University Library

MINERVA ACCESS

A gateway to Melbourne's research publications

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
Su, CH;Peterson, TJ;Costelloe, JF;Western, AW

Title:
A synthetic study to evaluate the utility of hydrological signatures for calibrating a base flow separation filter

Date:
2016-08-01

Citation:
Su, C. H., Peterson, T. J., Costelloe, J. F. & Western, A. W. (2016). A synthetic study to evaluate the utility of hydrological signatures for calibrating a base flow separation filter. Water Resources Research, 52 (8), pp.6526-6540. https://doi.org/10.1002/2015WR018177.

Persistent Link:
https://hdl.handle.net/11343/291578

A synthetic study to evaluate the utility of hydrological signatures for calibrating a baseflow separation filter

Chun-Hsu Su,[1] Tim J. Peterson,[1] Justin F. Costelloe,[1] Andrew W. Western[1]

Corresponding author: Chun-Hsu Su, Department of Infrastructure Engineering, University of Melbourne, Victoria 3010, Australia (csu@unimelb.edu.au)

[1]Department of Infrastructure Engineering, University of Melbourne, Victoria 3010, Australia.

Accepted Article

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version record. Please cite this article as doi:10.1002/2015WR018177.

This article is protected by copyright. All rights reserved.

Abstract. Estimation of baseflow from streamflow hydrographs has been a major challenge in hydrology for decades, leading to developments of baseflow separation filters. When without tracer or groundwater data to calibrate the filters, the standard approach to apply these filters in practice involves some degrees of subjectivity in choosing the filter parameters. This paper investigates the use of signature-based calibration in implementing baseflow filtering by testing seven possible hydrological signatures of baseflow against modelled daily baseflow produced by *Li et al.* [2014] for a range of synthetic catchments simulated with HydroGeoSphere. Our evaluation demonstrates that such a calibration method with few selected signatures as objectives is capable of calibrating a filter – Eckhardt filter – to yield satisfactory baseflow estimates at daily, monthly and long-term time scales, outperforming the standard approach. The best performing signatures can be readily derived from streamflow timeseries. While their performance depends on the catchment characteristics, the catchments where the signature method performs can be distinguished using commonly-used descriptors of flow dynamics.

Accepted Article

This article is protected by copyright. All rights reserved.

1. Introduction

Determination of the baseflow component of streamflow is a standard procedure in hydrology, and a valuable input for water resources management, flood frequency analysis, and for informing hydrological processes. Accordingly, many automated baseflow separation procedures have been developed [e.g. *Pettyjohn and Henning*, 1979; *Lyne and Hollick*, 1979; *Sloto and Crouse*, 1996; *Wittenberg*, 1999; *Van Dijk*, 2010], and some of the most common or physically-based methods rely on recursive digital filtering (RDF) [*Boughton*, 1993; *Chapman and Maxwell*, 1996; *Furey and Gupta*, 2001; *Huyck et al.*, 2005; *Eckhardt*, 2005; *Croke*, 2010; *Su et al.*, 2016]. To apply a filter to a streamflow hydrograph to retrieve a baseflow timeseries, the filter's parameters, ranging from one to five, need to be specified. For example, the maximum BFI (baseflow index) parameter of *Eckhardt* [2005] filter is assigned with considerable uncertainty based on flow regimes and aquifer types [*Eckhardt*, 2008]. The time delay parameter of *Furey and Gupta* [2001] filter is not objectively specified. The recession constant parameter of these two filters is determined through recession analysis of streamflow hydrographs, but the recession constants can vary considerably along the recession curves that cannot be identified with certainty [*Stoelzle et al.*, 2013].

This has led to increasing use of auxiliary datasets, such as geochemical or tracer data [*Stewart et al.*, 2007; *Zhang et al.*, 2013; *Gonzales et al.*, 2009; *Rimmer and Hartmann*, 2014] and groundwater levels [*Peters and Van Lanen*, 2005], to assist in the calibration of the filters. But these approaches are rarely done as they are limited by data scarcity in many catchments and their own large uncertainties due to complexity of hill

Accepted Article

This article is protected by copyright. All rights reserved.

slope hydrochemistry (distinction between and number of end-members [*Cartwright et al.*, 2014], and old-water paradox [*McDonnell et al.*, 2010]), and spatial heterogeneity.

Baseflow has traditionally been estimated using only streamflow data and further means of better using these data in the calibration of baseflow separation filters are of great practical value. When other independent datasets relating to groundwater dynamics are unavailable, the improved use of streamflow data provides a potentially attractive avenue for improving baseflow filter calibration. In particular, our understanding of baseflow dynamics in a catchment can be formalized as hydrological signatures, which represent typical, or expected, patterns of baseflow that can be quantitatively described. Examples of hydrological signatures are the limited rate of rise of the baseflow [*Peters and Van Lanen*, 2005] and lagged response of baseflow to causative rainfall [*Chapman*, 1999]. They are hydrologically meaningful patterns about the baseflow dynamics and in this work, our approach is to calibrate the RDF on these signatures that are defined in absence of auxiliary observations on baseflow. The intuition behind using hydrological signatures in RDF calibration is that a calibrated solution that manifests potential attributes of a true baseflow is closer to the truth than a solution that relies on arbitrary or poorly defined parameter values. This mirrors the approach of using hydrological signatures that reflect the functional behaviour of a catchment to evaluate or calibrate a watershed model [e.g. *Shafii and Tolson*, 2015]. Within this context, this work aims to provide insights to the following two questions: *Which hydrological signatures are relevant to calibrating a filter, and under what circumstances? Which hydrological signatures are most practically useful in terms of broad applicability to a wide range of catchments?*

Accepted Article

This article is protected by copyright. All rights reserved.

This study uses numerical experiments to test the utility of the hydrological signatures, which have been identified in the literature, for calibrating a baseflow filter. The filter of *Eckhardt* [2005] is chosen as a candidate method that is widely used by practitioners and has two a priori unknown parameters. The use of a hydrological model to provide synthetic hydrographs with known (modelled) baseflow component allows the definition of a 'truth' dataset to compare against baseflow filter outputs and is an approach effectively used in hydrology and hydrogeology [*Szilagyi*, 2004; *Fenicia et al.*, 2006; *Ferket et al.*, 2010; *Partington et al.*, 2012; *Li et al.*, 2013; *Shapoori et al.*, 2015; *Su et al.*, 2016]. The model can only provide a much simpler system than real-world catchments but allows a robust and repeatable test. If the approaches applied to the baseflow filters cannot replicate the outputs of a synthetic model, they are unlikely to work in practice. In our proof-of-concept experiments, we use simulation data from HydroGeoSphere (HGS), a fully integrated 3-dimensional surfacewater/groundwater physical model [*Therrien et al.*, 2009] produced by *Li et al.* [2014]. The model was run with a range of physical catchment characteristics and different hydrological forcing inputs, to simulate a range of baseflow dynamics. They are limited to temperate and tropical climates without modelling snowmelt and freeze/thaw soil processes; as such, this study similarly focuses on baseflow that is dominated by groundwater discharge. By evaluating the estimated baseflow against modelled baseflow, we distinguish the characteristics of the catchments where certain signatures are shown useful. Particular emphasis is directed at the measurable descriptors of the catchments, which will aid practical implementation of signature-based calibration of baseflow filters at real-world catchments.

Accepted Article

This article is protected by copyright. All rights reserved.

2. Method

2.1. Eckhardt filter

The Eckhardt filter is an archetype of the several RDFs proposed for estimating baseflow from daily streamflow hydrographs [e.g *Su et al.*, 2016]. The Eckhardt filter relates input streamflow timeseries y to baseflow timeseries through a recursive relation [*Eckhardt*, 2005],

$$\hat{b}_n = \min\Big[y_n, \frac{\alpha(1-\beta)}{1-\alpha\beta}\hat{b}_{n-1} + \frac{\beta(1-\alpha)}{1-\alpha\beta}y_n\Big]. \tag{1}$$

where $n = 1, 2, ..., N$ is the time step number, the filter parameters $\Theta = [\alpha, \beta]$ are generally assumed to be time-invariant, and the hat notation is adopted to distinguish baseflow estimates $\hat{b}$ from true baseflow values b. Thresholding exists to ensure that the estimated baseflow does not exceed streamflow, with an implicit assumption that streamflow measurements have no error. Here we follow *Furey and Gupta* [2001] to define the initial baseflow values as the minimum value of streamflow over the filtering period.

$\alpha \in [0, 1]$ is the recession constant because when quickflow q (i.e., $q = y - b$) is absent, Eq. 1 reduces to $\hat{b}_n = \alpha\hat{b}_{n-1}$ corresponding to a linear storage-discharge relation. $\beta \in [0, 1]$ is the maximum baseflow index. There are two limits to identify. First, as β tends to 1, $\hat{b}_n = y_n$, regardless of the value of α. Second, as α tends to 1, $\hat{b}_n = \hat{b}_{n-1}$ meaning that there is an increasing long-term memory and reduced variability in the estimated baseflow. This work considers simultaneous determination of α and β during the hydrological signature-based calibration process so that the cross covariances between the parameters are taken into consideration.

2.2. Simulated hydrological data

Accepted Article

This article is protected by copyright. All rights reserved.

Li et al. [2014] used the HydroGeoSphere (HGS) model to produce a test data set of synthetic daily streamflow and baseflow, for evaluating different baseflow filters. This data set is used in our proof-of-concept experiment to test the utility of the signature-based calibration. The fully integrated surfacewater/groundwater model describes several physical processes in three-dimensional catchments, including 3D saturated and unsaturated groundwater flow, overland flow, streamflow, infiltration, and exfiltration. The 2D St. Venant equations under diffusion wave approximation, 3D modified Richard's equation, and head continuity equation were solved simultaneously to simulate coupled surface and subsurface flow at each time steps, respectively [*Therrien et al.*, 2009]. As a post-processing step of the HGS simulation, the simulated streamflow at each stream cell is interpreted by the hydraulic mixing-cell method of *Partington et al.* [2011] to distinguish groundwater, overland flow, unsaturated flow and preferential flow entering or leaving the cell, and rainfall contribution to and evaporative loss from the storage within the cell. This enables the groundwater component of streamflow at each simulation time step in each cell of the stream to be determined, and in this study, this component is regarded as the baseflow.

The HGS models are a tilted V-shape catchment with uniform soils above an impermeable basement at 20 m depth. 66 representative catchments were set up with different combinations of rainfall and evapotranspiration (PET) forcings and model parameter values [*Li et al.*, 2014]. To generate 10-year simulation data, the catchments were forced by 10-year rainfall and potential evapotranspiration (PET) measurements from five Australian cities, namely Adelaide, Brisbane, Darwin, Melbourne and Sydney, which represent temperate, sub-tropical and tropical climate conditions. The daily time step of the sim-

Accepted Article

This article is protected by copyright. All rights reserved.

ulation is stipulated by the temporal (daily) resolution of the input forcing data. Latin Hypercube sampling strategy was used to randomly select model parameter values that control the geometry, including catchment area (A), hill slope ($S1$), channel slope ($S2$) and aspect ratio (AR). The catchment area is related to aquifer storage capacity that sustains baseflow generation, while the slopes influence the partitioning of rainfall to groundwater recharge and channel flow, and the rate at which soil water moves downslope. The aspect ratio, defined as the ratio between the catchment width to its length, affects the travel time of surface runoff and baseflow to the channel outlet. Other model parameters, namely saturated hydraulic conductivity K_s and van Genuchten parameters α_{vG} and β_{vG}, are also randomly sampled to vary soil properties between sand, sandy loam and loamy sand. Tab. 1 tabulates the annual means of forcing inputs and the model parameter values used in setting up the 66 catchments. Readers are referred to *Li et al.* [2014] for further details of the HGS simulation.

The catchments show a wide spectrum of discharge and baseflow dynamics as shown in Fig. 1, although they are predominantly perennial, with 64 catchments having continuous flow almost throughout the entire simulation period (i.e., $y_n > 0$ for $> 99.8\%$ of the period). The figure identifies 19 catchment descriptors, which include the HGS model parameters ($K_s, \alpha_{vG}, \beta_{vG}, S_1, S_2$, AR and A), BFI, QCV (quartile coefficient of variation) of daily [QCV(b)] and monthly (QCV[$m(b)$]) baseflow, model forcing statistics (mean annual rainfall, PET, and their ratio), runoff coefficient (RC), QCV of daily [QCV(y)] and monthly (QCV[$m(y)$]) streamflow, Hurst coefficients of baseflow $H(b)$ and of streamflow $H(y)$ (see later), and Baker-Richards flashiness index of streamflow F_{RB} (see later). Observe that RC of the 66 catchments vary between 4.6E-4 and 0.37 and the BFI, QCV(b)

Accepted Article

This article is protected by copyright. All rights reserved.

and QCV(y) show a full range from near zero to unity. These descriptors are divided into two categories, as partitioned by a dashed line, based on whether they can be estimated easily a priori. For instance, there are existing methods for estimating catchment area from a digital elevation map [*Mackay and Band*, 1998] and for daily rainfall and PET from satellite remote sensing [*Simpson et al.*, 1996; *Smith et al.*, 2007; *Kim and Hogue*, 2008].

2.3. Standard practice

The Eckhardt recession constant α is typically estimated using a recession analysis [e.g., *Tallaksen*, 1995], where a single estimate of the recession constant is defined as $\alpha_n = y_{n+1}/y_n$ during recession periods T_R. But α_n varies with n due to measurement uncertainty, nonlinear storage-outflow relationships, multiple baseflow pathways, and seasonal changes in processes [*Brutsaert and Nieber*, 1977; *Vogel and Kroll*, 1992; *Kirchner*, 2009; *Stoelzle et al.*, 2013]. *Eckhardt* [2005, 2008] recommended using the upper bound of the recession constant distribution to define a single α estimate to represent baseflow driven by groundwater discharge and this approach has been widely used in practice [e.g. *Gonzales et al.*, 2009; *Li et al.*, 2014; *Costelloe et al.*, 2015]. For comparing against the signature-based calibration scheme (Sec. 2.4), we regard this as the standard approach.

The second parameter β cannot be determined from the streamflow data alone, unless the recession characteristics are assumed to extrapolate forward and backward in time beyond the recession periods [*Linsley et al.*, 1958; *Wittenberg*, 1999; *Van Dijk*, 2010]. Eckhardt (2005) proposed values of β based on the hydrological and hydrogeological characteristics of the catchments, when tracer data are unavailable. Specifically, $\beta = 0.7 - 0.8$ for perennial streams on porous aquifers and ~ 0.25 for those on hard rock aquifers.

Accepted Article

This article is protected by copyright. All rights reserved.

Ephemeral streams on porous aquifers are expected to have $\beta = 0.2 - 0.5$. Accordingly, the standard approach entails the use of β values of 0.80 for the 64 perennial HGS catchments, and 0.50 for the remaining 2 ephemeral catchments. The HGS model structure and the chosen soil parameter values (corresponding to sand through to sandy loam) in *Li et al.* [2014] would generally be regarded as being representative of porous aquifers.

2.4. Hydrological signatures of baseflow

Distinguishing from the standard practice, a calibration exercise of the baseflow filter parameters is based on improving characteristics of the filter estimates $\hat{b}$, such as matching $\hat{b}$ to values independently derived from geochemical [e.g. *Rimmer and Hartmann*, 2014] or groundwater measurements [*Peters and Van Lanen*, 2005]. Here we consider the typical case where such baseflow measurements are not available at the catchment scale; we investigate the possibility of calibrating the filter based on requiring $\hat{b}$ to show hydrologically meaningful dynamics. In the following, we review some of these hydrological signatures of baseflow existing in the literature and used them to define calibrating metrics that capture hydrologically relevant aspects of the baseflow for the plausible estimations of baseflow. They are chosen in the view that the filter is commonly applied to daily streamflow records.

A widely accepted property of baseflow is that streamflow is entirely made up of baseflow during recession periods T_R. There are many algorithms that select recession segments from a streamflow hydrograph by defining starting and duration criteria for these segments (e.g., see the review by *Tallaksen* [1995]), and different algorithms can lead to different recession analyses. The practices of the recession analysis and baseflow filtering are further complicated by the fact that streamflow measurements are inaccurate by 5–10% [*Hirsch*

Accepted Article

This article is protected by copyright. All rights reserved.

and Costa, 2004]. Here we avoid introducing this uncertainty to our study by using modelled baseflow b to define T_R. With T_R selected, the difference between the baseflow estimates $\hat{b}$ and observed recession segments is used to characterise this aspect of baseflow,

$$S_{rec} = \frac{\mathrm{E}_{T_R}[(\hat{b} - y)^2]}{\mathrm{var}_{T_R}(y)}, \tag{2}$$

to vary from $S_{rec} \in [0, \infty)$. $\mathrm{E}_{T_R}(\circ)$ and $\mathrm{var}_{T_R}(\circ)$ compute the expectation and variance of the data from T_R only.

Furey and Gupta [2001] proposed two criteria to diagnose the performance of a baseflow filter. The first criterion is the percentage of baseflow estimates that violates the mass conservation condition of $\hat{b} \leq y$. The second criterion assumes that the annual minimum value of streamflow is predominantly baseflow. Thus when an annual minimum flow occurs at time n, the deviation $|\hat{b}_n - y_n|/y_n$ should be near zero. These two criteria are only relevant for a separation method that allows for $\hat{b} > y$; thus they are not useful for calibrating the Eckhardt filter.

Peters and Van Lanen [2005] suggest maximizing baseflow while minimizing the rate of rise of the baseflow. One implementation of this concept is to reduce the variability of baseflow timeseries by minimizing its QCV using the following calibrating metric,

$$S_{QCV} = \frac{Q_3(\hat{b}) - Q_1(\hat{b})}{Q_3(\hat{b}) + Q_1(\hat{b})}, \tag{3}$$

where $Q_1(b)$ computes the first quartile (25% percentile) of the baseflow, and $Q_3(b)$ its third quartile (75% percentile). This metric, ranged $[0, 1]$, is expected to be suitable for streams with slow discharge from groundwater stores. An alternative calibrating metric is to characterize the long-range memory or strong short-term dependence in baseflow

Accepted Article

This article is protected by copyright. All rights reserved.

timeseries using [*Hurst*, 1951; *O'Connell*, 1971]

$$S_{hurst} = 1 - H(\hat{b}). \tag{4}$$

where $H(\hat{b})$ returns the Hurst coefficient – a measure of persistence – of the baseflow, estimated via the rescaled range analysis of *Hurst* [1951].

Nathan and McMahon [1990] regard baseflow as a low-frequency signal in a streamflow timeseries. Accordingly, the Eckhardt filter operates as a low-pass filter that attenuates the high-frequency signal in the streamflow data [*Eckhardt*, 2005]. This is compatible with findings by *Zhang and Schilling* [2004] that the power spectral density of the observed hydraulic head fluctuations decreases linearity with frequency in the log-log scale. Such conceptualization can be translated into a calibrating metric that encourages the baseflow hydrographs to reflect the low-frequency variations in streamflow at a given time scale, which can be assumed to be of a seasonal time scale,

$$S_{low} = \frac{\mathrm{E}[(\hat{b} - y_{low})^2]}{\mathrm{var}(y_{low})}, \tag{5}$$

which varies between $[0, \infty]$. y_{low} is taken to be 6-month moving median of y that subjects to $y_{low} \leq y$. The median is taken rather than the mean because it is less sensitive to peak flows. Other moving window sizes, namely 3- and 12-month, were trialled but 6-month window yields better calibration outcomes (more later).

Next, baseflow has often been conceptualised by its delayed response to rainfall, distinguishing it from the quickflow [*Jakeman and Hornberger*, 1993]. The temporal association between precipitation (P) and these two flow constituents can be measured by their time-lag cross correlations $R(P, \hat{b}; \tau)$ and $R(P, \hat{q}; \tau)$ as functions of lag τ, where $\hat{q}$ is estimated quickflow ($\hat{q} = y - \hat{b}$). As such, the slow-vs-quick response will lead to a contrast between

Accepted Article

This article is protected by copyright. All rights reserved.

their lagged correlations. A more immediate response of quickflow to precipitation, within the time of cessation (T_C) of surface runoff and interflow at the catchment outlet, will manifest as $R(P, \hat{q}; \tau) > R(P, \hat{b}; \tau)$ for $\tau \in [-T_C, 0]$. Beyond this period, if groundwater recharge and increased discharge occurs, the delayed response of baseflow to rainfall will manifest as the opposite inequality, $R(P, \hat{q}; \tau) < R(P, \hat{b}; \tau)$ for $\tau < -T_C$. For simplicity, we use these correlation statistics at $\tau = 0$ to formalise the above dichotomy,

$$S_{quick} = 1 - |R(P, \hat{q}; \tau = 0)|, \tag{6}$$

$$S_{slow} = |R(P, \hat{b}; \tau = 0)|. \tag{7}$$

In other words, $S_{quick} \in [0, 1]$ corresponds to maximizing the correlation between precipitation and estimated quickflow at zero lag, under the assumption that precipitation on a catchment is coincident with interflow and/or runoff generation within the data sampling time interval. By contrast, $S_{slow} \in [0, 1]$ assumes that baseflow is not coincident. Hence, given that the Eckhardt filter effects time delays between streamflow and estimated baseflow timeseries, these metrics will inform the β value during the calibration process.

Despite the explicit choice of $\tau = 0$ here, other values of $\tau < 0$ could be used as a pre-defined input or another calibration objective to take into account a delayed response of quickflow to precipitation. However, these expressions are only useful at catchments where these statistics provide a distinction between the two flow components. Notably, these metrics will be problematic for basins with long routing times, losing streams, or persistent interflow during wet seasons. In particular, runoff and interflow can be delayed by routing and bank storage, lengthening their response time delay, while quick infiltration and quick discharge from shallow aquifers can shorten baseflow response time, leading to less distinctive lag-correlation characteristics by overlapping $R(P, q; \tau)$ and $R(P, b; \tau)$,

Accepted Article

This article is protected by copyright. All rights reserved.

where q and b denote the true flow components. In other cases where there is no runoff, $R(P, q; \tau)$ will be weak for all τ.

Finally, *Schoonover et al.* [2006] and *Deelstra et al.* [2014] have shown that the watersheds with high flashiness in streamflow have corresponding low BFI. Given that they capture opposing flow processes (fast c.f. slow processes), this is expected. One common indicator for flashiness is the Richards-Baker flashiness index (F_{RB}) that is derived by dividing the total path length of daily streamflow by the sum of average daily streamflow [*Baker et al.*, 2004]. To use this property in filter calibration, we use F_{RB} to provide a priori estimate of BFI and define the following calibrating metric,

$$S_{BFI} = \left|\frac{\mathrm{E}(\hat{b})}{\mathrm{E}(y)} - \max(0, 1 - F_{RB})\right|, \tag{8}$$

where $\max(\circ)$ returns the largest element in the bracket. The flashiness index of modelled streamflow from the HGS varies between 0.01 and 1.11 (Fig. 1). There is a strong inverse relationship between BFI and F_{RB}; their linear correlation is -0.98 and the relationship is close to 1:1 line such as the regression fit leads to BFI $= 1.04 - 0.96 F_{RB}$ with $R^2 = 0.97$.

2.5. Multiple signature-based calibration strategy

Multiple metrics are combined to construct single objective functions to capture different properties of the flow components in the following seven calibration setups ρ_i,

$$\rho_1 = S_{rec}, \tag{9}$$

$$\rho_2 = \max(S_{rec}, S_{QCV}), \tag{10}$$

$$\rho_3 = \max(S_{rec}, S_{hurst}), \tag{11}$$

$$\rho_4 = \max(S_{rec}, S_{low}), \tag{12}$$

$$\rho_5 = \max(S_{rec}, S_{quick}), \tag{13}$$

Accepted Article

This article is protected by copyright. All rights reserved.

$$\rho_6 = \max(S_{rec}, S_{slow}), \tag{14}$$

$$\rho_7 = \max(S_{rec}, S_{BFI}). \tag{15}$$

The maximum of the constituent calibrating metrics is taken as the means to arrive at a scalar objective function for optimization [e.g *Vrugt and Sadegh*, 2013]. For each calibration setup ρ_i, the Eckhardt α and β parameters are tuned to minimize ρ_i. The objective ρ_1 illustrates a conservative case where the recession limbs are dominated by baseflow. ρ_2, ρ_3 and ρ_4 parametrize the low variability in baseflow in three different manners. ρ_5 and ρ_6 capture the different temporal responses of baseflow and quickflow to rainfall, while ρ_7 uses a priori BFI estimate to condition the filter parameter values. S_{rec} is used in $\rho_2 - \rho_7$ because it would appear to provide information to condition the value of α. It is also necessary in defining ρ_2, ρ_6 and ρ_7 because using S_{QCV}, S_{quick} or S_{slow} alone leads to $\beta = 0$ (i.e., zero baseflow) as a trivial solution. Importantly, in practical application terms, $\rho_1 - \rho_4$ and ρ_7 rely only on streamflow data. ρ_5 and ρ_6 utilize streamflow and rainfall data; latter is now commonly available owing to satellite remote sensing [*Simpson et al.*, 1996; *Smith et al.*, 2007].

In addition to these calibration scenarios, we include a base case as a way to acknowledge the fact that the filter has structural limitations to reproduce modelled baseflow completely, even when the true b was used in its calibration [*Li et al.*, 2014]. Here the base case uses the synthetic 'truth' b in the calibration by matching $\hat{b}$ to b using the error measure $\mathrm{E}[(\hat{b}-b)^2]/\mathrm{var}(b)$ as the objective function. The base case defines the benchmark for what could be achievable with the Eckhardt filter, enabling appropriate attribution of skills to the signature-based calibration scheme.

Accepted Article

This article is protected by copyright. All rights reserved.

2.6. Two-fold cross validation

For each HGS catchment, its 10-year simulation data is partitioned into two 5-year datasets. The filter is trained on the first 5-year training data, and evaluated on the second data, and vice versa. Therefore, with 66 HGS catchment simulations, there are 132 evaluation samples in total. The candidate methods evaluated include the standard approach (Sec. 2.3) and the signature method with the various objectives ρ_i (Sec. 2.4). The following three statistical scores are used,

$$\mathrm{NSE}_d = 1 - \frac{\mathrm{E}_{T_p}[(\hat{b} - b)^2]}{\mathrm{var}_{T_p}(b)}, \tag{16}$$

$$\mathrm{NSE}_s = 1 - \frac{\mathrm{E}_{T_s}[(\hat{b} - b)^2]}{\mathrm{var}_{T_s}(b)}, \tag{17}$$

$$\mathrm{NSE}_m = 1 - \frac{\mathrm{E}\{[m(\hat{b}) - m(b)]^2\}}{\mathrm{var}[m(b)]}, \tag{18}$$

$$\mathrm{PBIAS} = \frac{\mathrm{E}(\hat{b}) - \mathrm{E}(b)}{\mathrm{E}(b)} \times 100\%, \tag{19}$$

where $\hat{b}$ still refers to the filter-estimated baseflow and b is the synthetic true baseflow from the model, and $m(\circ)$ computes the monthly accumulated baseflow. NSE_d (Nash-Sutcliffe efficiency) assesses the baseflow estimates at daily time steps, NSE_s assesses the estimates during high-flow or storm event periods, whereas NSE_m provides an indication of the accuracy at the monthly time scale. PBIAS provides a measure of an overall over- or under-estimation of baseflow over the 5-year period. To calculate NSE_d, only data during rainfall periods T_p are used, in accordance to *Partington et al.* [2012] and *Li et al.* [2014]. This is because the comparison between simulated and estimated baseflow is not meaningful during recession periods due to the thresholding condition in Eq. 1. The measure NSE_s is computed on periods where streamflow is greater than its 80th percentile. Further, we consider categorical performance with NSE > 0.5 and $|\mathrm{PBIAS}| <$

Accepted Article

This article is protected by copyright. All rights reserved.

25% as good [*Partington et al.*, 2012], and regard performance with $0 < \text{NSE} \leq 0.5$ and $25\% \leq |\text{PBIAS}| < 50\%$ as average, otherwise poor.

3. Results and analysis

3.1. Performance evaluation of filter calibration methods

Tab. 2 shows the percentages of the 132 evaluation samples, where the performance of a candidate method is considered good, average or poor. Recall that the base case is included to illustrate the achievable performance of the Eckhardt filter. The base-case calibration leads to good PBIAS scores at over 96% of the samples, and average PBIAS for the remaining samples, for both calibration and validation data The filter's monthly baseflow estimates show good or average skill ($\text{NSE}_m > 0$) at over 90% of the samples. The filter however shows more limited performance at the daily timescale and during high flows, with $\text{NSE}_d > 0$ and $\text{NSE}_s > 0$ for about 40–60% of the samples.

The signature method yields better performance than the standard approach. With the standard approach, only 7–23% of the samples show good-average NSE_d and NSE_m values > 0 and 72% with average or above $|\text{PBIAS}| \leq 50\%$. By contrast, the signature method with the objective ρ_4, which encourages the baseflow hydrograph to reflect the seasonal variations of streamflow, yields the best performance for NSE_d, NSE_m and PBIAS scores compared to the other objectives. Over 60% of the samples show good or average NSE_d, and there are significant numbers of the samples ($\sim$91% and $\sim$71%) show good PBIAS and NSE_m. Its categorical performance is comparable with that of the base case. In fact, we find good agreement in individual NSE_d values between ρ_4 and the base case for samples with lower BFI< 0.8. Their discrepancy becomes apparent in the higher BFI> 0.8 regime, suggesting that the use of 6-month moving window to define S_{low} (Eq. 5) is too

Accepted Article

This article is protected by copyright. All rights reserved.

wide for these catchments. The use of a narrower 3-month window improves upon this by lifting the percentages of the samples with good calibration and validation NSE_d by 2%. However, this occurs at the expense of reducing the numbers of catchments where it shows good PBIAS and NSE_m by 7–10% (not shown). There is also striking agreement between the skills of ρ_4 and the base case in terms of NSE_m and PBIAS values, particularly amongst samples with BFI$<$ 0.9.

The objective ρ_7, which uses F_{RB}-based BFI estimate, shows promising but less remarkable skills in NSE_d, NSE_m and PBIAS compared to ρ_4. It leads to average or higher NSE_d for about 44% of the evaluation samples, and over 56% and 76% for NSE_m and PBIAS scores, respectively. Despite the fact that F_{RB} provides a good estimate of BFI, it is not sufficient to produce good daily or monthly baseflow values in most samples. As a way to confirm this, we replace F_{RB} in S_{BFI} (Eq. 8) with the true BFI values (directly derived from modelled baseflow) in the signature calibration. This results in superior PBIAS scores that are comparable with ρ_4, but at the expense of diminished NSE skills (not shown).

Further, ρ_7 outperforms ρ_4 during high flows and show NSE_s scores are closer to those in the base case, with about a third of the samples showing good or average skill. Since that both ρ_4 and ρ_7 are skilful but for different aspects of baseflow, an aggregated objective $\max(S_{rec}, S_{low}, S_{BFI})$ was trialled. This led to 31% of the catchments with $\text{NSE}_s > 0$ (c.f. 22% for ρ_4), but diminished skills for the other measures (e.g., 27% and 7.6% of catchments show poor NSE_m and PBIAS) relative to ρ_4.

Accepted Article

This article is protected by copyright. All rights reserved.

The measure NSE_s provides a distinction between the categorical performance between ρ_7 and ρ_6, which are similar in other respects. The objective ρ_3 that reflects persistence in baseflow outperforms all but ρ_4 in terms of PBIAS and NSE_m.

The mere description of baseflow with the recession characteristics in the objective ρ_1 (i.e., S_{rec}) is clearly not sufficient to calibrate the filter for most samples. This is because in the limit of $\beta \to 1$ (regardless of the value of α), $\hat{b} \to y$ and S_{rec} tends to zero. In other words, the objective is only suitable for catchments with very high BFI. The objective ρ_5 also shows poor overall performance compared to the other objectives.

While Tab. 2 suggests similarities in categorical performance between the calibration and validation data, their individual scores can differ significantly. A large contrast in the scores can suggest hydrological non-stationarity between the two time periods, deficiency of the filter, or the lack of robustness in the calibration methods. Therefore, to single out the influence of the last factor, the 70 cross-validation samples that satisfy the following conditions are selected to examine the score differences, namely that the calibration and validation data sets show small differences in BFI (less than 0.05), and mean daily rainfall (less than 0.5 mm), and that the filter can be calibrated to achieve $\text{NSE}_d > 0$ in the base case. Tab. 3 calculates the percentages of these cross-validation samples where the scores on the calibration data are consistent with those of the validation data. The condition for judging consistency in the NSE scores (PBIAS) is based on having an absolute difference less than 0.2 (10%). We find that ρ_4 leads to consistent scores at over 94% of the samples for NSE_d, NSE_m and PBIAS, similar to the base case. The objective ρ_7 is more consistent than ρ_4 in terms of NSE_s (51% c.f. 37%), and shows robustness to some extents for other measures at over 70% of the samples.

Accepted Article

This article is protected by copyright. All rights reserved.

3.2. Differences in calibrated parameter values

Different objective functions in the signature method lead to different filter parameter values, as illustrated in Fig. 2 for ρ_4 and ρ_7. The optimized filter parameter values retrieved by the signature method are compared with the values derived in the base case. There are noticeable agreements between ρ_4-derived and the base-case values; their correlations in terms of α and β values are 0.44 ± 0.18 (95% confidence interval) and 0.52 ± 0.16, respectively. The relation does not exist for ρ_7, where the both correlations are not significant (p-value < 0.05).

The α values derived from the recession analysis (i.e., the standard approach) are also related to the base-case values but with a weaker correlation of 0.26 ± 0.20. The signature method captures a wider range of possible α values – similar to the base case – as it is not constrained by the conceptually-based upper percentile thresholds used in the standard practice. The ranges of β values retrieved by the base case and the signature method are also wider than the recommended values used in the standard approach.

3.3. Associating calibration performance with catchment characteristics

Fig. 3 illustrates the contrasts between the catchments where the signature method with ρ_4 or ρ_7 yield above average NSE scores and those where they do not. Readers are preferred to *Supplementary Materials* for the full set of boxplots for the other objectives. The differences are found amongst some of the 19 catchment descriptors identified in Sec. 2.2 and Fig. 1. Bearing in mind that the poor performance of the signature method at some catchments are partly attributed to the limitations of the Eckhardt filter.

Separating based on the lower or upper quartiles, the Eckhardt filter calibrated with ρ_4 (and in the base case) show $\mathrm{NSE}_d > 0$ at the catchments with BFI$\geq$ 0.20, RC$\leq$

Accepted Article

This article is protected by copyright. All rights reserved.

0.22, QCV$[m(y)] \leq 0.63$, $H(y) \geq 0.70$ and $F_{RB} \leq 0.59$ (Fig. 3(a)). The distinctions between the catchments in terms of the HGS model parameters are also apparent but less well defined. These catchments are associated with larger values of saturated hydraulic conductivity K_s, soil's van Genuchten β_{vG}, steeper slopes and lower aspect ratios. K_s has strong influence on BFI because larger K_s facilitate infiltration, leading to increased soil saturation and groundwater discharge [*Li et al.*, 2013]. At the same time, larger K_s values are associated with lower CQV(y) and flashiness, since there are sustained flows in the streams owing to groundwater discharge and/or small smaller temporal variations due to smaller surface runoff.

The use of the metric S_{low} in ρ_4 improves upon the objective ρ_1 by allowing the calibrated filter to perform better at 14–30 catchments that show lower BFI, QCV(b) and $H(y)$, higher QCV$[m(y)]$ and F_{RB}. This is similar when using other calibrating metrics in conjunction with S_{rec} but to variable (and lesser) extents.

The objective ρ_4 can also lead to skillful estimates of monthly baseflow values at significantly more catchments that are characterized by QCV$[m(y)] \leq 0.73$ and $F_{RB} \leq 0.92$, as shown in Fig. 3(c). The analysis for PBIAS is not shown because almost all the catchments (63) show above average PBIAS via using ρ_4.

The best objective, ρ_7, for baseflow estimates during high flows, leads to above average NSE$_s > 0$ at limited (22) catchments, with higher K_s values, RC ≤ 0.15, $H(y) \geq 0.72$, and $F_{RB} \leq 0.30$ (Fig. 3(b)). Most of these catchments are simulated with smaller rainfall forcings from Adelaide and Melbourne.

Accepted Article

This article is protected by copyright. All rights reserved.

4. Discussions

4.1. Improving the standard approach

The prescribed guidelines in the standard approach are not sufficiently flexible to reflect a wide range of flow behaviours found amongst the 66 HGS catchments (Fig. 1). We observed that there are many possible α and β values for the same catchment type under optimization (base case). This exemplifies the need for a better approach when tracer or groundwater data is absent to guide the calibration process. This work attempts to close this gap through using hydrological signatures of baseflow (and quickflow). Most of these signatures identified in Sec. 2.4 show relevance as they have led to calibrated filters that outperform the standard approach. In other words, they reflect some of the qualities of modelled baseflow observed in some of the catchments. The signature method therefore provides more capacity for the data to 'speak for itself', in comparison to choosing filter values based on catchment types.

4.2. Interpreting performances of the signatures

The utility of the signatures varies significantly across the HGS catchments and over different time scales. It is clear that certain signatures are only suitable at specific catchments; in particular, S_{rec} and S_{quick} are only relevant at less than 25% of the catchments at which there is little runoff (high BFI and low flashiness of streamflow). Combining S_{rec} with other metrics that reflect slow-varying behaviours (S_{QCV}, S_{low}) or persistence (S_{hurst}) in baseflow enables the signature method to perform at catchments with higher runoff. Amongst these metrics, S_{low} shows the best overall calibration outcomes at daily, monthly and long-term scales, and comes closest to the base-case performance. This suggests that a moving 6-month median statistics of streamflow hydrographs provide a

Accepted Article

This article is protected by copyright. All rights reserved.

reasonable representation of the seasonality of the baseflow at over 90% of the catchments. The variability measure QCV(b) and persistence measure $H(b)$ used in S_{QCV} and S_{hurst} are very weakly related; i.e., their correlation of -0.17 is insignificant at 95% confidence level. Hence they appear to capture different properties of the baseflow, leading to their contrasting performance at monthly and long-term scales.

The Eckhardt filter is a non-symmetric filter that naturally produces time shifts between streamflow and baseflow hydrographs during non-recession periods, and thus time delays between precipitation and baseflow timeseries. While one may therefore expect the minimization of the metric S_{slow} defined between coincident precipitation and baseflow can be easily satisfied, it has lead to skillful calibration outcomes at monthly and long-term time scales at more than 50% catchments.

The flashiness index F_{RB} shows surprisingly strong relationship with BFI amongst these synthetic catchments, although the simple linear relation used in defining S_{BFI} may not hold in more complex catchment setups (more later). However, the observed benefits of using S_{BFI} in estimating baseflow values during high-flow periods, which outperforms other signatures in this respect, reflects the deficiencies in the other signatures to properly reflect baseflow dynamics during these periods.

An accurate assessment of the utility of the signature method is undermined by the fundamental limitations of the Eckhardt filter in this study. In particular, the base-case performance of the Eckhardt filter is poor at 38–57% of the catchments at the daily time scale and during high-flow periods. NSE_s provides the most stringent measure that the Eckhardt filter is not suited for. Alternative RDF algorithms, which rely on physical equations or conceptualise many more flow pathways, offer more flexible filter structures

Accepted Article

This article is protected by copyright. All rights reserved.

with higher number of filter parameters (*Su et al.* [2016] and references therein). Importantly, the signature method that calibrates several filter parameters simultaneously can be easily applied to and tested on these RDFs.

Finally, the HGS data has enabled the objective and systematic evaluations at many catchments with diverse configurations and flow dynamics. Nevertheless, as the models represent approximations to reality, further evaluations with other catchment simulations and real-world catchments are warranted. In particular, the HGS models from *Li et al.* [2014] does not describe catchments with hard rock aquifers, high-latitude forcing and snowmelts. And, their data set does not show a variety of catchments with ephemeral streams. Further, the modelled catchments adopt simple symmetric topography (V-shape) with verifically uniform soils and possess uniformity in vegetation cover and geology. These should lead to more linear hydrological responses to meteorological forcing compared with real catchments. These may have contributed to the obvious strong linear relationship between F_{RB} and BFI. The simplified and uniform characteristics also limit the occurrence of other slow flow components, such as interflow and flows from bank storage, that occur more strongly in heterogeneous catchments. For instance, interflow will commonly occur where there are permeability contrasts at the base of the soil profile, and bank storage where there are contrasts between near river sediments and the regional unconfined aquifer.

4.3. Practical implications

Our study demonstrates that the objective ρ_4, which combines the signatures S_{rec} and S_{low}, yields promising and consistent calibration performance across short to long time scales, although the resultant baseflow estimates during high flows may be poor and

Accepted Article

This article is protected by copyright. All rights reserved.

biased. The objective can be practically implemented only with the streamflow record. It is duly noted that this work have not conducted a blinded recession analysis (required for calculating S_{rec} in practice) so that the generality of our results is not undermined by a particular choice of method for recession analysis. The sensitivity of S_{rec} to uncertainty in selecting recession limbs using different approaches is reserved for further work.

Acknowledging the limitations in the representativeness of the 66 HGS models, we distinguish the (modelled) catchments where the signature method can perform satisfactorily (above average NSE_d and NSE_m) based on their attributes, namely the RC, variability in monthly discharge QCV[$m(y)$], persistence in and flashiness of streamflow $H(y)$ and F_{RB}. Their values can also be easily derived from streamflow data and/or meteorological data and compared with the ranges identified in Sec. 3.3. This will inform the applicability of the method at a given catchment with perennial flows. The applicability of other objectives can also be determined similarly using these descriptors. Our results also show that a satisfactory estimate of BFI, which corresponds to an above average PIAS score, can retrieved for almost all the catchments using ρ_4.

5. Conclusions

This paper investigates the use of hydrological signatures of baseflow to calibrate a RDF. It aims to circumvent the subjectivity of a practitioner in choosing the filter parameters based on catchment and streamflow types, especially when the existing guidelines were established based on relatively small data samples. We applied the notion of signature-based calibration to RDF by testing seven prospective signatures on 66 synthetic catchments, leading to the following key conclusions. First, the signature method with particular sets of signatures, namely ρ_4 and ρ_7, is capable of calibrating the Eckhardt filter such that the

Accepted Article

This article is protected by copyright. All rights reserved.

calibrated filter is shown to perform satisfactorily at multiple time scales. In particular, ρ_4 is skillful in providing a good estimate of the long-term ($\geq$ 5-year) mean baseflow level at nearly all the catchments. Importantly, the signature method outperforms the standard approach and is a more objective and reproducible approach whose merits can be critiqued based on the suitability of the chosen signatures. Second, these signatures and the calibrating objectives can be computed with streamflow data and/or precipitation data, although the best performing signatures only require the former. Third, the performance of the signature method depends on the catchment characteristics, and the catchments compatible with the signature method can be distinguished by four descriptors (RC, QCV[$m(y)$], $H(y)$ and F_{RB}). The criteria are provided in Sec. 3.3. These descriptors are commonly used and can be easily derived to inform the practice of the method at a given site. For these reasons, the signature method is a promising and practical approach that can be easily implemented without the need for other independent observational data for the RDF calibration.

Further research is recommended on the following. There are opportunities to explore other ways of combining multiple signatures, since many of the identified signatures cannot be used alone as an objective. These signatures have different dynamic ranges and thus contribute to the objective function with varying influences. The sensitivity of the signature method to measurement errors in streamflow or ancillary (e.g., rainfall) data and uncertainty inherent to conducting recession analysis (e.g., for defining S_{rec}) should also be investigated. Moreover, given that the signature method has not reached the base-case performance in estimating baseflow during high-flow periods, it is prospective to explore other signatures that are relevant in this respect. Additionally, the utility of

Accepted Article

This article is protected by copyright. All rights reserved.

the signature method could be tested on other RDF algorithms with more than two filter parameters. Finally, it is clear that further evaluation is warranted, using other synthetic catchment simulations with different flow and climate conditions and by means of tracer-based separation at real catchments. These will provide further guidance for types of catchments where the signature method is satisfactory. The synthetic experiments could also provide an in-road to understanding the hydrological processes that may have led to the emergence of a baseflow signature and control its relevance.

Acknowledgments. The HGS streamflow and baseflow simulation data is available online as the Supplementary data to *Li et al.* [2014]; last accessed on April 30, 2016 via http://www.sciencedirect.com/science/article/pii/S1364815213003149. The rainfall and PET data were provided by L. Li through personal communication. We thank L. Li and H. Maier for their assistance with the data, and four anonymous reviewers and the associate editor for their valuable comments on the original manuscript. CHS thank Shelly Chua for valuable discussions. This research was conducted with financial support from the Australian Research Council (ARC Discovery Project No. DP120100253).

References

Baker, D. B., Richards, R. P., Timothy, T., Loftus, T. T., and Kramer, J. W. (2004), A new flashiness index: Characteristics and applications to midwestern rivers and streams, *J. Am. Water Resour. Assoc.*, 40, 503–522.

Boughton, W. C. (1993), A hydrograph-based model for estimating the water yield of ungauged catchments *In Hydrology and Water Resources Symposium*, Institution of Engineers Australia, Newcastle, NSW; 317-324.

Accepted Article

This article is protected by copyright. All rights reserved.

Brutsaert, W. and Nieber, J. L. (1977), Regionalized drought flow hydrographs from a mature glaciated plateau, *Water Resour. Res.*, *13*, 637–643.

Cartwright, I., Gilfedder, B., and Hofmann, H. (2014), Contrasts between estimates of baseflow help discern multiple sources of water contributing to rivers, *Hydrol. Earth Syst. Sci.*, *18*, 15–30.

Chapman, T. (1999), A comparison of algorithms for stream flow recession and baseflow separation, *Hydrol. Processes*, *13*, 701–714.

Chapman, T. G., and Maxwell, A. I. (1996), Baseflow separation comparison of numerical methods with tracer experiments. Institute Engineers Australia National Conference. Publ. 96/05, 539–545.

Costelloe, J., Peterson, T., Halbert, K., Western, A. W., and McDonnell, J. J. (2015), Groundwater surface mapping informs sources of catchment baseflow, *Hydrol. Earth Syst. Sci.*, *19(4)*, 1599–1613.

Croke, B.F.W. (2010), Exploring changes in catchment response characteristics: application of a generic filter for estimating the effective rainfall and unit hydrograph from an observed streamflow timeseries. In: BHS Third International Symposium: Role of Hydrology in Managing Consequences of a Changing Global Environment. Newcastle University, UK.

Deelstra, J., Iital, A., Povilaitis, A., Kyllmar, K., Greipsland, I., Blicher-Mathiesen, G., Jansons, V., Koskiaho, J., and Lagzdins, A. (2014), Hydrological pathways and nitrogen runoff in agricultural dominated catchments in Nordic and Baltic countries, *Agric. Ecosys. Environ.*, *198*, 65–73.

Accepted Article

This article is protected by copyright. All rights reserved.

Eckhardt, K. (2005), How to construct recursive digital filters for baseflow separation, *Hydrol. Processes*, *19*, 507–515.

Eckhardt, K. (2008), A comparison of baseflow indices, which were calculated using several different baseflow separation methods, *J. Hydrol.*, *352*(1-2), 168–173.

Eckhardt, K. (2012), Technical note: Analytical sensitivity analysis of a two parameter recursive digital baseflow separation filter, *Hydrol. Earth Syst. Sci.*, *16*, 451–455.

Fenicia, F., Savenije, H.H.G., Matgen, P., Pfister, L. (2006), Is the groundwater reservoir linear? Learning from data in hydrological modelling. *Hydrol. Earth Syst. Sci. 10*, 139–150.

Ferket, B.V.A., Samain, B., Pauwels, V.R.N. (2010), Internal validation of conceptual rainfallrunoff models using baseflow separation. *J. Hydrol.*, *381*, 158–173.

Furey, P. R., and Gupta, V. K. (2001), A physically based filter for separating base flow from streamflow time series, *Water Resour. Res.*, *37*(11), 2709–2722.

Gonzales, A. I., Nonner, J., Heijkers, J., and Uhlenbrrok, S. (2009), Comparison of different base flow separation methods in a lowland catchment, *Hydrol. Earth Sys. Sci.*, *13*, 2055–2068.

Hirsch, R.M. and Costa, J.E. (2004.), US stream flow measurement and data dissimination improve. *Eos*, *85*, 197–203.

Hurst, H. E. (1951), Long-term storage capacity of reservoirs. *Trans. Am. Soc. Civil Engin.*, *116*, 770–799.

Huyck, A. A. O., Pauwels, V. R. N., and Verhoest, N. E. C. (2005), A base flow separation algorithm based on the linearized Boussinesq equation for complex hillslopes, *Water Resour. Res.*, *41*, W08415.

Accepted Article

This article is protected by copyright. All rights reserved.

Jakeman, A. J., and G. M. Hornberger (1993), How much complexity is warranted in a rainfall-runoff model? *Water Resour. Res.*, *29*(8), 2637–2649.

Kim, J., and Hogue, T. S. (2008), Evaluation of a MODIS-based potential evapotanspiration product at the point scale, J. Hydrometeorol., *9*, 444–460.

Kirchner, J. W. (2009), Catchments as simple dynamical systems: Catchment characterization, rainfall-runoff modeling, and doing hydrology backward, *Water Resour. Res.*, *45*, W02429.

Li, L., Maier, H. R., Lambert, M. F., Simmons, C. T., and Partington, D. (2013), Framework for assessing and improving the performance of recursive digital filters for baseflow estimatin with application to the Lyne and Hollick filter, *Environ. Model. and Soft.*, *41*, 163–175.

Li, L., Maier, H. R., Partington, D., Lambert, M. F., and Simmons, C. T. (2014), Performance assessment and improvement of recursive digital baseflow filters for catchments with different physical characteristics and hydrological inputs, *Environ. Model. and Soft.*, *54*, 39–52.

Linsley, R. K., Kohler, M. A., Paulhus, J. L. H., and Wallace, J. S. (1958), *Hydrology for engineers*. McGraw Hill, New York.

Lyne, V. D., and Hollick, M. (1979), Stochastic time-variable rainfall-runoff modelling. In Hydrology and Water Resources Symposium, Institution of Engineers Australia, Perth, 89–92.

Mackay, D. S., and Band, L. E. (1998), Extraction and representation of nested catchment areas from digital elevation models in lake-dominated topography, *Water Resour. Res.*, *34(4)*, 897–901.

Accepted Article

This article is protected by copyright. All rights reserved.

McDonnell, J., et al. (2010), How old is the water? Open questions in catchment transit time conceptualization, modelling and analysis, *Hydrol. Processes*, 24(12), 17451754.

Nathan, R. J., and McMahon, T. A. (1990), Evaluation of automated techniques for base flow and recession analysis, *Water Resour. Res.*, *26*(7), 1465–1473.

O'Connell, P. E., (1971), A simple stochastic modeling of Hurst's law, paper presented at International Symposium on Mathematical Models in Hydrology, Int. Ass. Sci. Hydrol., Warsaw.

Peters, E., and Van Lanen, H. A. J. (2005), Separation of base flow from streamflow using groundwater levels illustrated for the Pang catchment (UK), *Hydrol. Processes*, *19*, 921–936.

Pettyjohn, W. A., and Henning, R. (1979), Preliminary estimate of groundwater recharge rates, related streamflow and water quality in Ohio. Ohio State University Water Resources Centre Project Completion Report No 552, 323.

Partington, D., Brunner, P., Simmons, C.T., Therrien, R., Werner, A.D., Dandy, G.C., and Maier, H.R. (2011), A hydraulic mixing-cell method to quantify the groundwater component of streamflow within spatially distributed fully integrated surface water–groundwater flow models, *Environ. Model. Softw.*, *26(7)*, 886–898.

Partington, D., Brunner, P., Simmons, C.T., Werner, A.D., Therrien, R., Maier, H.R., Dandy, G.C. (2012), Evaluation of outputs from automated baseflow separation methods against simulated baseflow from a physically based, surface water-groundwater flow model. *J. Hydrol.*, *458–459*, 28–39.

Rimmer, A., and Hartmann, A. (2014), Optimal hydrograph separation filter to evaluate transport routines of hydrological methods, *J. Hydrol.*, *514*, 249–257.

Accepted Article

This article is protected by copyright. All rights reserved.

Schoonover, J. E., Lockaby, B. G., and Helms, B. S. (2006), Impacts of land cover on stream hydrology in the West Georgia Piedmont, USA, *J. Environ. Qual.*, *35*, 2123–2131.

Shafii, M., and Tolson, B. A. (2015), Optimizing hydrological consistency by incorporating hydrological signatures into model calibration objectives, *Water Resour. Res.*, *51*(5), 3796–3814.

Shapoori, V. Peterson, T. J., Western, A. W., and Costelloe, J. F. (2015), Decomposing groundwater head variations into meteorological and pumping components: a synthetic study, *Hydrogeol. J.*, *23(7)*, 1431–1448.

Simpson, J., C. Kummerow, W.-K. Tao, and R. F. Adler (1996), On the Tropical Rainfall Measuring Mission (TRMM), *Meteor. Atmos. Phys.*, *60*, 19-36.

Sloto, R. A., and Crouse, M. Y. (1996), HYSEP: A computer program for streamflow hydrograph separation and analysis, US Geological Survey, Water Resources Investigations Report 96–40.

Smith, E., et al., 2007: The International Global Precipitation Measurement (GPM) program and mission: An overview, *Measuring Precipitation from Space: URAINSAT and the Future*, V. Levizzani and F. J. Turk, Eds., Springer, 611–653.

Stewart, M., Cimino, J., and Ross, M. (2007), Calibration of base flow separation methods with streamflow conductivity. *Ground Water*, *45*(1), 17–27.

Stoelzle, M., Stahl, K., and Weiler, M. (2013), Are streamflow recession characteristics really characteristic? *Hydrol. Earth Syst. Sci.*, *17*, 817–828.

Su, C.-H., Costelloe, J. F., Peterson, T. J., and Western, A. W (2016), On the structural limitations of recursive digital filters for baseflow estimation. *Water Resour. Res.*, *52*(6),

Accepted Article

This article is protected by copyright. All rights reserved.

4745–4764.

Szilagyi, J., 2004. Heuristic continuous base flow separation. *J. Hydrol. Eng. 9*, 311–318.

Tallaksen, L. M. (1995), A review of baseflow recession analysis, *J. Hydrol.*, *165*, 349–370.

Therrien, R., McLaren, R. G., Sudicky, E. A., Panday, S. M. (2009), HydroGeoSphere: A three-dimensional numerical model describing fully-integrated subsurface and surface flow and solute transport: Waterloo

Van Dijk, A. I. J. M. (2010), Climate and terrain factors explaining streamflow response and recession in Australian catchments, *Hydrol. Earth Syst. Sci*, *15*, 159–169.

Vogel, R. and Kroll, C. (1992), Regional geohydrologic-geomorphic relationships for the estimation of low-flow statistics, *Water Resour. Res.*, *28*, 2451–2458.

Vrugt, A., and Sadegh, M. (2013), Toward diagnostic model calibration and evaluation: Approximate Bayesian computation. *Water Resour. Res.*, 48, 4335–4345.

Wittenberg, H (1999), Baseflow recession and recharge as nonlinear storage processes, *Hydrol. Process.*, *13*, 715–726.

Zhang, Y.-K., and Schilling, K. (2004), Temporal scaling of hydraulic head and river base flow and its implication for groundwater recharge, *Water Resour. Res.*, *40*, W03504.

Zhang, R., Li, Qiang, Chow, T. L., Li, S., and Danielescu, S. (2013), Baseflow separation in a small watershed in New Brunswick, Canada, using a recursive digital filter calibrated with the conductivity mass balance method, *Hydrol. Processes*, *27*, 2659–2665.

Accepted Article

This article is protected by copyright. All rights reserved.

Table 1. Parameters for the synthetic HGS catchment models from *Li et al.* [2014]. The parameters A, $S1$, $S2$, AR, K_s, α_{vG}, and β_{vG} and forcing are varied to produce 66 test catchments used in this study.

Parameter	Value(s)
	Forcing
Mean annual rainfall (mm/a)	510 (Ade), 525 (Mel), 1096 (Syd), 1667 (Dar), 2238 (Bri)
Mean annual PET (mm/a)	911 (Mel), 921 (Syd), 1471 (Ade), 2056 (Dar), 2077 (Bri)
	Geometry
Catchment area A (km^2)	6, 48, 80, 120, 192
Hill slope $S1$	0.005, 0.008, 0.012, 0.016, 0.02
Channel slope $S2$	0.0025, 0.004, 0.006, 0.008, 0.01
Aspect ratio of width to length AR	0.5, 0.75, 1.0, 1.25, 1.5
	Surface
Manning's roughness (s/m$^{1/3}$)	0.015
Rill storage height (m)	0.001
Obstruction storage height (m)	0.0
Transpiration fitting parameter c1, c2, c3	0.3, 0.2, 10
Leaf area index	2.08
Wilting point	0.1
Field capacity	0.15
Oxic limit	0.9
Anoxic limit	1.0
Limiting saturation min, max	0.2, 0.32
Canopy storage parameter (m)	0.0
Initial interception storage (m)	0.0
	Subsurface
Evapotranspiration depth[a](m)	3
Root depth[a](m)	6
Saturated hydraulic conductivity (K_s) [m/s]	2.44E-5, 3.99E-5, 1.12E-4, 2.11E-4, 9.66E-4
van Genuchten α_{vG} (1/m)	0.572, 3.366, 6.161, 8.955, 11.75
van Genuchten β_{vG}	1.32, 1.56, 1.79, 2.03, 2.27
	Surface-subsurface coupling
Coupling length (m)	0.001

[a] Quadratic decay function

Accepted Article

This article is protected by copyright. All rights reserved.

Accepted Article

Table 2. Percentages of the 132 evaluation samples that show good, average or poor performance (as defined in Sec. 2.6) for a given calibration scheme.

Method	NSE_d			NSE_s			NSE_m			PBIAS		
	Good	Average	Poor	Good	Average	Poor	Good	Average	Poor	Good	Average	Poor
Calibration period												
Base case	48.5	13.6	37.9	38.6	4.5	56.8	68.9	25.8	5.3	97.0	3.0	0.0
Standard approach	3.8	3.0	93.2	0.0	0.0	100.0	9.1	14.4	76.5	45.5	26.5	28.0
ρ_1	13.6	3.0	83.3	5.3	3.0	91.7	30.3	6.1	63.6	34.8	6.8	58.3
ρ_2	25.0	12.9	62.1	12.9	3.8	83.3	39.4	11.4	49.2	48.5	18.9	32.6
ρ_3	28.8	11.4	59.8	9.1	5.3	85.6	61.4	12.1	26.5	79.5	9.1	11.4
ρ_4	44.7	17.4	37.9	12.1	11.4	76.5	71.2	18.2	10.6	97.0	1.5	1.5
ρ_5	12.1	12.9	75.0	3.0	5.3	91.7	18.2	22.0	59.8	40.2	23.5	36.4
ρ_6	36.4	5.3	58.3	13.6	9.8	76.5	50.0	6.1	43.9	68.2	6.1	25.8
ρ_7	39.4	4.5	56.1	31.1	2.3	66.7	53.0	3.0	43.9	70.5	8.3	21.2
Validation period												
Base case	47.7	13.6	38.6	38.6	3.0	58.3	68.9	23.5	7.6	96.2	3.8	0.0
Standard approach	3.8	3.0	93.2	0.0	0.0	100.0	9.1	14.4	76.5	45.5	26.5	28.0
ρ_1	13.6	3.0	83.3	5.3	3.0	91.7	30.3	6.1	63.6	34.8	6.8	58.3
ρ_2	26.5	11.4	62.1	12.1	4.5	83.3	37.1	12.9	50.0	50.0	18.2	31.8
ρ_3	27.3	10.6	62.1	9.1	3.8	87.1	60.6	14.4	25.0	75.8	13.6	10.6
ρ_4	42.4	18.2	39.4	15.2	6.8	78.0	70.5	18.2	11.4	97.0	2.3	0.8
ρ_5	15.2	9.1	75.8	3.0	4.5	92.4	20.5	16.7	62.9	40.9	24.2	34.8
ρ_6	33.3	7.6	59.1	13.6	9.8	76.5	47.7	10.6	41.7	65.9	9.8	24.2
ρ_7	39.4	3.8	56.8	30.3	3.0	66.7	51.5	5.3	43.2	68.9	7.6	23.5

This article is protected by copyright. All rights reserved.

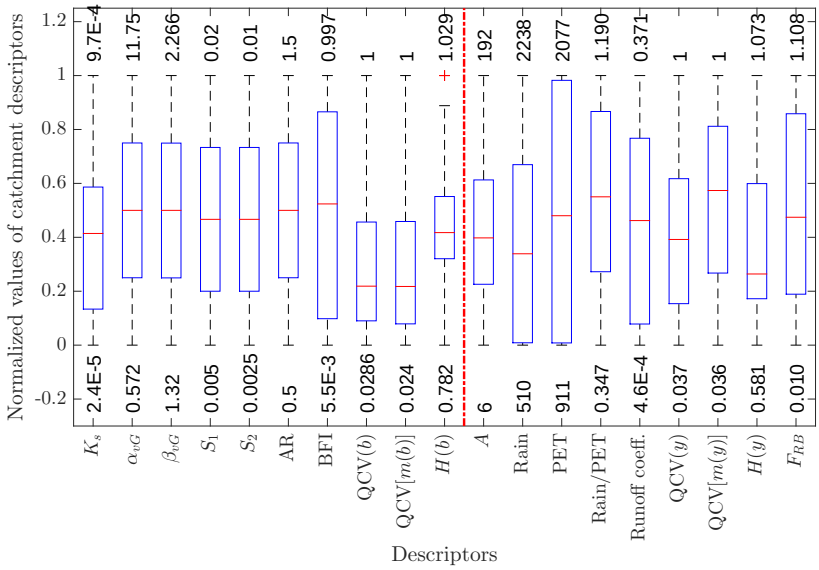


Figure 1. Boxplots showing the characteristics of the 66 HGS catchments based on a set of descriptors. The values of the descriptors are linearly normalized between 0 and 1, except for K_s that is log-normalized. The annotated numbers correspond to the minimum and maximum values for each descriptor. The dashed lines distinguish the descriptors that can be estimated relatively cheaply a priori (right) from those that cannot (left).

Table 3. Percentages of the 70 cross-validation samples where the performance scores NSE_d, NSE_s, NSE_m and PBIAS show consistency between the calibration and validation samples. Refer to text for our criterion for consistency.

Method	NSE_d	NSE_s	NSE_m	PBIAS
Base case	91.4	72.9	98.6	100.0
Standard approach	17.1	7.1	21.4	60.0
ρ_1	22.9	11.4	57.1	80.0
ρ_2	67.1	31.4	82.9	97.1
ρ_3	57.1	17.1	92.9	98.6
ρ_4	94.3	37.1	98.6	100.0
ρ_5	42.9	17.1	67.1	88.6
ρ_6	64.3	24.3	78.6	85.7
ρ_7	71.4	51.4	88.6	97.1

Accepted Article

This article is protected by copyright. All rights reserved.

Accepted

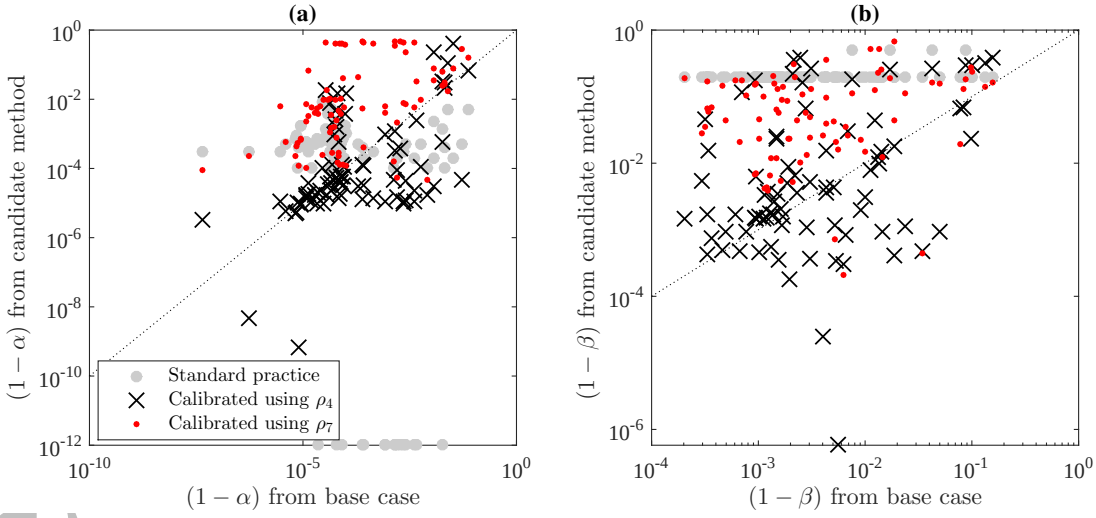


Figure 2. Comparisons of the optimized filter parameter values, α and β, derived in the base case and by the signature method with ρ_4 and ρ_7. Only the cross-validation samples where the base-case $\mathrm{NSE}_d > 0$ are shown. The values used in the standard practice are also superimposed.

This article is protected by copyright. All rights reserved.

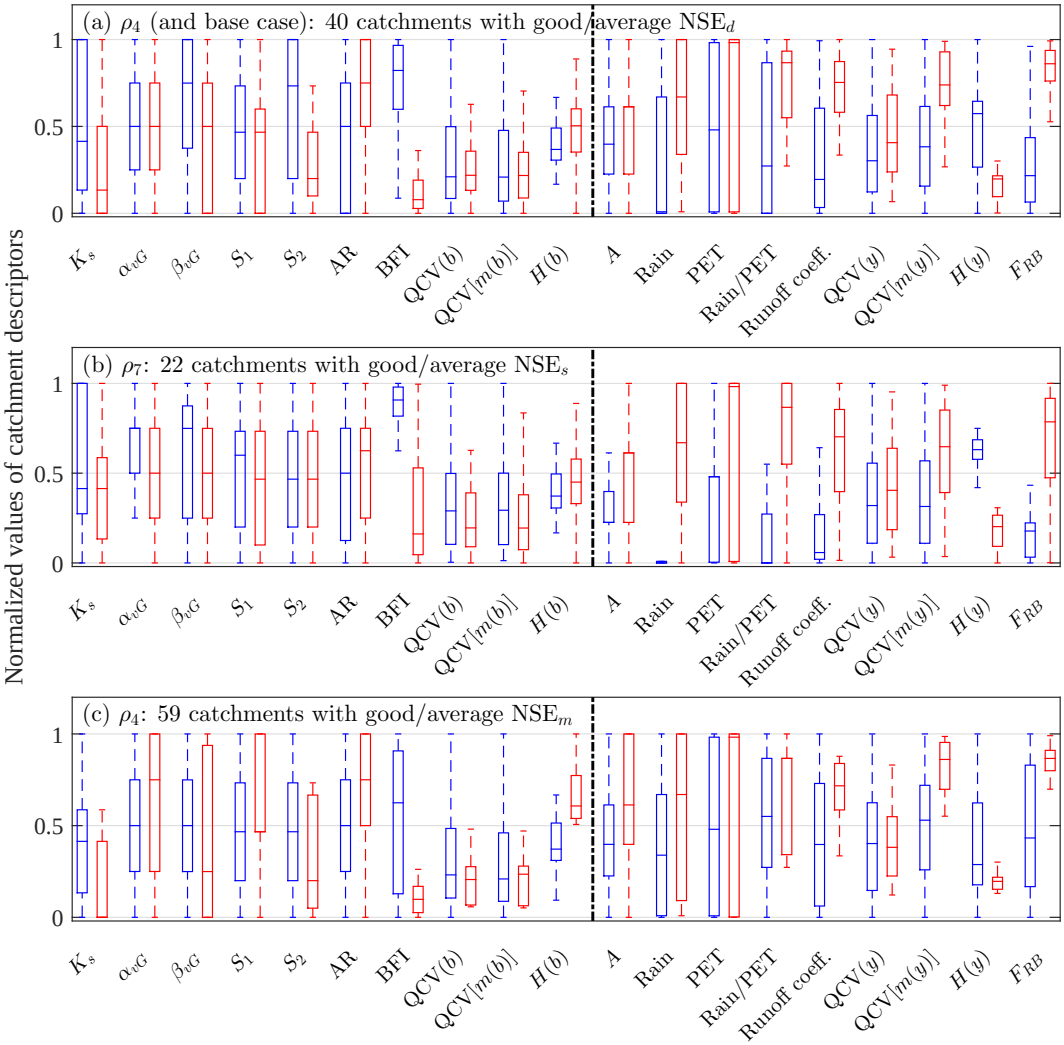


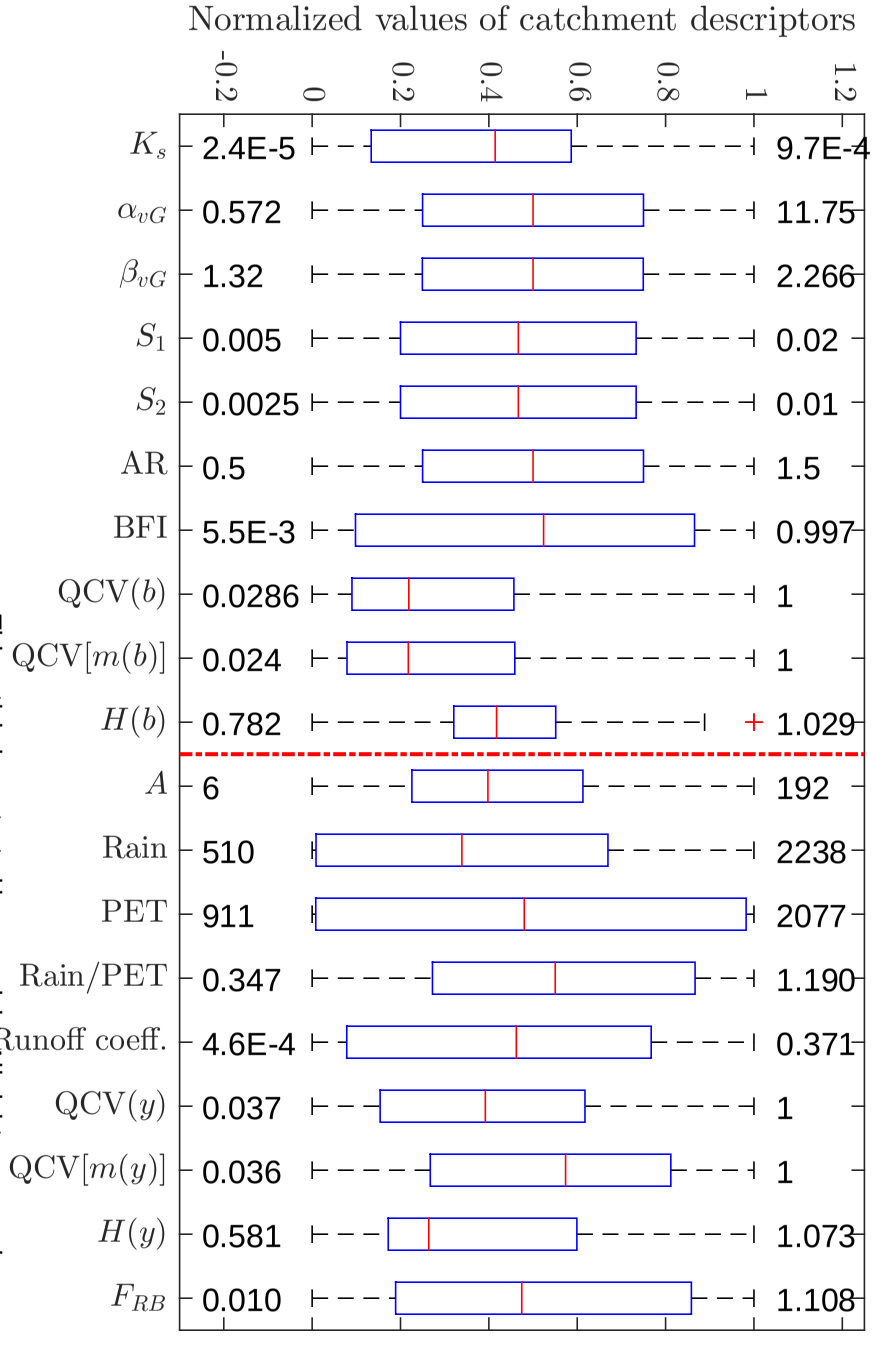
Figure 3. Blue boxplots show the characteristics of catchments where a given calibration method yields good or average scores during evaluation, and red boxplots show those with poor scores. The dashed lines distinguish the descriptors that can be estimated relatively cheaply a priori (right) from those that cannot (left).

This article is protected by copyright. All rights reserved.

Figure 1. Figure

Accepted Article

This article is protected by copyright. All rights reserved.

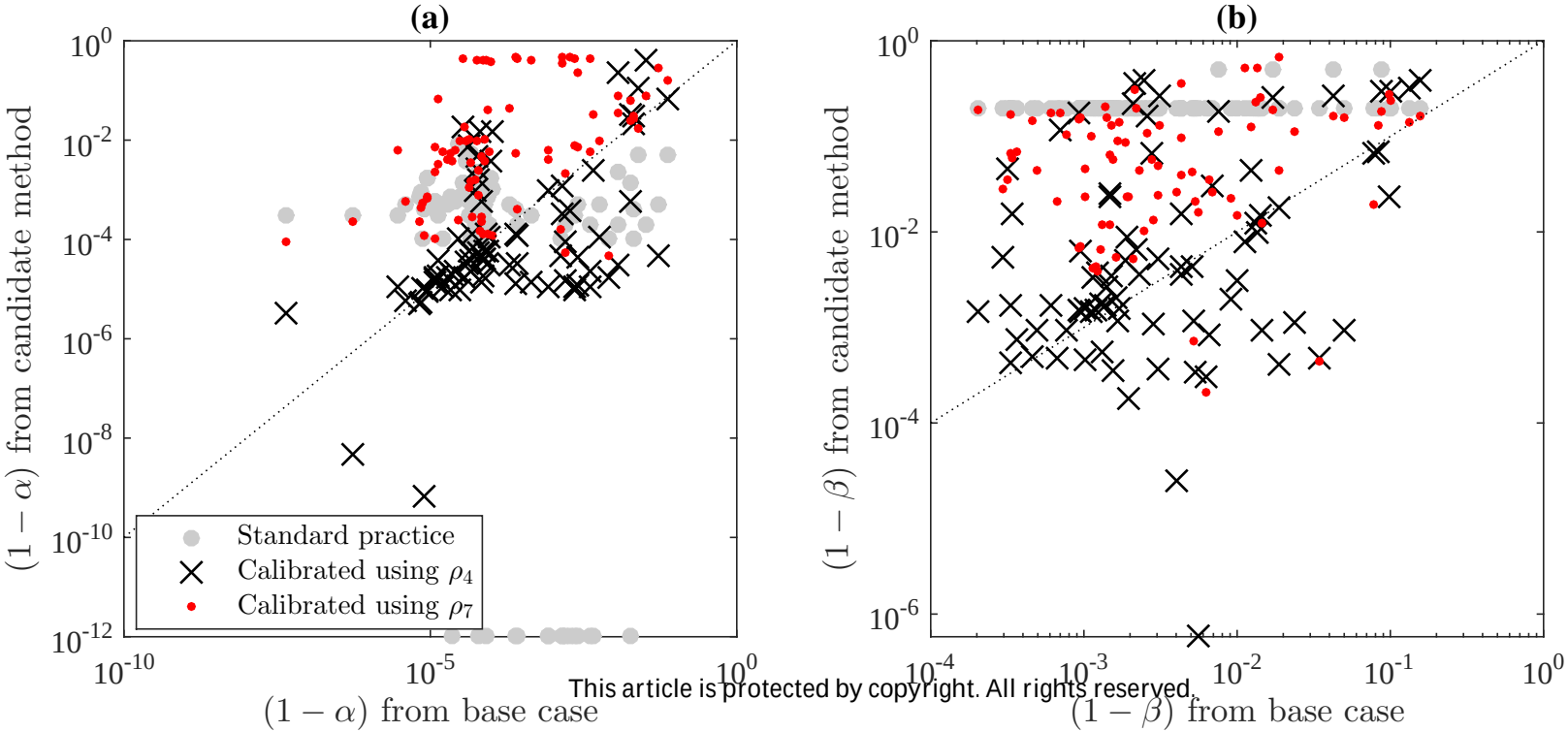


This article is protected by copyright. All rights reserved.

Figure 2. Figure

Accepted Article

This article is protected by copyright. All rights reserved.

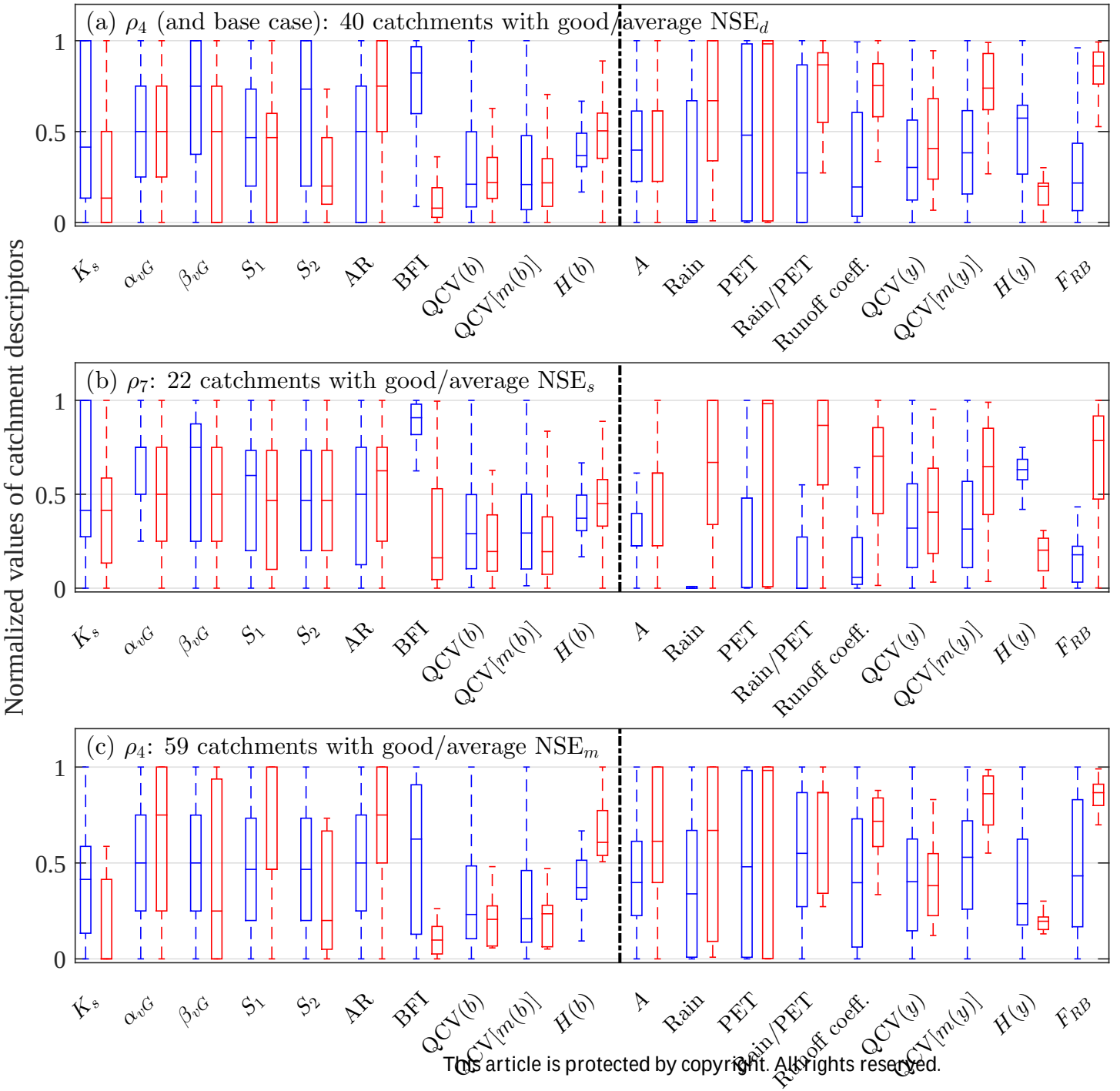


This article is protected by copyright. All rights reserved.

Figure 3. Figure

Accepted Article

This article is protected by copyright. All rights reserved.



This article is protected by copyright. All rights reserved.