurvey.aspx (accessed 7.12.17).

Olden, J.D., Poff, N.L., 2003. Redundancy and the choice of hydrologic indices for characterizing streamflow regimes. River Res. Appl. 19, 101–121. https://doi.org/10.1002/rra.700.

Olsen, M., Troldborg, L., Henriksen, H.J., Conallin, J., Refsgaard, J.C., Boegh, E., 2013. Evaluation of a typical hydrological model in relation to environmental flows. J. Hydrol. 507, 52–62. https://doi.org/10.1016/j.jhydrol.2013.10.022.

Oudin, L., Andréassian, V., Mathevet, T., Perrin, C., Michel, C., 2006. Dynamic averaging of rainfall-runoff model simulations from complementary model parameterizations.

Water Resour. Res. 42, 1–10. https://doi.org/10.1029/2005WR004636.
Patrick, C.J., Yuan, L.L., 2017. Modeled hydrologic metrics show links between hydrology and the functional composition of stream assemblages. Ecol. Appl. 27, 1605–1617. https://doi.org/10.1002/eap.1554.
Pfannerstill, M., Guse, B., Fohrer, N., 2014. Smart low flow signature metrics for an improved overall performance evaluation of hydrological models. J. Hydrol. 510, 447–458. https://doi.org/10.1016/j.jhydrol.2013.12.044.
Poff, N.L., Allan, J.D., Bain, M.B., Karr, J.R., Prestegaard, K.L., Richter, B.D., Sparks, R.E., Stromberg, J.C., 1997. The natural flow regime. Bioscience 47, 769–784. https://doi.org/10.2307/1313099.
Poff, N.L., Richter, B.D., Arthington, A.H., Bunn, S.E., Naiman, R.J., Kendy, E., Acreman, M., Apse, C., Bledsoe, B.P., Freeman, M.C., Henriksen, J., Jacobson, R.B., Kennen, J.G., Merritt, D.M., O'Keeffe, J.H., Olden, J.D., Rogers, K., Tharme, R.E., Warner, A., 2010. The ecological limits of hydrologic alteration (ELOHA): a new framework for developing regional environmental flow standards. Freshwater Biol. 55, 147–170. https://doi.org/10.1111/j.1365-2427.2009.02204.x.
Poff, N.L., Zimmerman, J.K.H., 2010. Ecological responses to altered flow regimes: a literature review to inform the science and management of environmental flows. Freshwater Biol. 55, 194–205. https://doi.org/10.1111/j.1365-2427.2009.02272.x.
Price, K., Purucker, S.T., Kraemer, S.R., Babendreier, J.E., 2012. Tradeoffs among watershed model calibration targets for parameter estimation. Water Resour. Res. 48, 1–16. https://doi.org/10.1029/2012WR012005.
Pushpalatha, R., Perrin, C., Moine, N. Le, Andréassian, V., 2012. A review of efficiency criteria suitable for evaluating low-flow simulations. J. Hydrol. 420–421, 171–182. https://doi.org/10.1016/j.jhydrol.2011.11.055.
Reed, P.M., Hadka, D., Herman, J.D., Kasprzyk, J.R., Kollat, J.B., 2013. Evolutionary multiobjective optimization in water resources: the past, present, and future. Adv. Water Resour. 51, 438–456. https://doi.org/10.1016/j.advwatres.2012.01.005.
Richter, B.D., Baumgartner, J.V., Powell, J., Braun, D.P., 1996. A method for assessing hydrologic alteration within ecosystems. Conserv. Biol. 10, 1163–1174. https://doi.org/10.2307/2387152.
Sanborn, S.C., Bledsoe, B.P., 2006. Predicting streamflow regime metrics for ungauged streamsin Colorado, Washington, and Oregon. J. Hydrol. 325, 241–261. https://doi.org/10.1016/j.jhydrol.2005.10.018.
Seada, H., Deb, K., 2016. A unified evolutionary optimization procedure for single, multiple, and many objectives. IEEE Trans. Evol. Comput. 20, 358–369. https://doi.org/10.1109/TEVC.2015.2459718.
Shafii, M., De Smedt, F., 2009. Multi-objective calibration of a distributed hydrological model (WetSpa) using a genetic algorithm. Hydrol. Earth Syst. Sci. Discuss. 6, 243–271. https://doi.org/10.5194/hessd-6-243-2009.
Shafii, M., Tolson, B.A., 2015. Optimizing hydrological consistency by incorporating hydrological signatures into model calibration objectives. Water Resour. Res. 51, 3796–3814. https://doi.org/10.1002/2014WR016520.
Shrestha, R.R., Peters, D.L., Schnorbus, M.A., 2014. Evaluating the ability of a hydrologic model to replicate hydro-ecologically relevant indicators. Hydrol. Process. 28, 4294–4310. https://doi.org/10.1002/hyp.9997.
Shrestha, R.R., Schnorbus, M.A., Peters, D.L., 2016. Assessment of a hydrologic model's reliability in simulating flow regime alterations in a changing climate. Hydrol. Process. 30, 2628–2643. https://doi.org/10.1002/hyp.10812.
Sindhya, K., Miettinen, K., Deb, K., 2013. A hybrid framework for evolutionary multi-objective optimization. IEEE Trans. Evol. Comput. 17, 495–511. https://doi.org/10.1109/TEVC.2012.2204403.
The Nature Conservancy, 2009. Indicators of Hydrologic Alteration Version 7.1 User's Manual. The Nature Conservancy doi:10.1002/0470848944.
US EPA, 2011. A Primer on Using Biological Assessments to Support Water Quality Management. EPA 810-R-11-01. doi:10.1007/s13398-014-0173-7.2.
USGS, 2017. National Water Information System: Web Interface [WWW Document]. https://waterdata.usgs.gov/nwis (accessed 7.12.17).
van Griensven, A., Meixner, T., Grunwald, S., Bishop, T., Diluzio, M., Srinivasan, R., 2006. A global sensitivity analysis tool for the parameters of multi-variable catchment models. J. Hydrol. 324, 10–23. https://doi.org/10.1016/j.jhydrol.2005.09.008.
van Werkhoven, K., Wagener, T., Reed, P., Tang, Y., 2009. Sensitivity-guided reduction of parametric dimensionality for multi-objective calibration of watershed models. Adv. Water Resour. 32, 1154–1169. https://doi.org/10.1016/j.advwatres.2009.03.002.
Vigiak, O., Lutz, S., Mentzafou, A., Chiogna, G., Tuo, Y., Majone, B., Beck, H., de Roo, A., Malagó, A., Bouraoui, F., Kumar, R., Samaniego, L., Merz, R., Gamvroudis, C., Skoulikidis, N., Nikolaidis, N.P., Bellin, A., Acuňa, V., Mori, N., Ludwig, R., Pistocchi, A., 2018. Uncertainty of modelled flow regime for flow-ecological assessment in Southern Europe. Sci. Total Environ. 615, 1028–1047. https://doi.org/10.1016/j.scitotenv.2017.09.295.
Vis, M., Knight, R., Pool, S., Wolfe, W., Seibert, J., 2015. Model calibration criteria for estimating ecological flow characteristics. Water (Switzerland) 7, 2358–2381. https://doi.org/10.3390/w7052358.
Vörösmarty, C.J., McIntyre, P.B., Gessner, M.O., Dudgeon, D., Prusevich, A., Green, P., Glidden, S., Bunn, S.E., Sullivan, C.A., Liermann, C.R., Davies, P.M., 2010. Global threats to human water security and river biodiversity. Nature 467, 555–561. https://doi.org/10.1038/nature09440.
Wenger, S.J., Luce, C.H., Hamlet, A.F., Isaak, D.J., Neville, H.M., 2010. Macroscale hydrologic modeling of ecologically relevant flow metrics. Water Resour. Res. 46, 1–10. https://doi.org/10.1029/2009WR008839.
While, L., Bradstreet, L., Barone, L., 2016. Walking Fish Group: Hypervolume Project [WWW Document]. http://www.wfg.csse.uwa.edu.au/hypervolume/ (accessed 10.10.17).
While, L., Bradstreet, L., Barone, L., 2012. A fast way of calculating exact hypervolumes. IEEE Trans. Evol. Comput. 16, 86–95. https://doi.org/10.1109/TEVC.2010.2077298.
Woznicki, S.A., Nejadhashemi, A.P., Abouali, M., Herman, M.R., Esfahanian, E., Hamaamin, Y.A., Zhang, Z., 2016a. Ecohydrological modeling for large-scale environmental impact assessment. Sci. Total Environ. 543, 274–286. https://doi.org/10.1016/j.scitotenv.2015.11.044.
Woznicki, S.A., Nejadhashemi, A.P., Ross, D.M., Zhang, Z., Wang, L., Esfahanian, A.-H.H., 2015. Ecohydrological model parameter selection for stream health evaluation. Sci. Total Environ. 511, 341–353. https://doi.org/10.1016/j.scitotenv.2014.12.066.
Woznicki, S.A., Nejadhashemi, A.P., Tang, Y., Wang, L., 2016b. Large-scale climate change vulnerability assessment of stream health. Ecol. Indic. 69, 578–594. https://doi.org/10.1016/j.ecolind.2016.04.002.
Yang, H.C., Suen, J.P., Chou, S.K., 2016. Estimating the ungauged natural flow regimes for environmental flow management. Water Resour. Manage. 30, 4571–4584. https://doi.org/10.1007/s11269-016-1437-0.
You, G.J.Y., Thum, B.H., Lin, F.H., 2014. The examination of reproducibility in hydro-ecological characteristics by daily synthetic flow models. J. Hydrol. 511, 904–919. https://doi.org/10.1016/j.jhydrol.2014.02.047.
Zhang, Y., Shao, Q., Zhang, S., Zhai, X., She, D., 2016. Multi-metric calibration of hydrological model to capture overall flow regimes. J. Hydrol. 539, 525–538. https://doi.org/10.1016/j.jhydrol.2016.05.053.