AGU ADVANCING EARTH AND SPACE SCIENCE

Water Resources Research

RESEARCH ARTICLE
10.1029/2021WR030800

A Calibration-Free Groundwater Module for Improving Predictions of Low Flows

Arik Tashie[1], Tamlin Pavelsky[2], and Mukesh Kumar[1]

[1]Department of Civil, Construction, and Environmental Engineering, University of Alabama, Tuscaloosa, AL, USA, [2]Department of Geological Sciences, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

Key Points:

- Develop a calibration-free groundwater/low flow module that is compatible with land surface models and rainfall-runoff models
- Without calibration, the module's performance compares favorably with an existing, fully calibrated alternative
- Module predictions of low flows are robust to drastic changes in the overlying hydrologic model

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

A. Tashie,
arik@climate.ai

Citation:

Tashie, A., Pavelsky, T., & Kumar, M. (2022). A calibration-free groundwater module for improving predictions of low flows. *Water Resources Research*, *58*, e2021WR030800. https://doi.org/10.1029/2021WR030800

Received 9 JUL 2021
Accepted 8 MAR 2022

Author Contributions:

Conceptualization: Arik Tashie, Tamlin Pavelsky
Formal analysis: Arik Tashie
Investigation: Arik Tashie
Methodology: Arik Tashie, Mukesh Kumar
Project Administration: Tamlin Pavelsky, Mukesh Kumar
Resources: Tamlin Pavelsky, Mukesh Kumar
Software: Arik Tashie
Supervision: Tamlin Pavelsky, Mukesh Kumar
Validation: Arik Tashie, Mukesh Kumar
Visualization: Arik Tashie, Tamlin Pavelsky, Mukesh Kumar
Writing – original draft: Arik Tashie
Writing – review & editing: Arik Tashie, Tamlin Pavelsky, Mukesh Kumar

Abstract Groundwater modules are critically important to the simulation of low flows in physically based land surface models (LSMs) and conceptual rainfall-runoff models (HBV). Here, we develop a Groundwater for Ungauged Basins (GrUB) module that uses only physically based properties for which data are widely available, thus allowing its application without the need for calibration. GrUB is designed to be computationally simple and readily adaptable to a wide variety of LSMs and rainfall-runoff models. We assess the performance of GrUB in 84 United States watersheds by incorporating it into HBV, a popular rainfall-runoff model. We compare predictions of low flows by the native (calibrated) HBV groundwater module with those by the (uncalibrated) GrUB module and find that GrUB generates error metrics that are equivalent or superior to those generated by the (calibrated) HBV groundwater module. To assess whether predictions by GrUB are robust to changes in the structure and parameterization of the overlying hydrologic model, we run tests for two artificial scenarios: Slow Recharge with rates of percolation below 0.1 mm/day, and Fast Recharge with rates of percolation of up to 1,000 mm/day. GrUB proves to be robust to these extreme changes, with mean absolute error (MAE) of predictions of low flows only increasing by an average of up to 17%, while average MAE increases by up to 158% when the same tests are performed on HBV without the GrUB module. We suggest GrUB as a potential tool for improving predictions of low flows in LSMs, as well as rainfall-runoff models when calibration data are sparse.

1. Introduction

1.1. Background

Shallow groundwater supplies a majority of streamflow to most watersheds (Beck et al., 2013) and is the primary source of streamflow during seasonal low flows (Smakhtin, 2001). This mobile water resource provides an essential buffer to changes in temperature, nutrients, and precipitation (Ficklin et al., 2015) and sustains evapotranspiration (ET; Yang et al., 2011) during low flow periods. Even in snow-dominated climates, the hydrologic response to snowmelt is mediated by groundwater (Enzminger et al., 2019) with important implications for the response of these systems to global warming (Tague & Grant, 2009). Because shallow groundwater contributions to low flows are disproportionately sensitive to changes in near-term (i.e., years and decades) climate signals (Hare et al., 2021), it is essential that our hydrologic models are able to predict groundwater contributions to low flows accurately. Unfortunately, accurate simulation of low flows and groundwater contributions to them has proven to be notoriously difficult in distributed, physically based land surface models (LSMs; Clark et al., 2015, Holtzman et al., 2020) as well as lumped, conceptual "bucket" models (i.e., rainfall-runoff models; Fowler et al., 2020).

Early LSMs described low flows as dependent on one-dimensional drainage below a soil column (Clark et al., 2015). This "free" drainage led to well-documented inaccuracies: too-fast drainage during wet periods, underestimates of seasonal storage, and the cessation of ET during short dry periods (Baker et al., 2008; Brunke et al., 2016; Fan et al., 2019; Kuppel et al., 2017; Miguez-Macho & Fan, 2012a, 2012b; Milly & Shmakin, 2002; Pokhrel et al., 2013). Though LSM researchers have introduced several mechanisms for constraining low flows, this additional model complexity (NOAA, 2016) has failed to generate consistent improvements in model predictions (Gan et al., 2019; Yang et al., 2011). Indeed, it is often necessary to incorporate a calibrated groundwater module in LSMs (Fang et al., 2019; Holtzman et al., 2020) when using LSM projections in a predictive context. The requirement of calibration both increases the complexity of applying LSMs and (importantly) limits their utility in ungauged basins.

The failure to clearly improve low flows in LSMs has been partially attributed to three issues. First, the simple (linear) groundwater reservoirs (e.g., Niu et al., 2007) and quasi-TOPMODEL modules (e.g., Niu et al., 2005 or Oleson et al., 2010) that are commonly used may lack the complexity to characterize heterogeneous watershed features that drive highly nonlinear streamflow processes during periods of recession (Clark et al., 2009, 2015; Fan et al., 2019; Rahman et al., 2019; Tashie et al., 2019). Second, the data used to parameterize these low flow modules are highly uncertain. For example, the hydraulic properties of the deep subsurface are typically estimated according to the texture of the shallow overlying soil using a pedotransfer function (Gedney & Cox, 2003). Even where it is appropriate to assume a strong correlation between hydraulic properties of shallow soils and those of the deeper soils, regolith, and bedrock, this approach ignores macropores (Mendoza et al., 2003; Zecharias & Brutsaert, 1988) and is biased in non-temperate climates (Hengl et al., 2017; Huscroft et al., 2018). Finally, the overreliance on only one or two datasets (i.e., areal average hydraulic conductivity and/or topographic wetness index) to characterize a complex process heightens the likelihood that the biases and uncertainties that are inherent in all large datasets express themselves in model predictions (Hariri et al., 2019).

These issues are not limited to LSMs, but are also increasingly being recognized as an area of needed improvement in conceptual rainfall-runoff models (Fowler et al., 2020; Seibert & van Meerveld, 2016). Computationally simple rainfall-runoff models are often applied in climate change impact studies to inform policy and decision-making in a wide variety of arenas (Balkovič et al., 2018; Cui et al., 2018; de Jong et al., 2018, Emanuel, 2018; Flörke et al., 2018; Iqbal et al., 2018; Mahmoud & Gan, 2018). However, rainfall-runoff models tend to become increasingly inaccurate during periods of hydroclimatic variability (Saft et al., 2016; Seibert & van Meerveld, 2016), which inhibits their utility in predicting watershed response to climate. Furthermore, most rainfall-runoff models require calibration of several parameters on a single objective function, which raises the issue of equifinality (Beven, 2006) and limits their utility in ungauged and poorly gauged basins (Boughton, 2006).

2. Objectives

We propose to improve predictions of low flows in LSMs and conceptual rainfall-runoff models by developing a portable, data-driven module called Groundwater for Ungauged Basins (GrUB). To facilitate easy and broad adoption of this module by the modeling community, we are guided by four key practical principles:

1. *No calibration required*: The module must be useable directly "out of the box" and easily applied in ungauged basins.
2. *Simple data requirements*: The module must depend only on widely available (continental- or global-scale) data that requires minimal processing on the part of the model user.
3. *Modular*: To be adaptable to a variety of LSMs and rainfall-runoff models, the module must be driven by a single flux, that is, common in most LSMs and rainfall-runoff models (i.e., deep recharge), and it must operate independently of the rest of the model structure (i.e., no feedback mechanisms).
4. *Computational simplicity*: The module must not substantially increase the run time of LSMs (which are already computationally complex) or rainfall-runoff models (for which short run times are a major source of appeal).

Following the module development (detailed below), we incorporate GrUB into a common rainfall-runoff model (HBV) and assess its performance by posing the following questions:

1. Does incorporating the uncalibrated GrUB module into the otherwise calibrated HBV model negatively affect its overall performance?
2. Does GrUB reduce error and bias in predictions of low flows, especially during historically long dry periods?
3. Are GrUB predictions of low flows robust to changes in the parameterization and performance of the overlying hydrologic model?

The overall goal of this analysis is to use the performance of the *calibrated* HBV model as a benchmark against which to judge the relative performance of the *uncalibrated* GrUB module. We stress that we are not arguing that the *uncalibrated* GrUB module should be assessed as a potential replacement for the *calibrated* HBV module. Instead, we analyze the relative performance of GrUB across a wide variety of climatic and physiographic settings as a first step toward the future incorporation of GrUB into rainfall-runoff models in settings where calibration is

not possible, or into more complex, physically based LSMs that are cannot be calibrated. Following this analysis, we discuss the limitations of GrUB and identify potential avenues for improvement.

3. Model Development

3.1. Conceptual Model

In individual watersheds (or hillslopes), actual groundwater flow patterns are exceedingly complex, varying in time, place, and with antecedent conditions (Aulenbach et al., 2021; Tashie et al., 2019; Zimmer & McGlynn, 2017a, 2017b). Which mechanisms are dominant (and when) varies according to the specific geophysical properties and climatic fluxes of a watershed (Tashie et al., 2020). Unfortunately, the practical realities of large-scale hydrologic and land surface modeling demand that low flow calculations must be generic (i.e., the same for all watersheds) and computationally efficient.

In this research, we apply a bottom-up, data driven approach to module development. Broadly, we propose that models that leverage a wide variety of empirical data are more likely to be transferrable among watersheds without the need for calibration. We begin by identifying a class of conceptual models that incorporates a wide variety of physically based properties and mechanisms for which extensive empirical data are readily available. Then, we build a module that is simple and generic, so that it is applicable across a range of watersheds using only these widely available data.

Clark et al. (2015) identify five categories of conceptual models for groundwater behavior for application in LSMs: (a) free or restricted drainage; (b) conceptual buckets; (c) TOPMODEL; (d) grid-to-grid lateral flux; and (e) hillslope flow based on Darcy's Law. (Grid-to-grid lateral flux does not directly generate runoff and is therefore not further considered here.) Free drainage and conceptual buckets generate subsurface runoff (Q_{sb}) [L^3/t] as a function of storage (S) and saturated hydraulic conductivity (K_s) [LT^{-1}]:

$$Q_{sb} = f(S,\ K_s) \tag{1}$$

where S is a function of modeled inputs to and outputs from the lower boundary and K_s is usually derived from the properties of lowest layer of the soil column. Free and restricted drainage modules were the most common approaches in early LSMs and remain popular (e.g., CABLE, TESSEL, Noah, and ORCHIDEE; Kowalczyk et al., 2013; Krinner et al., 2005; Niu et al., 2011; van den Hurk et al., 2000).

To address the poor performance of free and restricted drainage models, many LSMs (e.g., Catchment, CLM, JULES, MATSIRO, and Noah-MP) have conceptually incorporated variable contributing source areas (i.e., TOPMODEL); Beven & Kirkby, 1979), where Q_{sb} is again a function of S and K, but also depends on catchment-average topographic slope and curvature, that is, the topographic wetness index (TWI):

$$Q_{sb} = f(S,\ K_s, TWI) \tag{2}$$

In TOPMODEL approaches, S is again a function of modeled inputs to and outputs from the lower boundary, while TWI and K_s are largely derived from empirical data. TWI is calculated directly from digital elevation models (DEMs) and K_s varies with S, such that maximum K_s is estimated from the properties of the lowest layer of the soil column and decays exponentially with declining values of S (e.g., Niu et al., 2005).

Finally, flow from a representative hillslope based on Darcy's Law has been explicitly identified as a potential mechanism for improving representation of groundwater and low flows in LSMs (Clark et al., 2015; Fan et al., 2019), though only one version of this mechanism has been included in any LSM to date (i.e., LM3; Subin et al., 2014). Broadly, a representative hillslope flow model describes Q_{sb} as a function of S, K_s, and several topographic and geomorphic variables:

$$Q_{sb} = f(S,\ K_s, i,\ B,\ L, Th) \tag{3}$$

where i is aquifer slope, L is stream network length [L], B is characteristic aquifer breadth [L], and Th is aquifer thickness [L]. (Note that B is traditionally denoted by the letter L in hillslope hydraulics equations, and Th is often denoted with the letter D; however, in GrUB, these terms are reserved for the naming of other parameters.) Once again, S [L] is a function of modeled inputs to and outputs from the lower boundary. K_s, i, B, L, and Th represent physical properties of the watershed for which empirical data is widely available. These empirical data

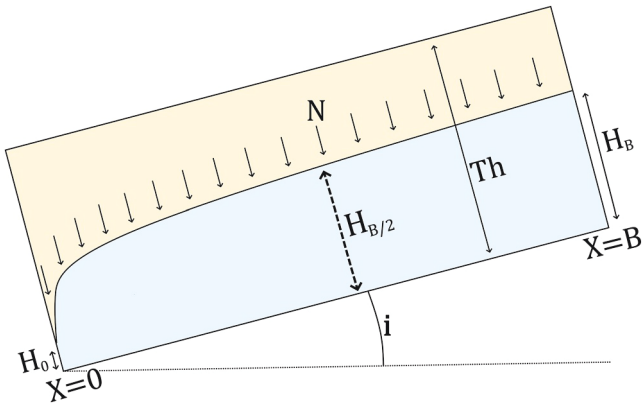

Figure 1. A definition sketch of the cross section of a hillslope aquifer atop a sloping impermeable layer under a state of steady recharge. Note that the sketch is distorted for easier display (in reality Th << B).

represent an opportunity to uniquely parameterize watersheds without calibration, and without overreliance on any single dataset or set of simplifying assumptions that may be prone to bias. Therefore, we proceed to develop the GrUB module according to this approach.

3.2. Model Equations

We begin with a continuity equation and add complexity where hydrological realism demands further refinement:

$$dS = N - q_{sb} \tag{4}$$

where N [LT^{-1}] is recharge per unit area (i.e., modeled drainage from the unsaturated zone in Figure 1), q_{sb} [LT^{-1}] is subsurface discharge from the hillslope per unit area, and dS [LT^{-1}] is change in groundwater storage. Volumetric subsurface discharge (Q_{sb}) [L^3] is a product of q_{sb}, the thickness of the saturated aquifer at the stream-hillslope interface (H_0) [L], and the total length of stream network (L) [L], thus:

$$Q_{sb} = q_{sb} H_0 L = K_s \frac{dh}{dx} H_0 L \tag{5}$$

where dh/dx [-] is the hydraulic gradient and K_s is the effective saturated hydraulic conductivity of the components of the subsurface that are actively sustaining streamflow (i.e., the hillslope aquifer). Watershed-scale estimates of L can be derived from topography, remote sensing, and modeling experiments, and these data are widely available in public databases like the National Hydrography Dataset (NHD) (USGS, 2004). Because stream networks in the NHD tend to underestimate L by a factor of up to 2 in many watersheds (Godsey & Kirchner, 2014; Jensen et al., 2017), we rescale L by a factor of 1.5 in all watersheds for this study. This leaves K_s, dh/dx, and H_0 to be solved for.

As noted above, K_s in LSMs is typically estimated according to the soil texture at the bottom of the soil column according to a pedotransfer function and modeled to decay exponentially with depth to water table (Niu et al., 2005). However, pedotransfer functions are known to be highly uncertain (Freeze & Cherry, 1979; Zhang & Schaap, 2019) and to systematically underestimate K_s by ignoring macropores (Mendoza et al., 2003; Zecharias & Brutsaert, 1988). Furthermore, the exponential decay of K_s is a purely empirical relationship, that is, often calibrated in hydrologic models (e.g., Tague & Band, 2004). To address these pitfalls, Tashie et al. (2021) developed maps of a quantity they term *effective* K_s for the coterminous United States that: (a) accounts for the effects of macropores on watershed-average values of K_s; and (b) explicitly represents values of K_s as a function of shallow groundwater storage (S). Importantly, these empirical data illustrate that watershed-scale effective values of K_s covary with S, such that K_s decreases exponentially as S decreases linearly. The database gives two values of K_s ($K_{s_{dry}}$ and $K_{s_{wet}}$) which can be used to model K_s as a function of S in log-linear space, with K_s taking some minimum value (K_0) during extremely dry conditions ($S = 0$): $\log_{10} K_s = \log_{10} K_0 + \left(\frac{S}{f}\right) m$, where m is an empirically derived constant unique to each watershed and f [-] is effective aquifer drainable porosity, which is set to a typical value of 0.1 where empirical data are not available. We can then solve K_s for each watershed as a function of static properties (K_0, f, and m) and time-varying S:

$$K_s = 10^{\left(\log_{10} K_0 + (S/f) m\right)} \tag{6}$$

This dataset also has the benefit of providing an estimate of maximum saturated aquifer thickness (Th) which is useful for approximating the two remaining terms to be solved: dh/dx and H_0.

Extensive literature exists deriving analytical solutions to dh/dx and H_0 as functions of Th, S, N, and drawdown time, with each set of solutions applying its own set of assumptions and approximations (Troch et al., 2013). Broadly, the water table elevation profile ($h(x)$) of a fully recharged hillslope aquifer is "flat" ($H_0 = \mathrm{Th} = H_B$), then drains more quickly nearer the stream ($X \to 0$) generating a curvilinear $h(x)$ profile. However, each set of assumptions and approximations for solving $h(x)$ may be more (or less) reasonable depending on the specific

properties of the aquifer and recharge pathways under consideration (see Troch et al., 2013). We apply the simplifying assumption that the hydraulic gradient (dh/dx) at the outlet ($x = 0$) may be roughly approximated as the average hydraulic gradient along the hillslope:

$$\frac{dh}{dx} = \frac{dH + dz}{dx} \approx \frac{S/f\cos i + (B/2)\sin i}{(B/2)\cos i} \tag{7}$$

where dH is the hydraulic head due to saturated aquifer thickness, dz is elevation head relative to the underlying impermeable layer, and $B/2$ is the midpoint of the hillslope.

We take a similarly practical approach to approximating H_0. H_0 increases linearly with N immediately following a pulse of recharge, such that initial conditions of a dry aquifer ($S = 0$) following a pulse of recharge (N) are $H_0 \approx N \approx H_B$, after which the effect of N on H_0 decays exponentially in time. Therefore, at any time t:

$$H_0 \approx \sum_{t=0}^{-\infty} \frac{N_t}{f} v^{-t} \tag{8}$$

where N_t is total recharge during the previous time step t (i.e., t days ago) and v [-] is an empirically derived constant that takes a positive value <1. Because LSMs and rainfall-runoff models are solved sequentially, Equation 8 can be simplified for computational efficiency as:

$$H_0 = \frac{N_0}{f} + H_{0_{t=-1}} v \tag{9}$$

where $H_{0_{t=-1}}$ is equal to H_0 during the previous time step. To account for well-documented monthly- to seasonal-scale hysteresis in low flow recession signatures (Bart & Hope, 2014; Shaw & Riha, 2012), we selected a value of v such that N from more than a month previous ($t < -31$) has a negligible effect on H_0 ($v^{31} < 0.01$). Therefore, we set $v = 0.86$ for all watersheds. Because v cannot be calculated directly, we provide sensitivity analysis for this variable in Section 4.4.

However, H_0 does not reach 0 during periods without recharge, but instead asymptotically approaches some minimum value (H_{min}). To approximate H_{min}, we rely on estimates of aquifer storage during historically dry periods (S_{dry}) for all watersheds in the coterminous United States from Tashie et al. (2021). Because S_{dry} is an empirical estimation of the minimum total storage necessary to sustain stream low flows during historically dry periods, therefore S_{dry}/f represents the correlated minimum average aquifer thickness during historically dry periods. We estimate H_{min} as some fraction of S_{dry}/f such that $H_{min} = \frac{S_{dry}}{f} d$, where d [-] is an empirical constant. While values of S_{dry}/f are available for the coterminous United States in Tashie et al. (2021), values of d are not. Therefore, we estimate d according to the following criteria: (a) in actual hillslopes, a reasonable absolute minimum thickness of the stream-aquifer interface during extraordinarily dry periods in headwaters is likely to be on the scale of (tens of) millimeters; (b) the median S_{dry}/f for the coterminous United States is about 2,000 mm (Tashie et al., 2021); and (c) therefore $d = 0.001$ provides a reasonable estimation on H_{min}. Combining this estimate with Equation 9 yields:

$$H_0 = \frac{N_0}{f} + H_{0_{t=-1}} v + \frac{S_{dry}}{f} d \tag{10}$$

Solving Equation 5 with Equation 6, Equation 7, and Equation 10 gives

$$Q_{sb} = \left(10^{(\log_{10} K_0 + (S/f)m)}\right)\left(\frac{S/f\cos i + (B/2)\sin i}{(B/2)\cos i}\right)\left(\frac{N_0}{f} + H_{0_{t=-1}} v + \frac{S_{dry}}{f} d\right) L \tag{11}$$

This computationally efficient model structure contains seven parameters that are watershed-scale empirical values (K_0, m, S_{dry}, B, f, L, and i), two parameters that are universal estimates (d and v), one variable, that is, supplied by the overlying LSM or rainfall-runoff model (N_0), and two variables that are solved iteratively by the GrUB low flow module (S and $H_{0_{t=-1}}$). As noted above, we provide sensitivity analysis for v and d in Section 4.4 because their values are neither analytically derived nor empirically estimated.

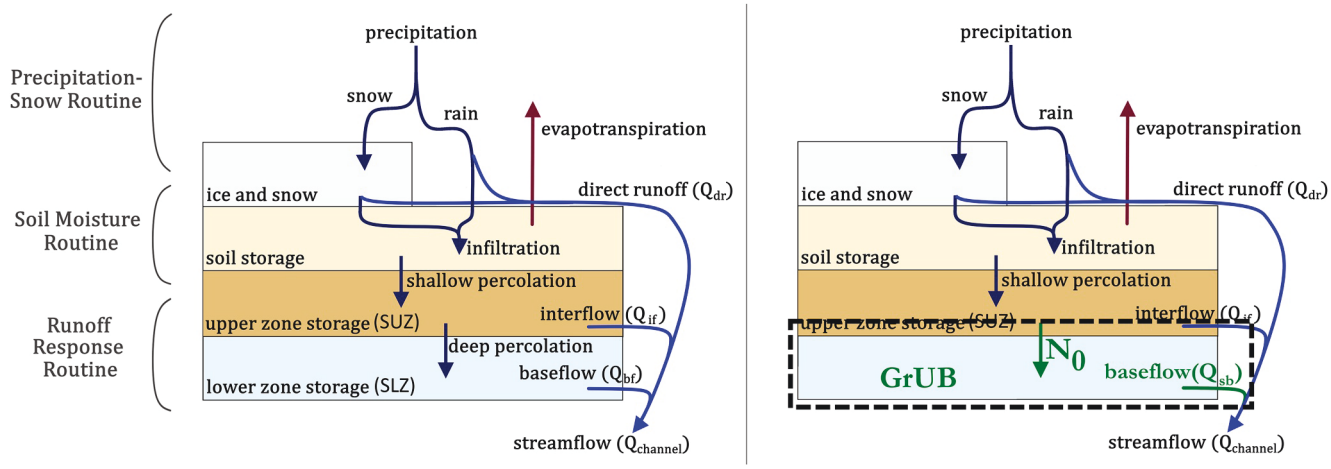

Figure 2. Conceptual model of HBV (left) and of HBV with GrUB replacing the groundwater module (right). GrUB stores and fluxes are highlighted in green.

4. Model Testing

4.1. Description of HBV

An initial assessment of the potential uncertainties and biases of the GrUB module necessitates analysis of a variety of flow conditions (i.e., long time periods) across a variety of physioclimates (i.e., a large number of basins). We chose to incorporate GrUB within the HBV model (Bergstrom, 1976, Bergström & Lindström, 2015) because of its modular structure that is parsimonious yet adaptable to a wide range of physical and climatic settings (Bergstrom, 1992). Specifically, we adapted HBV.IANIGLA (Toum, 2021) which interfaces with R software ("R Core Team," 2019). Apart from its appeal as being computationally simple and broadly applicable, we chose to implement HBV.IANIGLA because the subroutines it uses to generate streamflow are conceptually similar to those used in Noah-MP and other common LSMs (Figure 2). Here, we give a brief outline of HBV model structure, though for a detailed description we direct readers to Bergström (1976, 1992) and Bergström and Lindström (2015).

HBV operates as a sequence of subroutines with parameters within each subroutine requiring calibration (Sælthun, 1996; Figure 2). HBV.IANIGLA requires two time series of climate data (mean temperature [T_{mean}] and precipitation [PPT]) to drive the model, and a single time series (generally streamflow (Q)) for calibration. The first routine ("Precipitation-Snow Routine") calculates rainfall, snowfall, snowmelt, and potential evapotranspiration (PET) from T_{mean} and PPT. The second routine ("Soil Moisture Routine") calculates actual evapotranspiration (AET) and infiltration according to inputs from the first routine. The third and final routine ("Runoff Response Routine") calculates direct runoff (Q_{dr}), interflow (Q_{if}), and baseflow (Q_{bf}) according to inputs from the first two routines. The three runoff components are each independently calculated according to linear discharge from each of three conceptual buckets, with total streamflow ($Q_{channel}$) calculated as their sum. The parameters requiring calibration for each subroutine are listed in Table 2.

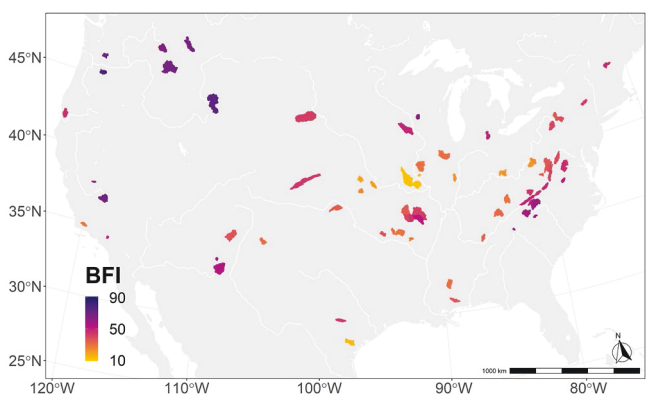

Figure 3. Map of MOPEX watersheds used in this study. Watersheds are highlighted according to their baseflow index (BFI [-]) as calculated using an automated hydrograph separation program (Wahl & Wahl, 1995) and reported by Wolock (2003).

A benefit of the minimal data requirements of HBV is that a large number of watersheds are available for model calibration and assessment. We chose to use data from the Model Parameter Estimation Experiment (MOPEX) data set (Schaake et al., 2006) due to their strict data standards. Of the 438 MOPEX basins for which hydrometeorological data are available, we selected 84 based on: (a) being designated as minimally impacted by human interference in the GAGES-II dataset (Falcone et al., 2010) and (b) having a minimum of 20 yr of data without data quality flags. These watersheds are illustrated in Figure 3.

Table 1
List of Parameters and Variables Used in GrUB

Parameter or Variable	Description	Derivation
N_0	Depth of recharge at the current time step	Input from overlying model
K_{dry}, K_{wet}	Effective saturated hydraulic conductivity during historically dry and wet periods	Tashie et al. (2021)
K_S, K_0	Effective saturated hydraulic conductivity at a particular value of storage (S), with K_0 being K at $S = 0$	Calculated using Tashie et al. (2021)
m	Recession constant (slope) relating S to $\log_{10}(K_S)$	Calculated using Tashie et al. (2021)
S_{dry}, S_{wet}	Effective drainable storage during historically dry and wet periods	Tashie et al. (2021)
S	Storage per unit area in the hillslope aquifer	Varies with model
f	Drainable porosity [-]	Set to 0.1
i	Slope of the impermeable layer underlying a hillslope aquifer (approximated as topographic slope)	USGS [2004]
L	Length of the stream network	USGS [2004]
B	Breadth of hillslope aquifer; note this term is often identified by the letter "L" (or "length") in hillslope hydraulics equations	USGS [2004]
H_0	Saturated aquifer thickness at the stream-hillslope interface	Varies with model
d	Ratio for approximating the minimum H_0 in the absence of recharge	Set to 0.001
v	Parameter for approximating the effects of antecedent recharge on H_0	Set to 0.95

For each watershed, we calibrated the 14 parameters from Table 1 using the R-software package hydroGOF (Zambrano-Bigiarini, 2020). First, we generated a uniform distribution for each variable, with each distribution being constrained to a realistic value per HBV model description (see Table 2). For calibration, we applied a Monte-Carlo approach, generating 10,000 random samples from each distribution. We removed the first 3 yr of data from each watershed before calculating the Kling-Gupta efficiency (KGE; Gupta et al., 2009) scores to allow the groundwater stores to equilibrate. Model parameterization was chosen based optimal KGE calibration; because model *performance* was not assessed with KGE but instead with low flow metrics (see below) no data were reserved for validation using KGE.

4.2. Implementing GrUB in HBV

We implemented the GrUB module in HBV directly, by replacing the lower bucket (SLZ, see Table 1) and the lower bucket storage constant ($K2$) with Equation 11 (Figure 2). Specifically:

1. N_0 from Equation 11 was calculated according to the recharge term (Perc) from the intermediate bucket (SUZ) to the lower storage bucket (SLZ) in HBV.IANIGLA.
2. Q_{sb} from Equation 11 replaced Q_{bf} in HBV.IANIGLA.
3. S from Equation 11 replaced SLZ in HBV.IANIGLA and was calculated as $S = S_{t-1} + N_0 - Q_{sb}$

While the groundwater module native to HBV was fully calibrated for each watershed, the GrUB module was never calibrated in any watershed. When incorporating GrUB into HBV, we first calibrated HBV using all 14 parameters from Table 1, then replaced the HBV groundwater module with GrUB. In subsequent sections, we refer to this implementation of the GrUB module within the HBV model simply as "GrUB" for convenience (Figure 4).

4.3. Model Evaluation

Broadly, we evaluated the performance of HBV and GrUB in two ways: (a) according to their ability to represent streamflow during the entire period of record (i.e., "general model performance"); and (b) according to their ability to accurately represent streamflow during dry periods (i.e., "low flow performance"). We identified

Table 2
Parameters Used to Calibrate HBV

Parameter name	Parameter description	Routine	Minimum value	Maximum value	Median calibrated value
SFCF	Snowfall correction factor [-]	Precipitation-snow routine	0.2	3	0.98
TR	Solid-liquid PPT threshold temperature [°C]	Precipitation-snow routine	−6	6	−2.4
TT	Melt temperature [°C]	Precipitation-snow routine	−6	6	0.76
FM	Snowmelt factor [mm/°C]	Precipitation-snow routine	0.2	8	1.6
FI	Icemelt factor [mm/°C]	Precipitation-snow routine	0.2	10	1.3
FIC	Debris-covered icemelt factor [mm/°C]	Precipitation-snow routine	2	10	6.0
FC	Soil field capacity [mm]	Soil moisture routine	25	1200	140
LP	AET correction factor [-]	Soil moisture routine	0.2	1	0.55
Beta	Soil storage-runoff exponential [-]	Soil moisture routine	1	3	2.2
K0	Top bucket (STZ) storage constant [t^{-1}]	Runoff response routine	0.05	1	0.46
K1	Intermediate bucket (SUZ) storage constant [t^{-1}]	Runoff response routine	0.005	0.5	0.22
K2	Lower bucket (SLZ) storage constant [t^{-1}]	Runoff response routine	0.0001	0.1	0.021
UZL	Max flux rate between STZ and SUZ [mm/day]	Runoff response routine	0.2	40	12
Perc	max flux rate between SUZ and SLZ [LT^{-1}]	Runoff response routine	0.1	20	0.29

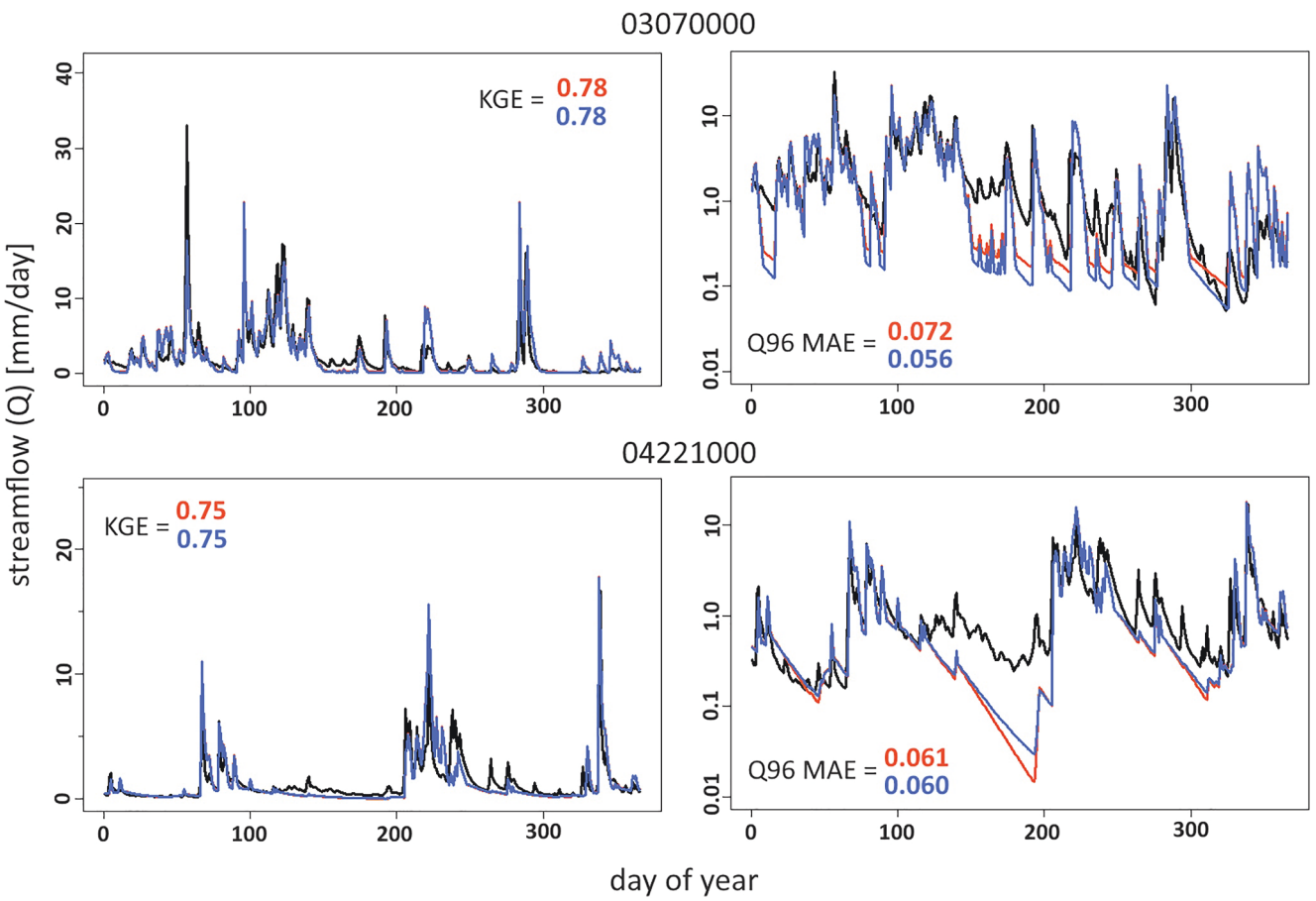

Figure 4. Example of model results for a single water year in two watershed: (1) USGS 03070000 Cheat River at Rowlesburg, WV; (2) USGS 04221000 Genesee River at Wellsville, NY. Gaged discharge is illustrated in black, HBV results in red, and GrUB in blue. Plots on the right are in log space to highlight low flows.

periods of "low flow" according to two different methods. First, we applied the Tennant method (Tennant, 1976), which is the most widely used method for defining minimum environmental instream flow in the United States (Jowett, 1997). Minimum environmental instream flows are calculated as 30% of the mean annual flow for the period of record. Second, we applied a quantile threshold approach (Praskievicz et al., 2018). Historical Q was ranked from largest to smallest, and streamflow records were selected based on an exceedance threshold of 96% (Q_{96}), which is often also referred to as the lower fourth percentile of flow. We also calculated the time required to run GrUB in each watershed using the base R function "system.time()" with the average user time being 0.065 s/yr per watershed.

To assess general model performance, we relied of four common metrics. The Nash-Sutcliffe model efficiency coefficient (NSE; Nash & Sutcliffe, 1970), which is traditionally the most common metric for evaluating hydrologic model performance. NSE varies from negative infinity to 1, with one indicating perfect correlation and 0 representing the use of mean streamflow as the benchmark predictor. We also used the Kling-Gupta efficiency coefficient (KGE; noted above), which is becoming increasingly popular as a more balanced measure than NSE (Knoben et al., 2019). We also report mean absolute error (MAE) and absolute bias (Bias) both for general model performance and for low flow performance.

To assess whether the error metric distributions generated by HBV and GrUB were significantly different ($p < 0.1$) from one another, we relied on the non-parametric Wilcoxon signed-rank test (Wilcoxon, 1945). In the supplemental tables (Tables S1, S2, S3, S4, S5, and S6), we also report significance values from the two-sample Kolmogorov-Smirnov (KS) test (Smirnov, 1948). However, in our primary analysis (i.e., main text and figures) we rely on the Wilcoxon test as opposed to the KS test, as the former prioritizes (possible) differences in the median values of the distributions while the latter prioritizes differences in the shapes of the distributions.

4.4. Sensitivity to Rate of Recharge

As our primary goal is to develop a groundwater module that is generally adaptable to a variety of LSMs and rainfall-runoff models without the need for fine tuning or calibration, it is essential that we "stress test" our module to uncertainty in the structure of the overlying hydrologic model. Because recharge (N) is the only component of the overlying hydrologic model that interacts directly with GrUB, we chose to focus on components of the HBV model structure that constrain rates and magnitude of N. Specifically, we recalculated the rate at which infiltrated water percolates through the subsurface according to two end-member scenarios:

1. Slow Recharge: UZL and Perc (see Table 2) set to maximum rates of 0.1 mm/day.
2. Fast Recharge: UZL and Perc set to minimum rates of 1,000 mm/day.

Recall (Table 2 and Figure 2) that UZL constrains the maximum rate of movement from soil storage to the upper storage zone (SUZ) and Perc similarly constrains the maximum rate of movement from the SUZ to the lower storage zone (SLZ). Therefore, in the Slow Recharge scenario N is constrained to a small, nearly constant rate of 0.1 mm/day at maximum. Meanwhile, the Fast Recharge scenario effectively represents all water reaching the water table immediately following percolation below the shallow soils store. For all three models, we altered the values of UZL and Perc but otherwise kept all parameter values the same (i.e., the values listed in Table 2).

5. Model Results

5.1. Model Performance

For the entire simulation period, HBV generated general evaluation results that were quite good (KGE: median = 0.74; Figure 5). Average across all watersheds, absolute bias (Bias) was extremely low (Bias: median = 0.002 mm/day, mean = −0.008 mm/day) indicating high accuracy in the overall water balance. MAE was below 1 mm/day in most watersheds, with a median value of 0.61 mm/day. Because the models were optimized on KGE, NSE evaluation results (which place a relatively higher priority on representing periods of high flow) were somewhat lower though still in a generally acceptable range (NSE: median = 0.53).

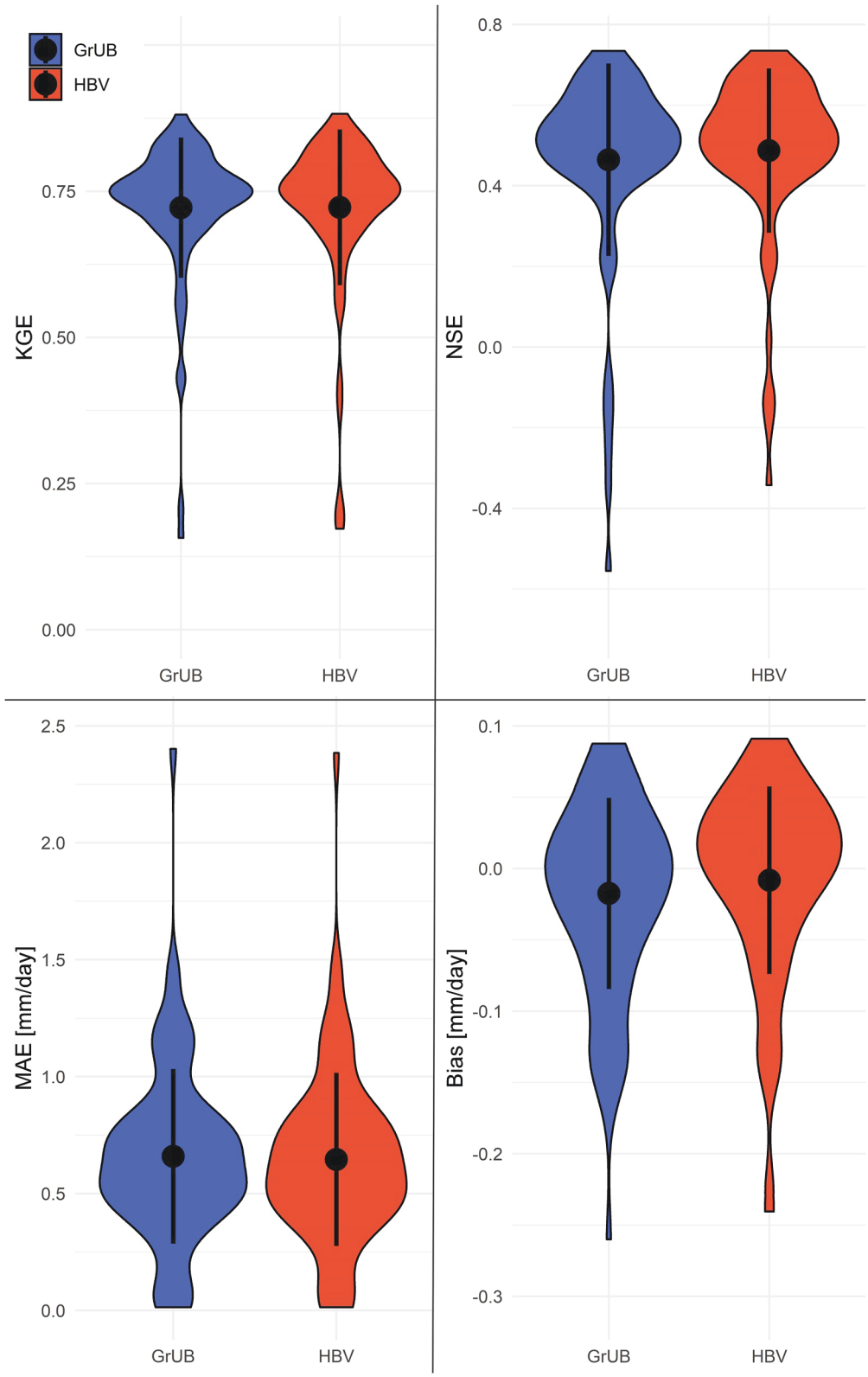

Figure 5. Violin plots of model evaluation results. Black dots represent mean values and black bars represent one standard deviation around the mean. Blue illustrates results from GrUB and red from HBV with the native calibrated groundwater module. Model performance results are (clockwise from top left) the Kling-Gupta efficiency (KGE), Nash-Sutcliffe efficiency (NSE), absolute bias (Bias), and mean absolute error (MAE). *Note.* That no GrUB error metric distributions are significantly ($p < 0.1$) different from those of the basic HBV model.

For the entire simulation period, the uncalibrated GrUB module generated nearly identical error metric distributions to the calibrated HBV model, with no distribution generated by GrUB being significantly ($p < 0.1$) different from those generated by HBV (Figure 5; Table S1).

During periods of low flow, HBV generated MAE that were usually below 0.2 mm/day (Q_{96} MAE: median = 0.11; Tennant MAE: median = 0.18; Figure 6). However, HBV low flow estimates exhibited a consistently positive Bias for nearly all watersheds (Q_{96} Bias: median = 0.07; Tennant Bias: median = 0.10).

During periods of low flow, GrUB generated marginally lower MAE during extremely dry periods (Q_{96} MAE: median = 0.09) although Tennant low flow MAEs were marginally higher (Tennant MAE: median = 0.19) and none of these differences was statistically significant ($p > 0.1$). However, according to both methods of low flow delineation, Bias was significantly improved ($p < 0.1$). Typical Bias of Q_{96} dropped to 0.02, and for the Tennant methods Bias dropped to 0.06. That is, the uncalibrated GrUB module generates low flow predictions that are similarly precise as and significantly more accurate than those produced by the fully calibrated HBV model.

5.2. Alternative Scenario: Slow Recharge

For the entire period of record, HBV generated lower KGE and higher error metrics under the Slow Recharge scenario (Figure 7) than it did under the optimized scenario (Figure 5). Median KGE dropped to 0.63, MAE jumped to 0.97 mm/day, and NSE dropped to 0.25, while Bias remained about the same at 0.003. This indicates that artificially slowing recharge had a minimal impact on the overall water balance but caused a major decline in the ability of HBV to predict the timing of streamflow, as should be expected.

The overall effects were nearly identical for GrUB (Figure 7), with no error metric distribution generated by GrUB being significantly different than that generated by HBV.

For low flows, the effects on HBV were mixed. Somewhat surprisingly, Q_{96} MAE improved marginally (median = 0.09) while median Bias dropped by two-thirds to 0.02 (Figure 8). Meanwhile, Tennant median MAE worsened to 0.22 while Bias again dropped (median = 0.06). Recall that the rates of recharge in this scenario are artificially and unrealistically low. This indicates that there may be an inherent trade-off in HBV in terms of making accurate predictions of the timing of high flows versus low flows, with the chosen calibration routine being geared toward better representation of the former.

GrUB predictions of low flows changed very little under conditions of artificially slow recharge (Figure 8). Q_{96} median MAE stayed about the same (0.09) and Tennant median MAE worsened slightly to 0.21, while Bias also stayed about the same (Q_{96} Bias: 0.001; Tennant Bias: 0.05). Under this scenario, Bias of GrUB predictions of Q_{96} low flows (median = 0.0001) remained significantly better than that of HBV. For all other error metrics, the distributions of GrUB were not statistically different from those of HBV.

5.3. Alternative Scenario: Fast Recharge

For the entire period of record, HBV generated worse KGE, NSE, MAE, and Bias under the Fast Recharge scenario (Figure 9) than it did under the optimized scenario (Figure 5). In particular, the declines in values of KGE and NSE were extreme (KGE: median = 0.08; NSE: median = 0.10). Median MAE rose to 0.85 mm/day and median Bias quadrupled to 0.01 mm/day.

While error metric distributions also worsened for GrUB, the changes were generally less extreme. Median Bias went slightly negative to −0.0006 mm/day, while MAE rose to 0.75 mm/day. Median NSE values dropped precipitously to 0.11, but median KGE remained at a somewhat healthier value of 0.26. Under this scenario, KGE values for GrUB were significantly better than those generated by HBV (Figure 9).

For low flows, all HBV error metrics were two to five times worse under the Fast Recharge scenario (Figure 10) than they were under the optimized scenario (Figure 6). For both Q_{96} and Tennant, median MAE more than doubled (Q_{96}: 0.28; Tennant: 0.44) while Bias quadrupled (Q_{96}: 0.26; TNT: 0.42).

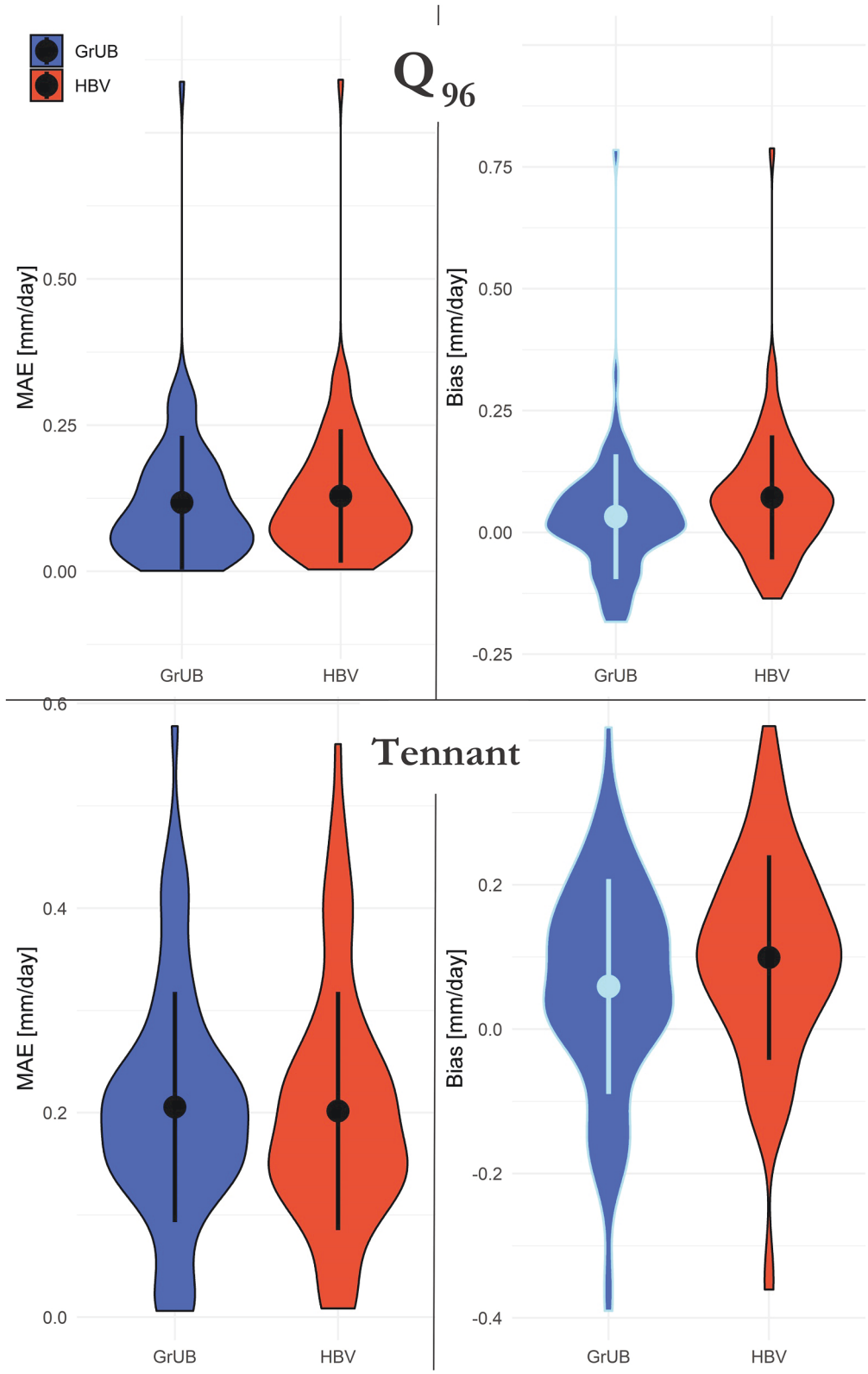

Figure 6. Violin plots of model predictions of low flows defined as 96% exceedance (Q_{96}) flows (top row) and according to the Tennant method (bottom row). Blue illustrates results from GrUB and red from HBV with the native calibrated groundwater module. Model performance results are mean absolute error (MAE—left) and absolute bias (Bias—right). GrUB error metric distributions that are significantly ($p < 0.1$) different from those of the basic HBV model are highlighted in cyan.

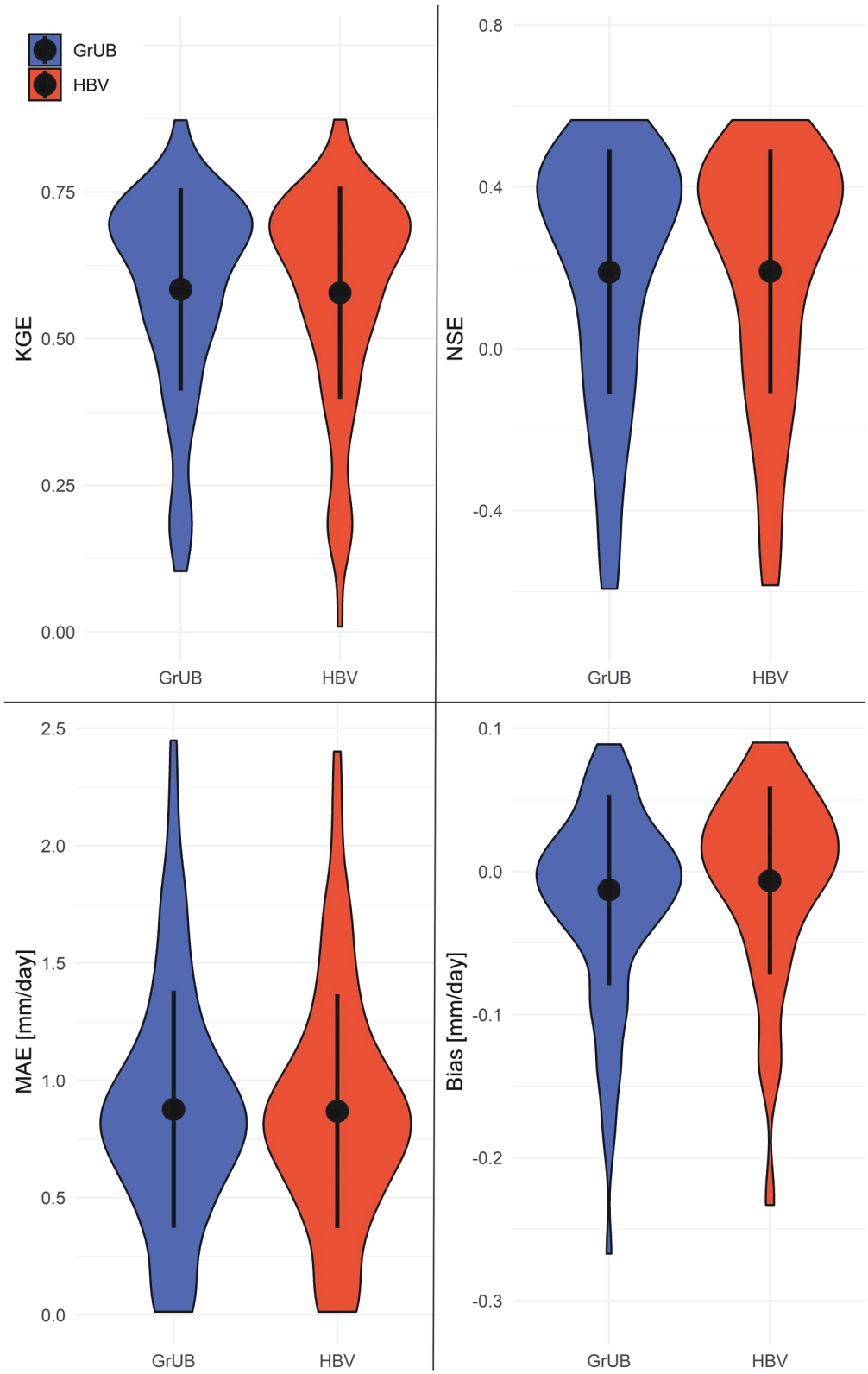

Figure 7. Violin plots of model performance under the alternative scenario of Slow Recharge. Black dots represent mean values and black bars represent one standard deviation around the mean. Blue illustrates results from GrUB and red from HBV with the native calibrated groundwater module. Model performance results are (clockwise from top left) the Kling-Gupta efficiency (KGE), Nash-Sutcliffe efficiency (NSE), absolute bias (Bias), and mean absolute error (MAE). *Note.* That no GrUB error metric distributions are significantly ($p < 0.1$) different from those of the basic HBV model.

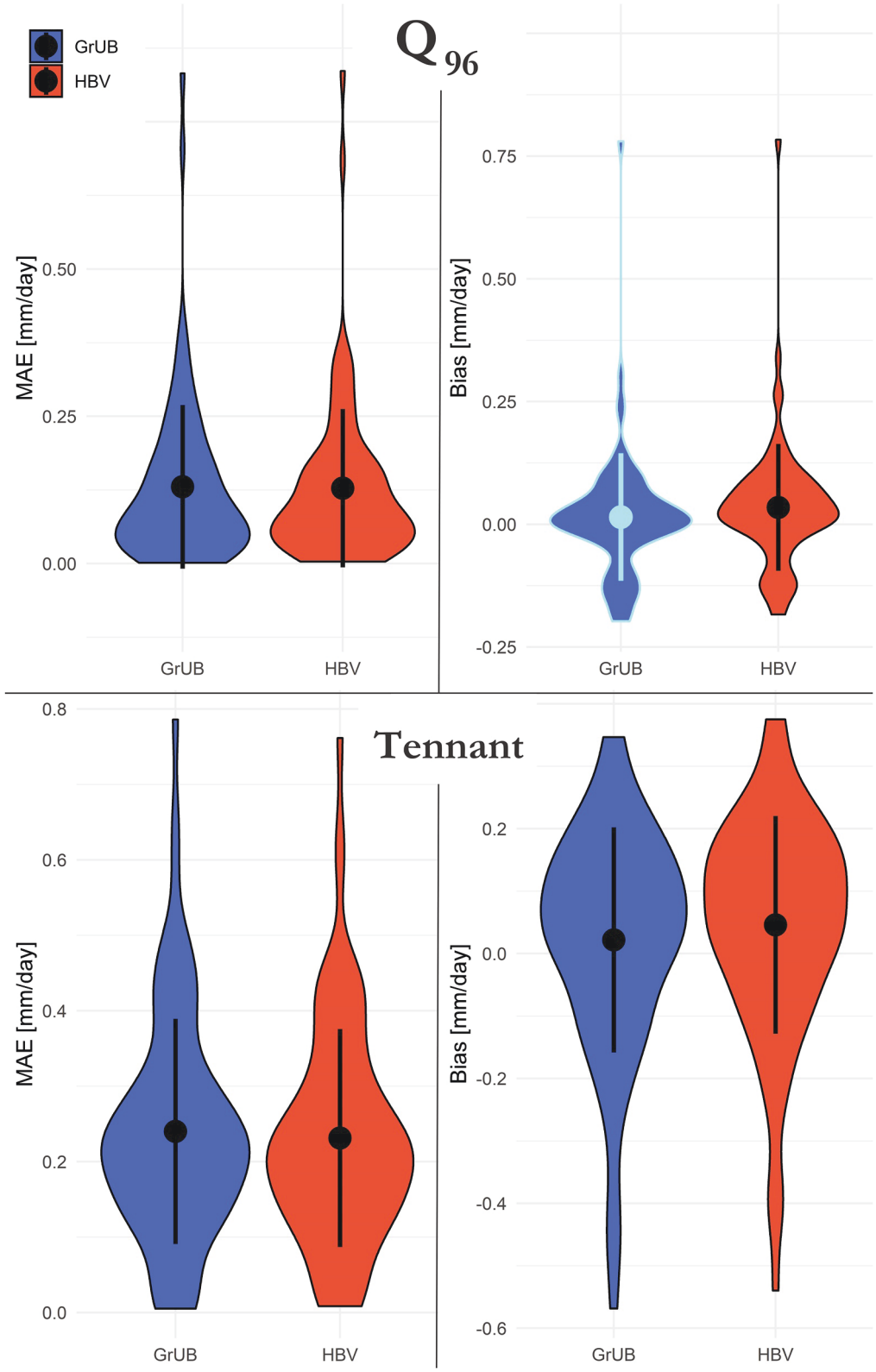

Figure 8. Violin plots of model predictions of low flows under the alternative scenario of Slow Recharge. Low flows are defined as 96% exceedance flows (Q_{96}—top row) and according to the Tennant method (bottom row). Blue illustrates results from GrUB and red from HBV with the native calibrated groundwater module. Model performance results are mean absolute error (MAE—left) and absolute bias (Bias—right). GrUB error metric distributions that are significantly ($p < 0.1$) different from those of the basic HBV model are highlighted in cyan.

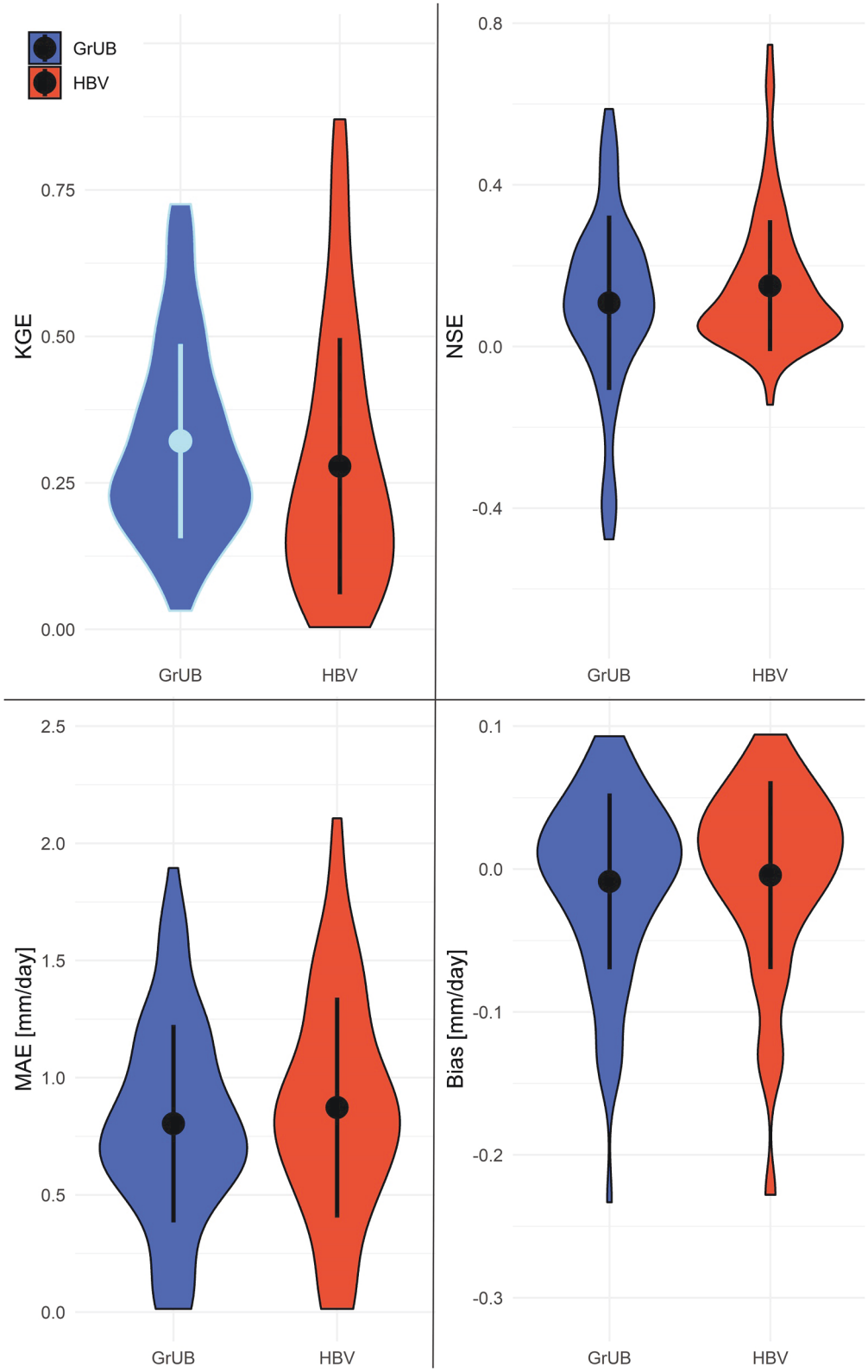

Figure 9. Violin plots of model performance under the alternative scenario of Fast Recharge. Black dots represent mean values and black bars represent one standard deviation around the mean. Blue illustrates results from GrUB and red from HBV with the native calibrated groundwater module. Model performance results are (clockwise from top left) the Kling-Gupta efficiency (KGE), Nash-Sutcliffe efficiency (NSE), absolute bias (Bias), and mean absolute error (MAE). GrUB error metric distributions that are significantly ($p < 0.1$) different from those of the basic HBV model are highlighted in cyan.

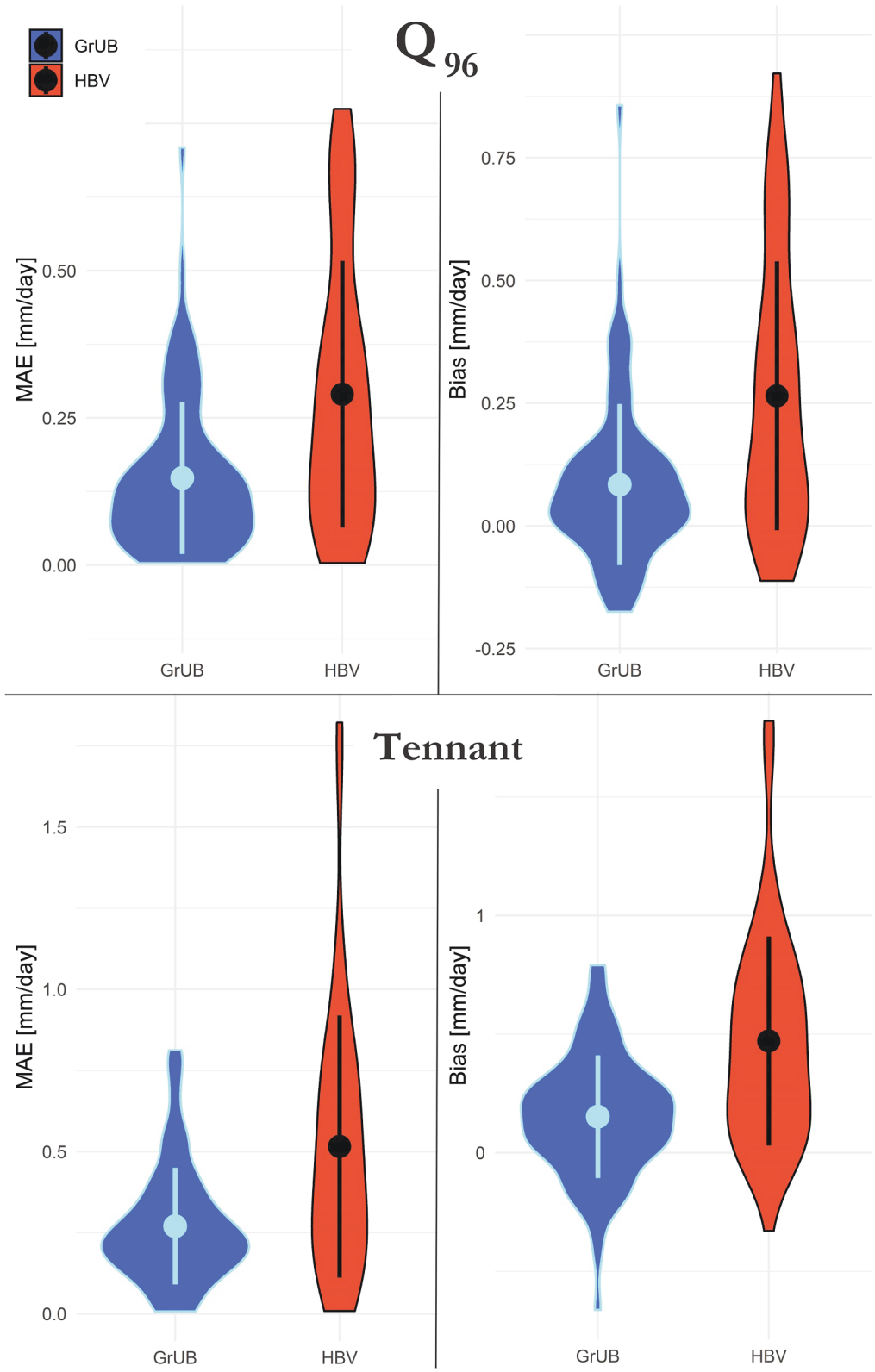

Figure 10. Violin plots of model predictions of low flows under the alternative scenario of Fast Recharge. Low flows are defined as 96% exceedance (Q_{96}) flows (top row) and according to the Tennant method (bottom row). Blue illustrates results from GrUB and red from HBV with the native calibrated groundwater module. Model performance results are mean absolute error (MAE—left) and absolute bias (Bias—right). GrUB error metric distributions that are significantly ($p < 0.1$) different from those of the basic HBV model are highlighted in red.

GrUB predictions of low flows also worsened, but to a much lesser degree. For Q_{96} and Tennant, median MAE only increased by 15%–35% (Q_{96}: 0.12; Tennant: 0.22) while median Bias doubled (Q_{96}: 0.05; Tennant: 0.15). For all low flow metrics, GrUB error metric distributions were significantly ($p < 0.1$) better than those produced by HBV (Figure 10).

5.4. Sensitivity to Model Parameterization

GrUB relies on two parameters (v and d) that are not derived from empirical data, but instead estimated as universal constants. We tested sensitivity to these parameters by re-running GrUB for all watersheds with a wide range of parameter values. The range of parameter values was chosen to extend to (or beyond) values that are physically realistic.

For v, which constrains the length of time that recharge (N) has a non-negligible effect on H_0 ($v^{-t} > 0.01$), we chose values of 0.52, 0.86, 0.95, 0.963, and 0.9875. These represent time periods of 1 week, 1 month, 1 season, 2 seasons, and 1 yr. For d, which represents the absolute minimum fraction of the hillslope aquifer thickness (Th) that remains activated at the stream-aquifer interface (H_o) during long periods without any recharge, we selected values 0.02, 0.005, 0.001, 0.0005, and 0.0001. Because S_{dry}/f represents the maximum value that H_{max} can take during dry periods, and median S_{dry}/f for the coterminous United States is estimated at about 2 m (Tashie et al., 2021), a value of $d = 0.02$ represents a typical $H_{minimum}$ of 40 mm, and a value of $d = 0.0001$ represents a typical $H_{minimum}$ of 0.2 mm.

Parameter values did have noticeable effects on the hydrographs of individual watersheds, though these differences did not express themselves as significant changes in the overall performance of GrUB across all watersheds (Figure 11). We assessed the significance ($p < 0.1$) of their effects on KGE (for the entire period of record) and MAE (during periods of low flow) using the Wilcoxon rank sum test (Wilcoxon, 1945) and the two-sample Kolmogorov-Smirnov test (Smirnov, 1948). None were significant, with the lowest p-values ($p = 0.45$) registering for changes in the variable v.

We also ran a limited test to assess the capacity of each of these parameters to be optimized via calibration. For each of the five values of v and d, we identified the best (and worst) parameter value for each watershed, as illustrated in Figures 11 and 12. Parameter value choice did not significantly affect the error metric distributions for KGE for either v or d. However, "calibration" of v does result in significantly improved predictions of low flow MAE ($p < 0.001$). Interestingly, optimal values of d are at the end member range of possible values (0.02 or 0.0001) with no values in the middle range (0.005, 0.001, and 0.0005) being the optimal choice for any watershed. Optimal parameter values also show distinct regional clustering (Figure 12), with lower values of both v and d being optimal in the southern Appalachian Mountains and in portions of the Western United States, and higher values of each being optimal in most other regions.

6. Model Performance

We assess the overall performance of the GrUB module not in terms of its absolute performance in maximizing the objective function (i.e., KGE) as this feat is largely achieved by the overlying hydrologic model (i.e., HBV). Instead, we assess its capacity to conform to the four key development objectives (a–d) while achieving the three key performance metrics (i–iii) outlined in the introduction. The four key development objectives that guided the development of the GrUB module are as follows:

1. *No calibration required*: The wholly uncalibrated GrUB module is fully compatible with the otherwise calibrated HBV model.
2. *Simple data requirements*: Parameterization of GrUB only requires data from several freely available continental- and global-scale datasets and requires minimal processing on the part of the model user. Unfortunately, three of the required parameters (K_0, m, and $S_{max/min}$) are, to our knowledge, currently only available for the coterminous United States. Until these data are developed globally, the practical application of GrUB may be limited to this region.
3. *Modularity*: GrUB requires input from only a single, common flux term from the driving hydrologic model (i.e., deep recharge) and otherwise operates independently from the rest of the model structure.

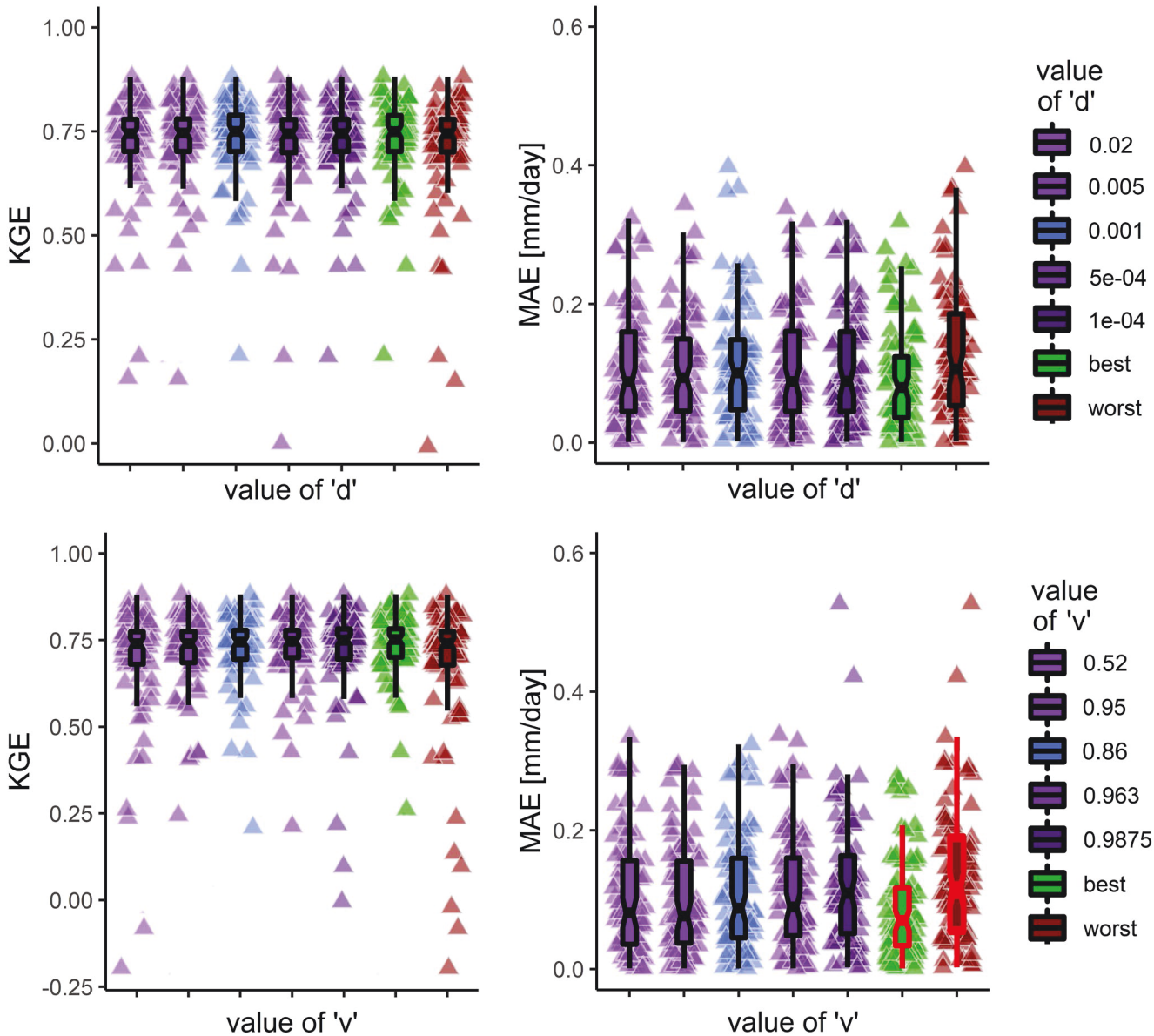

Figure 11. Boxplots of GrUB sensitivity to parameter selection. The left column illustrates KGE for the entire period of record. The right column illustrates mean absolute error (MAE) during low flows as defined by the Q_{96} method. The top row is for different values of "*d*" and the bottom row for different values of "*v*." The parameterization used in the primary analysis is shown in blue (in the third position), with each triangle representing a single value from a single watershed. Alternative parameter values are illustrated in shades of purple (in the first, second, fourth, and fifth positions). Optimal or "best" parameter values are aggregated then illustrated in green (sixth position) and "worst" parameter values are aggregated then illustrated in red (seventh position). Error metric distributions that are significantly different ($p < 0.1$) from those used in the primary analysis are highlighted in red.

4. *Computational simplicity*: Though GrUB (Equation 11) is more complex than the simple storage-discharge used in HBV, this added complexity is negligible in terms of the added computation time.

GrUB also largely achieves the three performance metric objectives, with some minor exceptions as outlined below. When comparing the performance of the fully calibrated HBV groundwater module with that of the wholly uncalibrated GrUB module:

1. Performance metrics for the entire period of record: overall model performance (i.e., high flows) is not significantly affected by incorporating the uncalibrated GrUB module into the calibrated HBV model.

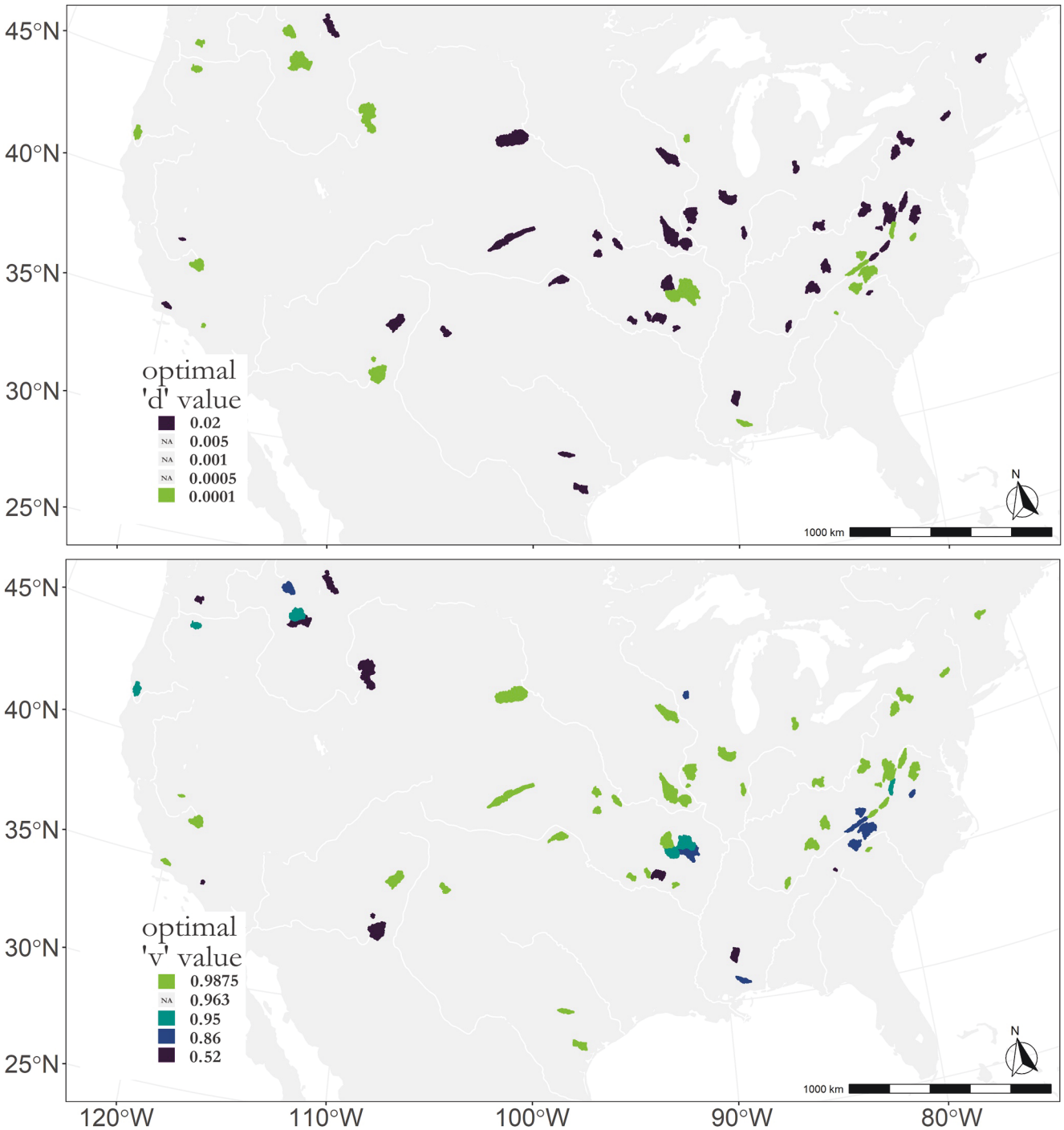

Figure 12. Optimal values of each parameter ("*d*" on top and "*v*" on bottom) for each watershed. *Note.* That several parameter values are not the optimal (i.e., lowest MAE during low flows as defined by Q_{96}) for any watershed.

2. Performance metrics for periods of low flows: the uncalibrated GrUB module generates low flow error metrics that tend to be the same as or better than those generated by the calibrated HBV module.
3. Sensitivity to overlying model inputs: GrUB is more robust to changes in the overlying hydrologic model parameterization than is the standard HBV storage-discharge module.

For the potential incorporation of GrUB into LSMs, these results are generally promising, though with some important caveats (see Section 6). That the wholly uncalibrated GrUB predicts low flows as well as (or better than) a simple calibrated module may prove beneficial to LSMs, which are broadly seen as describing low flows poorly without post hoc calibration (Fan et al., 2019; Holtzman et al., 2020). More important than the general performance, however, is the consistency with which a groundwater module predicts low flows when that module is forced by uncertain inputs from the overlying model structure. This "robustness" to changes in inputs from the overlying model is essential as different LSMs are structured and parameterized quite differently from each other (Clark et al., 2015).

GrUB is indeed extremely consistent in its prediction of low flows, as illustrated in the artificially extreme Slow Recharge and Fast Recharge scenarios (Sections 4.2 and 4.3). Recall that the Slow Recharge scenario represents percolation rates of about 0.1 mm/day, compared with up to 1,000 mm/day in the Fast Recharge scenario. Despite these potential rates of recharge spanning four orders of magnitude, error in GrUB predictions of extreme low flows did not vary significantly, with median MAE values differing by only about 20% across all three recharge scenarios (Figures 6, 8, and 10). Conversely, HBV error during low flows more varied by a factor of 2.8 across the same three scenarios. Even more importantly, GrUB Bias during low flows was both smaller on average and much more consistent across the three scenarios (GrUB median Q_{95} for original, Slow, and Fast Recharge: 0.02, 0.0001, and 0.05; HBV median Q_{95} for original, Slow, and Fast Recharge: 0.07, 0.02, and 0.26).

Presented results have mixed implications for the potential incorporation of GrUB into rainfall-runoff models. While low flow Bias in GrUB was slightly lower than that in HBV, GrUB did not significantly improve any other error metric. Therefore, we do not expect GrUB to enhance the performance of other calibrated rainfall-runoff models where sufficient data for calibration are available. However, rainfall-runoff models are commonly applied in data sparse regions with "regionalization" approaches replacing parameter tuning via calibration (e.g., Swain & Patra, 2017). Regionalization approaches tend to be subject to high uncertainty, with no single regionalization method yet accepted as a standard (Oudin et al., 2008; Samuel et al., 2011). Because GrUB is robust to changes in the parameterization of the overlying hydrologic model, it may prove useful in constraining model characterization of low flows in basins where regionalization approaches are necessary.

7. Limitations and Future Efforts

GrUB depends on three parameter values that are not currently available globally (K_0, m, and S_{wet}—see Table 1 for definitions). Though the methods to estimate these variables globally have been established (Tashie et al., 2021), until those data are available GrUB may not be applied outside the coterminous United States. GrUB also relies on L (stream network length), estimates of which are known to be extremely uncertain and which is known to vary with antecedent conditions by a factor of two or more (Godsey & Kirchner, 2014). While incorporating improved estimates of average L (Lin et al., 2021) may improve model performance, no large-scale data sets yet exist for describing its dynamic response to catchment conditions.

GrUB further relies on two parameters (v and d) that are treated as empirical constants. In many hydrologic models, these parameters might present an opportunity for fine tuning model behavior. However, fine-tuning and unique calibrations are antithetical to the purpose of GrUB. Even though GrUB is not overly sensitive to changes in the value of these parameters (Section 4.4), the fact that these parameters are estimated according to hydrologic "intuition" rather than directly calculated represents an unfortunate deviation from the otherwise empirically based structure of GrUB. Promisingly, the optimal value for each of these parameters tends to show distinct regional clustering (Figure 12) indicating possible underlying physical mechanisms that may be uncovered in future research.

Though GrUB does incorporate seasonal-scale "watershed memory" by calculating H_0 (saturated aquifer thickness at the stream-hillslope interface) as a function of recent recharge (Equation 10), it does not account for potential multiannual long-term memory of climatic forcings that have been documented in many watersheds (for example, Fowler et al., 2020). Climate change is expected to induce long-term changes in watershed storage (and related ecohydrological responses) that are not directly accounted for in current generation LSMs (Argus et al., 2017; Enzminger et al., 2019). In its current form, GrUB does not make significant strides toward addressing this deficiency. Current generation LSMs are also known to underestimate ET during dry periods. A potential

explanation for this is that the representation of watersheds according to a representative soil column fails to account for the heterogeneity in soil moisture and depth to water table that is induced in actual watersheds by topography, geology, and soil structure. Though a hillslope hydraulics model of groundwater has been invoked as a potential tool for better capturing ET during dry periods (Clark et al., 2015; Fan et al., 2019), we do not attempt to model such behavior here.

8. Conclusion

We develop a calibration-free, computationally simple module called GrUB to predict groundwater contributions to streamflow. GrUB may be readily incorporated into a variety of model structures that may be defined as lumped, conceptual rainfall-runoff models and distributed, physically based LSMs (Pechlivanidis et al., 2011). GrUB requires no calibration, but instead depends entirely on empirical data that is available for the entire coterminous United States and could potentially be derived globally. We assess the performance of GrUB in over 80 United States watersheds by incorporating it into HBV, a popular rainfall-runoff model, and comparing overall performance metrics as well as error in predictions of low flows by the native (calibrated) HBV groundwater module and those by the (uncalibrated) GrUB module. GrUB generates error metrics that are equivalent to (or superior to) those generated by the calibrated HBV groundwater module. To ensure that predictions by GrUB are robust to changes in the structure and parameterization of the overlying hydrologic model, we run tests according to two artificial scenarios: Slow Recharge at a rate of up to 0.1 mm/day, and Fast Recharge at a rate of up to 1,000 mm/day. GrUB proves to be very robust to these extreme changes, with MAE of predictions of low flows only increasing by an average of 10% and 15%, respectively, in the Slow and Fast recharge scenarios. We suggest GrUB as a potential tool for improving predictions of low flows in LSMs that cannot be calibrated, as well as rainfall-runoff models when applied in locations where calibration data are sparse.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Data Availability Statement

All data used in this study are publicly available on: United States Geological Survey (USGS) National Water Information System at https://waterdata.usgs.gov/nwis/rt (for stream discharge); CUAHSI's HydroShare platform at https://www.hydroshare.org/resource/115409dbe8354e78a2c2219d32e2b9de (for hydraulic properties) and at https://www.hydroshare.org/resource/99d5c1a238134ea6b8b767a65f440cb7 (for MOPEX data); and the USGS National Hydrography repository at https://www.usgs.gov/core-science-systems/ngp/national-hydrography/nhdplus-high-resolution (for catchment characteristics). All code used to read, analyze, and plot data is publicly available under doi: 10.5281/zenodo.5019561 at https://zenodo.org/record/5019561#.YNNpYUwpBhE. A standalone copy of the code used to run GrUB in a modular fashion is available on GitHUB at https://github.com/atashie/LSM_gw_module/blob/main/GrUB_module.txt.

Acknowledgments
This research was funded by NSF grants EAR-1920425 and OIA-2019561.

References

Argus, D. F., Landerer, F. W., Wiese, D. N., Martens, H. R., Fu, Y., Famiglietti, J. S., et al. (2017). Sustained water loss in California's mountain ranges during severe drought from 2012 to 2015 inferred from GPS. *Journal of Geophysical Research: Solid Earth*, *122*(12), 10–559. https://doi.org/0.1002/2017JB014424

Aulenbach, B. T., Hooper, R. P., van Meerveld, H. J., Burns, D. A., Freer, J. E., Shanley, J. B., et al. (2021). The evolving perceptual model of streamflow generation at the Panola Mountain Research Watershed. *Hydrological processes*, *35*(4), e14127. https://doi.org/10.1002/hyp.14127

Baker, I. T., Prihodko, L., Denning, A. S., Goulden, M., Miller, S., & da Rocha, H. R. (2008). Seasonal drought stress in the Amazon: Reconciling models and observations. *Journal of Geophysical Research*, *113*, G00B01. https://doi.org/10.1029/2007JG000644

Balkovič, J., Skalský, R., Folberth, C., Khabarov, N., Schmid, E., Madaras, M., et al. (2018). Impacts and uncertainties of +2°C of climate change and soil degradation on European crop calorie supply. *Earth's Future*, *6*(3), 373–395. https://doi.org/10.1002/2017EF000629

Bart, R., & Hope, A. (2014). Inter-seasonal variability in baseflow recession rates: The role of aquifer antecedent storage in central California watersheds. *Journal of Hydrology*, *519*, 205–213. http://doi.org/10.1016/j.jhydrol.2014.07.020

Beck, H. E., van Dijk, A. I., Miralles, D. G., De Jeu, R. A., Bruijnzeel, L. S., McVicar, T. R., & Schellekens, J. (2013). Global patterns in base flow index and recession based on streamflow observations from 3394 catchments. *Water Resources Research*, *49*(12), 7843–7863. https://doi.org/10.1002/2013WR013918

Bergstrom, S. (1976). *Development and application of a conceptual runoff model for Scandinavian catchments*. Swedish Meteorological and Hydrological Institute.

Bergstrom, S. (1992). *The HBV model—Its structure and applications* (Report RH 4). Swedish Meteorological and Hydrological Institute.
Bergström, S., & Lindström, G. (2015). Interpretation of runoff processes in hydrological modelling—Experience from the HBV approach. *Hydrological Processes*, *29*(16), 3535–3545. https://doi.org/10.1002/hyp.10510
Beven, K. J. (2006). A manifesto for the equifinality thesis. *Journal of Hydrology*, *320*, 18–36. https://doi.org/10.1016/j.jhydrol.2005.07.007
Beven, K. J., & Kirkby, M. J. (1979). A physically based, variable contributing area model of basin hydrology. *Hydrological Sciences Bulletin*, *24*(1), 43–69. https://doi.org/10.1080/02626667909491834
Boughton, W. (2006). Calibrations of a daily rainfall-runoff model with poor quality data. *Environmental Modelling & Software*, *21*(8), 1114–1128. http://doi.org/10.1016/j.envsoft.2005.05.011
Brunke, M. A., Broxton, P., Pelletier, J., Gochis, D., Hazenberg, P., Lawrence, D. M., et al. (2016). Implementing and evaluating variable soil thickness in the Community Land Model, version 4.5 (CLM4.5). *Journal of Climate*, *29*, 3441–3461. https://doi.org/10.1175/JCLI-D-15-0307.1
Clark, M. P., Fan, Y., Lawrence, D. M., Adam, J. C., Bolster, D., Gochis, D. J., et al. (2015). Improving the representation of hydrologic processes in Earth System Models. *Water Resources Research*, *51*(8), 5929–5956. https://doi.org/10.1002/2015WR017096
Clark, M. P., Rupp, D. E., Woods, R. A., Tromp-van Meerveld, H. J., Peters, N. E., & Freer, J. E. (2009). Consistency between hydrological models and field observations: Linking processes at the hillslope scale to hydrological responses at the watershed scale. *Hydrological Processes: International Journal*, *23*(2), 311–319. https://doi.org/10.1002/hyp.7154
Cui, T., Yang, T., Xu, C. Y., Shao, Q., Wang, X., & Li, Z. (2018). Assessment of the impact of climate change on flow regime at multiple temporal scales and potential ecological implications in an Alpine river. *Stochastic Environmental Research and Risk Assessment*, *32*(2). https://doi.org/10.1007/s00477-017-1475-z
de Jong, P., Tanajura, C. A. S., Sánchez, A. S., Dargaville, R., Kiperstok, A., & Torres, E. A. (2018). Hydroelectric production from Brazil's São Francisco River could cease due to climate change and inter-annual variability. *The Science of the Total Environment*, *634*, 1540–1553. https://doi.org/10.1016/j.scitotenv.2018.03.256
Emanuel, R. E. (2018). Climate change in the Lumbee River watershed and potential impacts on the Lumbee tribe of North Carolina. *Journal of Contemporary Water Research & Education*, *163*(1), 79–93. https://doi.org/10.1111/j.1936-704X.2018.03271.x
Enzminger, T. L., Small, E. E., & Borsa, A. A. (2019). Subsurface water dominates Sierra Nevada seasonal hydrologic storage. *Geophysical Research Letters*, *46*(21), 11993–12001. https://doi.org/10.1029/2019GL084589
Falcone, J. A., Carlisle, D. M., Wolock, D. M., & Meador, M. R. (2010). Gages: A stream gage database for evaluating natural and altered flow conditions in the conterminous United States: Ecological archives E091-045. *Ecology*, *91*(2), 621. https://doi.org/10.1890/09-0889.1
Fan, Y., Clark, M., Lawrence, D. M., Swenson, S., Band, L. E., Brantley, S. L., et al. (2019). Hillslope hydrology in global change research and Earth system modeling. *Water Resources Research*, *55*(2), 1737–1772. https://doi.org/10.1029/2018WR023903
Fang, K., Ji, X., Shen, C., Ludwig, N., Godfrey, P., Mahjabin, T., & Doughty, C. (2019). Combining a land surface model with groundwater model calibration to assess the impacts of groundwater pumping in a mountainous desert basin. *Advances in Water Resources*, *130*, 12–28. https://doi.org/10.1016/j.advwatres.2019.05.008
Ficklin, D. L., Maxwell, J. T., Letsinger, S. L., & Gholizadeh, H. (2015). A climatic deconstruction of recent drought trends in the United States. *Environmental Research Letters*, *10*(4), 044009. http://doi.org/10.1088/1748-9326/10/4/044009
Flörke, M., Schneider, C., & McDonald, R. I. (2018). Water competition between cities and agriculture driven by climate change and urban growth. *Nature Sustainability*, *1*(1), 51–58. https://doi.org/10.1038/s41893-017-0006-8
Fowler, K., Knoben, W., Peel, M., Peterson, T., Ryu, D., Saft, M., et al. (2020). Many commonly used rainfall-runoff models lack long, slow dynamics: Implications for runoff projections. *Water Resources Research*, *56*(5), e2019WR025286. https://doi.org/10.1029/2019WR025286
Freeze, R. A., & Cherry, J. A. (1979). *Groundwater* (p. 604). Prentice-Hall.
Gan, Y., Liang, X. Z., Duan, Q., Chen, F., Li, J., & Zhang, Y. (2019). Assessment and reduction of the physical parameterization uncertainty for Noah-MP land surface model. *Water Resources Research*, *55*(7), 5518–5538. https://doi.org/10.1029/2019WR024814
Gedney, N., & Cox, P. M. (2003). The sensitivity of global climate model simulations to the representation of soil moisture heterogeneity. *Journal of Hydrometeorology*, *4*, 1265–1675. https://doi.org/10.1002/eco.1362
Godsey, S. E., & Kirchner, J. W. (2014). Dynamic, discontinuous stream networks: Hydrologically driven variations in active drainage density, flowing channels and stream order. *Hydrological Processes*, *28*(23), 5791–5803. https://doi.org/10.1002/hyp.10310
Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, *377*(1–2), 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003
Hare, D. K., Helton, A. M., Johnson, Z. C., Lane, J. W., & Briggs, M. A. (2021). Continental-scale analysis of shallow and deep groundwater contributions to streams. *Nature Communications*, *12*, 1450. https://doi.org/10.1038/s41467-021-21651-0
Hariri, R. H., Fredericks, E. M., & Bowers, K. M. (2019). Uncertainty in big data analytics: Survey, opportunities, and challenges. *Journal of Big Data*, *6*, 44. https://doi.org/10.1186/s40537-019-0206-3
Hengl, T., de Jesus, J. M., Heuvelink, G. B., Gonzalez, M. R., Kilibarda, M., Blagotić, A., et al. (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PLoS One*, *12*(2), e0169748. https://doi.org/10.1371/10.1371/journal.pone.0169748
Holtzman, N. M., Pavelsky, T. M., Cohen, J. S., Wrzesien, M. L., & Herman, J. D. (2020). Tailoring WRF and Noah-MP to improve process representation of Sierra Nevada runoff: Diagnostic evaluation and applications. *Journal of Advances in Modeling Earth Systems*, *12*(3), e2019MS001832. https://doi.org/10.1029/2019MS001832
Huscroft, J., Gleeson, T., Hartmann, J., & Börker, J. (2018). Compiling and mapping global permeability of the unconsolidated and consolidated Earth: GLobal HYdrogeology MaPS 2.0 (GLHYMPS 2.0). *Geophysical Research Letters*, *45*(4), 1897–1904. https://doi.org/10.1002/2017GL075860
Iqbal, M., Dahri, Z., Querner, E., Khan, A., & Hofstra, N. (2018). Impact of climate change on flood frequency and intensity in the Kabul River Basin. *Geosciences*, *8*(4), 114. https://doi.org/10.3390/geosciences8040114
Jensen, C. K., McGuire, K. J., & Prince, P. S. (2017). Headwater stream length dynamics across four physiographic provinces of the Appalachian Highlands. *Water Resources Research*, *31*(19), 3350–3363. https://doi.org/10.1002/hyp.11259
Jowett, I. G. (1997). Instream flow methods: A comparison of approaches. *Regulated Rivers: Research & Management*, *13*(2), 115–127. http://doi.org/10.1002/(SICI)1099-1646(199703)13:23.0.CO;2-6
Knoben, W. J., Freer, J. E., & Woods, R. A. (2019). Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores. *Hydrology and Earth System Sciences*, *23*(10), 4323–4331. https://doi.org/10.5194/hess-23-4323-2019
Kowalczyk, E. A., Stevens, L., Law, R. M., Dix, M., Wang, Y. P., Harman, I. N., et al. (2013). The land surface model component of ACCESS: Description and impact on the simulated surface climatology. *Australian Meteorological and Oceanographic Journal*, *63*(1), 65–82. http://doi.org/10.22499/2.6301.005
Krinner, G., Viovy, N., de Noblet-Ducoudré, N., Ogée, J., Polcher, J., Friedlingstein, P., et al. (2005). A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system. *Global Biogeochemical Cycles*, *19*(1). https://doi.org/10.1029/2003GB002199

Kuppel, S., Fan, Y., & Jobbagy, E. G. (2017). Seasonal hydrologic buffer on continents: Patterns, drivers and ecological benefits. *Advances in Water Resources*, *102*, 178–187. https://doi.org/10.1016/j.advwatres.2017.01.004

Lin, P., Pan, M., Wood, E. F., Yamazaki, D., & Allen, G. H. (2021). A new vector-based global river network dataset accounting for variable drainage density. *Scientific Data*, *8*, 28. https://doi.org/10.1038/s41597-021-00819-9

Mahmoud, S. H., & Gan, T. Y. (2018). Impact of anthropogenic climate change and human activities on environment and ecosystem services in arid regions. *The Science of the Total Environment*, *633*, 1329–1344. https://doi.org/10.1016/j.scitotenv.2018.03.290

Mendoza, G. F., Steenhuis, T. S., Walter, M. T., & Parlange, J.-Y. (2003). Estimating basin-wide hydraulic parameters of a semi-arid mountainous watershed by recession-flow analysis. *Journal of Hydrology*, *279*, 57–69. https://doi.org/10.1016/S0022-1694(03)00174-4

Miguez-Macho, G., & Fan, Y. (2012a). The role of groundwater in the Amazon water cycle: 1. Influence on seasonal streamflow, flooding and wetlands. *Journal of Geophysical Research*, *117*, D15113. https://doi.org/10.1029/2012JD017539

Miguez-Macho, G., & Fan, Y. (2012b). The role of groundwater in the Amazon water cycle: 2. Influence on seasonal soil moisture and evapotranspiration. *Journal of Geophysical Research*, *117*, D15114. https://doi.org/10.1029/2012JD017540

Milly, P. C., & Shmakin, A. B. (2002). Global modeling of land water and energy balances. Part I: The land dynamics (laD) model. *Journal of Hydrometeorology*, *3*, 283–299. https://doi.org/10.1175/1525-7541(2002)003<0283:GMOLWA>2.0.CO;2

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I—A discussion of principles. *Journal of Hydrology*, *10*(3), 282–290. https://doi.org/10.1016/0022-1694(70)90255-6

Niu, G. Y., Yang, Z. L., Dickinson, R. E., & Gulden, L. E. (2005). A simple TOPMODEL-based runoff parameterization (SIMTOP) for use in global climate models. *Journal of Geophysical Research: Atmospheres*, *110*(D21). https://doi.org/10.1029/2005JD006111

Niu, G. Y., Yang, Z. L., Dickinson, R. E., Gulden, L. E., & Su, H. (2007). Development of a simple groundwater model for use in climate models and evaluation with Gravity Recovery and Climate Experiment data. *Journal of Geophysical Research: Atmospheres*, *112*, D07103. https://doi.org/10.1029/2006JD007522

Niu, G. Y., Yang, Z. L., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., et al. (2011). The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. *Journal of Geophysical Research: Atmospheres*, *116*(D12). https://doi.org/10.1029/2010JD015139

NOAA. (2016). *National water model: Improving NOAA's water prediction services* (p. 2). Retrieved from http://water.noaa.gov/documents/wrn-national-water-model.pdf

Oleson, K. W., Lawrence, D. M., Bonan, G. B., Flanner, M. G., Kluzek, E., Lawrence, P. J., et al. (2010). *Technical description of version 4.0 of the Community Land Model (CLM)*. University Corporation for Atmospheric Research. Retrieved from https://opensky.ucar.edu/islandora/object/technotes:493

Oudin, L., Andréassian, V., Perrin, C., Michel, C., & Le Moine, N. (2008). Spatial proximity, physical similarity, regression and ungaged catchments: A comparison of regionalization approaches based on 913 French catchments. *Water Resources Research*, *44*(3). https://doi.org/10.1029/2007WR006240

Pechlivanidis, I. G., Jackson, B. M., Mcintyre, N. R., & Wheater, H. S. (2011). Catchment scale hydrological modelling: A review of model types, calibration approaches and uncertainty analysis methods in the context of recent developments in technology and applications. *Global NEST Journal*, *13*(3), 193–214. https://doi.org/10.30955/gnj.000778

Pokhrel, Y. N., Fan, Y., Miguez-Macho, G., Yeh, P. J. F., & Han, S. C. (2013). The role of groundwater in the Amazon water cycle: 3. Influence on terrestrial water storage computations and comparison with GRACE. *Journal of Geophysical Research: Atmospheres*, *118*, 3233–3244. https://doi.org/10.1002/jgrd.50335

Praskievicz, S., Luo, C., Bearden, B., & Ernest, A. (2018). Evaluation of low-flow metrics as environmental instream flow standards during long-term average and 2016 drought conditions: Tombigbee River Basin, Alabama and Mississippi, USA. *Water Policy*, *20*(6), 1240–1255. https://doi.org/10.2166/wp.2018.023

Rahman, M., Rosolem, R., Kollet, S. J., & Wagener, T. (2019). Towards a computationally efficient free-surface groundwater flow boundary condition for large-scale hydrological modelling. *Advances in Water Resources*, *123*, 225–233. https://doi.org/10.1016/j.advwatres.2018.11.015

R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. URL Retrieved from https://www.R-project.org/

Saft, M., Peel, M. C., Western, A. W., & Zhang, L. (2016). Predicting shifts in rainfall-runoff partitioning during multiyear drought: Roles of dry period and catchment characteristics. *Water Resources Research*, *52*(12), 9290–9305. https://doi.org/10.1002/2016WR019525

Sælthun, N. R. (1996). *The "Nordic" HBV model. Description and documentation of the model version developed for the project Climate Change and Energy Production*. Retrieved from https://www.osti.gov/etdeweb/servlets/purl/515690

Samuel, J., Coulibaly, P., & Metcalfe, R. A. (2011). Estimation of continuous streamflow in Ontario ungauged basins: Comparison of regionalization methods. *Journal of Hydrologic Engineering*, *16*(5), 447–459. http://doi.org/10.1061/(ASCE)HE.1943-5584.0000338

Schaake, J., Cong, S., & Duan, Q. (2006). *US MOPEX data set* (No. UCRL-JRNL-221228). Lawrence Livermore National Laboratory.

Seibert, J., & van Meerveld, H. J. (2016). Hydrological change modeling: Challenges and opportunities. *Hydrological Processes*, *30*(26), 4966–4971. https://doi.org/10.1002/hyp.10999

Shaw, S. B., & Riha, S. J. (2012). Examining individual recession events instead of a data cloud: Using a modified interpretation of dQ/dt–Q streamflow recession in glaciated watersheds to better inform models of low flow. *Journal of Hydrology*, *434*, 46–54. https://doi.org/10.1016/j.jhydrol.2012.02.034

Smakhtin, V. U. (2001). Low flow hydrology: A review. *Journal of Hydrology*, *240*(3–4), 147–186. https://doi.org/10.1016/S0022-1694(00)00340-1

Smirnov, N. (1948). Table for estimating the goodness of fit of empirical distributions. *The Annals of Mathematical Statistics*, *19*(2), 279–281. https://doi.org/10.1214/aoms/1177730256

Subin, Z., Milly, P., Sulman, B., Malyshev, S., & Shevliakova, E. (2014). Resolving terrestrial ecosystem processes along a subgrid topographic gradient for an Earth-system model. *Hydrology and Earth System Sciences Discussions*, *11*, 8443–8492. https://doi.org/10.5194/hessd-11-8443-2014

Swain, J. B., & Patra, K. C. (2017). Streamflow estimation in ungauged catchments using regionalization techniques. *Journal of Hydrology*, *5548*(19), 420–433. https://doi.org/10.1016/j.jhydrol.2017.08.054

Tague, C., & Band, L. (2004). Rhessys: Regional hydro-ecologic simulation system—An object-oriented approach to spatially distributed modeling of carbon, water, and nutrient cycling. *Earth Interactions*, *8*(19), 1–42. https://doi.org/10.1175/1087-3562(2004)8<1:RRHSSO>2.0.CO;2

Tague, C., & Grant, G. E. (2009). Groundwater dynamics mediate low-flow response to global warming in snow-dominated alpine regions. *Water Resources Research*, *45*(7). https://doi.org/10.1029/2008WR007179

Tashie, A., Pavelsky, T., Band, L. E., & Topp, S. (2021). Watershed-scale effective hydraulic properties of the continental United States. *Journal of Advances in Modeling Earth Systems*, *13*(6), e2020MS002440. https://doi.org/10.1029/2020MS002440

Tashie, A., Pavelsky, T., & Emanuel, R. E. (2020). Spatial and temporal patterns in baseflow recession in the continental United States. *Water Resources Research*, *56*(3), e2019WR026425. https://doi.org/10.1029/2019WR026425

Tashie, A., Scaife, C. I., & Band, L. E. (2019). Transpiration and subsurface controls of streamflow recession characteristics. *Hydrological Processes*, *33*(19), 2561–2575. http://doi.org/10.1002/hyp.13530

Tennant, D. L. (1976). Instream flow regimens for fish, wildlife, recreation and related environmental resources. *Fisheries*, *1*(4), 6–10. https://doi.org/10.1577/1548-8446(1976)001<0006:IFRFFW>2.0.CO;2

Toum, E. (2021). *HBV.IANIGLA: Modular hydrological model. 0.2.0*. Retrieved from https://cran.r-project.org/web/packages/HBV.IANIGLA/index.html

Troch, P. A., Berne, A., Bogaart, P., Harman, C., Hilberts, A. G., Lyon, S. W., et al. (2013). The importance of hydraulic groundwater theory in catchment hydrology: The legacy of Wilfried Brutsaert and Jean-Yves Parlange. *Water Resources Research*, *49*(9), 5099–5116. https://doi.org/10.1002/wrcr.20407

United States Geological Survey [USGS]. (2004). *National hydrography dataset [map]*. U.S. Department of the Interior, U.S. Geological Survey.

van den Hurk, B. J., Viterbo, P., Beljaars, A. C. M., & Betts, A. K. (2000). *Offline validation of the ERA40 surface scheme*. Retrieved from https://www.ecmwf.int/en/elibrary/12900-offline-validation-era40-surface-scheme

Wahl, K. L., & Wahl, T. L. (1995, August 16–17). Determining the flow of Comal Springs at New Braunfels, Texas. In *Proceedings of Texas Water '95* (pp. 77–86). American Society of Civil Engineers. Retrieved from http://www.usbr.gov/pmts/hydraulics_lab/twahl/bfi/texaswater95/comalsprings.html

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics*, *1*, 80–83. https://doi.org/10.2307/3001968

Wolock, D. M. (2003). *Flow characteristics at U.S. Geological survey streamgages in the conterminous United States* [open-file report 03-146, digital dataset]. U.S. Geological Survey. Retrieved from https://water.usgs.gov/lookup/getspatial?qsitesdd

Yang, Z.-L., Cai, X., Zhang, G., Tavakoly, A. A., Jin, Q., Meyer, L. H., & Guan, X. (2011). *The community Noah land surface model with multi-parameterization options (Noah-MP)*. The University of Texas at Austin.

Zambrano-Bigiarini, M. (2020). *hydroGOF: Goodness-of-fit functions for comparison of simulated and observed hydrological time series* [R package version 0.4-0]. Retrieved from https://github.com/hzambran/hydroGOF

Zecharias, Y. B., & Brutsaert, W. (1988). Recession characteristics of groundwater outflow and base flow from mountainous watersheds. *Water Resources Research*, *24*(10), 1651–1658. https://doi.org/10.1029/WR024i010p01651

Zhang, Y., & Schaap, M. G. (2019). Estimation of saturated hydraulic conductivity with pedotransfer functions: A review. *Journal of Hydrology*, *575*, 1011–1030. https://doi.org/10.1016/j.jhydrol.2019.05.058

Zimmer, M. A., & McGlynn, B. L. (2017a). Ephemeral and intermittent runoff generation processes in a low relief, highly weathered catchment. *Water Resources Research*, *53*(8), 7055–7077. https://doi.org/10.1002/2016WR019742

Zimmer, M. A., & McGlynn, B. L. (2017b). Time-lapse animation of hillslope groundwater dynamics details event-based and seasonal bidirectional stream–groundwater gradients. *Hydrological Processes*, *31*(10), 1983–1985. http://doi.org/10.1002/hyp.11124