Representing the Dupuit-Boussinesq Aquifer in the National Water Model: Catchment-Scale Application of Hydraulic Groundwater Theory

Minki Hong[1] and Binayak Mohanty[1]

[1]Texas A&M University

November 23, 2022

Abstract

While hydraulic groundwater theory has been understood as a viable approach for representing the role of the aquifer(s) in the surface-subsurface hydrologic cycle, the integrated modeling community still lacks a proper hydrologic structure to utilize the well-studied theory for large-scale hydrologic predictions. This study aims to present a novel hydrologic modeling framework that enables the Boussinesq equation-based depiction of hillslope-channel connectivity for applying hydraulic groundwater theory to large-scale model configurations. We integrated the BE3S's [Hong et al., 2020] representation scheme of the catchment-scale Boussinesq aquifer into the National Water Model (NWM) and applied the NWM-BE3S model to three major basins in Texas (i.e., the Trinity, Brazos, and Colorado River basins). Since the NWM currently relies on a single reservoir model for baseflow estimation, theory-based evaluation was performed as the efficacies that the Boussinesq aquifer has relative to the single reservoir model should be consistent with hydraulic groundwater theory. We identified that the implemented Boussinesq aquifer(s) showed 'more' pronouced improvements in capturing streamflow dynamics than the original NWM as aquifers exhibited higher nonlinearities in the observed recessions. The varying degree of improvements in streamflow outputs according to the recession nonlinearities demonstrates (1) the applicability of the theory-based depiction of hillslope-channel connectivity and (2) the technical enhancement of model structure. We also examined the river states of all the reaches based on the represented bidirectional lateral hydraulic connections between the stream-aquifer and thus identified the dominant processes between the stream-aquifer (i.e., either river infiltration or baseflow) were spatially variable roughly following climatic gradients.

Representing the Dupuit-Boussinesq Aquifer in the National Water Model: Catchment-Scale Application of Hydraulic Groundwater Theory

M. Hong[1], B. P. Mohanty[1]

[1]Texas A&M University, Department of Biological and Agricultural Engineering

Corresponding author: Binayak P. Mohanty (mailto:bmohanty@tamu.edu)

Key Points:

- Catchment-scale bidirectional lateral hydraulic connections between the stream-aquifer were newly represented in the National Water Model
- The Boussinesq aquifer yielded improved streamflow prediction than the single reservoir model as the nonlinearity of recessions increases
- The state of river reaches was evaluated based on bidirectional processes by the lateral hydraulic gradient between the stream-aquifer

Abstract

Although hydraulic groundwater theory has been recognized as a promising tool for understanding the role of the aquifer(s) in the surface-subsurface hydrologic cycle, the integrated modeling community still lacks a proper hydrologic structure to apply the well-studied theory to large-scale hydrologic predictions. This study aims to present a novel hydrologic structure that enables the Boussinesq equation-based depiction of hillslope-channel connectivity for applying hydraulic groundwater theory to large-scale model configurations. We integrated the BE3S's [*Hong et al.*, 2020] representation scheme of the catchment-scale Boussinesq aquifer into the National Water Model (NWM) and applied the NWM-BE3S model to three major basins in Texas (i.e., the Trinity, Brazos, and Colorado River basins). Since the NWM currently relies on a single reservoir model for baseflow simulation, theory-based evaluation was performed as the efficacies that the Boussinesq aquifer has relative to the single reservoir model should be consistent/explained with hydraulic groundwater theory. We identified the implemented Boussinesq aquifer(s) yielded 'more' pronounced improvements in predicting streamflow than the NWM's bucket model as aquifers exhibited higher nonlinearities in the observed recessions. The varying degree of improvements in streamflow outputs according to the recession nonlinearities demonstrates (1) the applicability of the theory-based depiction of hillslope-channel connectivity as well as (2) the technical enhancement of model structure. We also diagnosed the river states of all the reaches based on the represented bidirectional lateral hydraulic connections between the stream-aquifer and identified the dominant processes between the stream-aquifer (i.e., either river infiltration or baseflow) were spatially heterogenous roughly following climatic gradients.

1 Introduction

Streamflow forecasts have been increasingly gaining attention because of their potential uses for such as water resources management, reservoir operations, and flood-risk mitigation in the climate change era [*Alfieri et al.*, 2013; *Maurer*, 2005; *Zhao et al.*, 2011]. Future water demand expected to be growing (e.g., crop production) [*Rinaudo*, 2015] also calls for the improved streamflow predictability for sustainable use of water resources [*Barthel*, 2014]. Effective management of water resources should necessitate viable tools, beyond observational data, to assist decision making, and process-based understanding of dynamic hydrologic systems has been a fundamental approach for forecasting [*Baroni et al.*, 2019; *Jasper et al.*, 2002; *Maxwell et al.*, 2015; *Viterbo et al.*, 2020]. Particularly, comprehensive knowledge about large-scale water cycle/movement can be a basis of holistic communication between different dimensions (i.e., resources, human, and policy) for socio-economic development [*Savenije and Van der Zaag*, 2008].

Groundwater and river water are two hydraulically connected reservoirs [*Fleckenstein et al.*, 2010], and thus physically-based simulated streamflow cannot expect significant qualitative improvements without explicitly accounting for the role of the aquifer in interacting with the river [*Huntington and Niswonger*, 2012; *Karki et al.*, 2021; *Nijssen et al.*, 1997]. Over decades, a variety of efforts has been made to represent the interactive processes between the stream and the aquifer. The efforts range from building equation-based theoretical method [*Boussinesq*, 1904; *Hornberger and Remson*, 1970; *Lockington*, 1997; *Rorabaugh*, 1964; *Rupp and Selker*, 2005] to developing integrated hydrologic models with a particular focus on groundwater (GW) - surface water (SW) interactions [*Bisht et al.*, 2017; *Gochis*, 2018; *Kim et al.*, 2008; *S. J. Kollet and Maxwell*, 2006; *C P Shen and Phanikumar*, 2010; *Tesfa et al.*, 2014]. Although significant advances in understanding the effects of GW-SW interactions on streamflow generation have been achieved by the developed large-scale hydrologic models, particularly, the heterogeneity in properties and process complexity at different scales still remains as a significant challenge for an adequate description of exchange processes [*Gauthier et al.*, 2009; *Gómez-Hernández and Gorelick*, 1989; *Maxwell and Condon*, 2016; *McDonnell et al.*, 2007]. Even in (relatively) high-resolution (e.g., 1-km) integrated models, subsurface heterogeneities ignored below modelspecific resolution is one of the primary factors exacerbating the uncertainty in modeled data [*Baroni et al.*, 2019; *S Kollet et al.*, 2017; *Maxwell et al.*, 2014; *Tijerina et al.*, 2021].

To deal with the uncertainty associated with subsurface heterogeneity and dynamics, the concept of the equivalent parameter (i.e., effective parameter) in reproducing the effects of subsurface spatial variability has been of critical importance [*Berg and Illman*, 2011; *Binley et al.*, 1989; *Gauthier et al.*, 2009]. Unlike data-driven parameterization scheme used in several distributed hydrologic models [*Kim et al.*, 2008; *Lautz and Siegel*, 2006; *Maxwell et al.*, 2015; *C Shen et al.*, 2013], no need to characterize (all) relevant small-scale subsurface variations, which is infeasible due to technological constraints, exists with the working assumption that large-scale aquifer heterogeneity can be lumped into effective parameter values [*Binley et al.*, 1989; *McDonnell et al.*, 2007]. However, the scale-dependent nature of aquifer hydraulic properties (such as saturated hydraulic conductivity K_s and porosity f) requires flux variables observable at corresponding scale to test if such equivalent properties properly represent natural heterogeneous system [*Dewandel et al.*, 2017; *Fallico et al.*, 2016; *Gómez-Hernández and Gorelick*, 1989; *Shin et al.*, 2013; *Zhu and Mohanty*, 2003]. Effective parameterization schemes should thus have limited practical value if flux observations are not readily available at the scale where they pursue to decide aquifer properties for numerical predictions.

The importance of representing catchment-scale hydrologic processes has been studied in great detail, primarily based on the mathematical relationship between power-law streamflow recession model and the Boussinesq equation for aquifer outflow [*Brutsaert and Nieber*, 1977; *Clark et al.*, 2009; *Fan et al.*, 2019; *Troch et al.*, 2003; *Troch et al.*, 2013; *Wagener et al.*, 2007]. By providing the physically explicit method to explain observed hydrologic responses on catchment-scale (e.g., groundwater release from hillslope) with catchment (flux) observations (e.g., streamflow), the theory has enabled not only inferring dominant catchment processes but also effective parameterization of catchment-scale aquifer [*Brutsaert and Lopez*, 1998; *Vannier et al.*, 2014]. While the hillslope-channel exchange representation based on the hydraulic groundwater theory provides conceptual clarity and parametric efficiency [*Fan et al.*, 2019; *Troch et al.*, 2013] with (readily available) streamflow observations, integrated modeling community still lacks an appropriate hydrologic structure (i.e., physics) able to perform theory-based depiction of hillslope-channel connectivity (e.g., GW-SW interactions). Furthermore, the

GW-SW interactions described based on (variably saturated) 2D/3D Richards' equation is assumed to occur in the vertical direction only [*Bisht et al.*, 2017; *Kim et al.*, 2008; *S. J. Kollet and Maxwell*, 2006; *Seo et al.*, 2007; *Wu et al.*, 2021], which should be considered a structural limitation that cannot represent the actual GW-SW interactions driven by lateral hydraulic gradients in real hydrologic systems [*Basha*, 2013; *Boussinesq*, 1904; *Hornberger et al.*, 1970; *Liang et al.*, 2018; *Paniconi et al.*, 2003; *Rupp and Selker*, 2006].

Therefore, this study aims to represent the catchment-scale hillslope-channel connectivity (i.e., bidirectional interactions) based on the Dupuit-Boussinesq assumption, hereafter referred to as the Boussienq aquifer, in an integrated framework. As the theoretical basis of hydraulic groundwater theory, the Boussinesq aquifer can describe the hillslope storage and release dynamics as the response to streamflow variations (i.e., fluctuations) based on catchment-scale lateral hydraulic gradients [*Hong et al.*, 2020; *Hornberger et al.*, 1970; *Lockington*, 1997; *Rupp and Selker*, 2006]. To understand the effects of the implemented Boussinesq aquifer with regard to parametric efficiency and prediction accuracy, we specifically selected the National Water Model (NWM) (i.e., the WRF-Hydro NWM configuration) and attempted to integrate (i.e., full coupling) the representation scheme of the catchment-scale Boussinesq aquifer presented in the recent Bidirectional Exchange Scheme in Surface and Subsurface (BE3S) [*Hong et al.*, 2020] into the NWM.

The National Water Model (NWM), as the Next Generation Water Resources Modeling Framework (Nextgen), has been operating over the conterminous United States (CONUS) since 2016. Streamflow is one of the primary variables forecasted at various time intervals (i.e., 18 hr (short), 10 d (medium), and 30 d (long)), and thus the importance of proper representation of aquifer system is pronounced. Currently, the aquifer-channel connectivity in the NWM relies on a conceptual (i.e., not physically-explicit) storage (S) – discharge (Q) reservoir model, which yields baseflow fluxes (Q) as the function of groundwater storage (S) [*Gochis*, 2018]. While the conceptual single reservoir model provides computational efficiency, the (almost) linear behavior between S and Q (of the catchment-scale reservoir(s)) in the NWM [*Karki et al.*, 2021] hampers the predictive capability of the NWM for streamflow with distinct temporal dynamics (i.e., high nonlinearity in streamflow recessions). Also, since no hydrologic structure exists to support physically explicit parameterization of aquifer properties, the fitting parameters (to define the exponential relationship between S – Q) have to be empirically derived, and the lateral hydraulic connections between the stream and the aquifer were inevitably ignored [*Gochis*, 2018].

We presented the coupled NWM-BE3S integrated framework to complement the addressed structural limitations in the (current) single reservoir baseflow module in the NWM. To comparatively understand the effects of (different) physics configurations of the aquifer system in large-scale hydrologic modeling, we applied both the NWM-BE3S and the (original) NWM to three major basins in Texas (i.e., Trinity River, Brazos River, and Colorado River basins). Streamflow was used as a primary comparison variable considering the availability of observations and its implication as the main result of aquifer processes in the (hydraulically-connected) catchment system. As the working hypothesis ((1), (2), and (3)), the effects of the Boussinesq equation-based depiction of hillslope-channel connectivity in the NWM-BE3S coupled model (compared to the original NWM) are evaluated based on the following criteria:

(1) The Boussinesq aquifer enables the utilization of effective aquifer parameters (from streamflow observations) to depict catchment-scale hillslope-channel connectivity.

(2) The baseflow fluxes derived from the Boussinesq aquifer should capture a broader range of streamflow recession characteristics than the single reservoir model.

(3) The river states (i.e., gaining/losing) are temporally dynamic and can be diagnosed based on the lateral hydraulic gradients per the stream-aquifer head difference.

2 Methods

2.1 Algorithmic Description of the BE3S

The BE3S couples 1-dimensional governing equations of Richards' equation, Boussinesq equation, and Saint-Venant equation to represent flow processes in the vadose zone, the phreatic aquifer, and the open channel (i.e., river reach), respectively. A complete coupled surface-subsurface flow system should include surface and subsurface hydrologic components, interfacial/external boundary conditions, and initial conditions in the forward modeling framework [*Furman*, 2008]. Time-dependent hydrologic states in each flow domain (e.g., groundwater depth and river stage) were explicitly used in establishing interfacial boundary conditions between adjacent domains. The BE3S connects two types of interfaces (i.e., the interface between vadose zone-aquifer and aquifer-river) and simulates the potential-driven bidirectional exchanges at the interfaces through (hydraulic) head-based boundary conditions [*Hong et al.*, 2020]. Since the BE3S handles multiple processes that involve different temporal scales ranging from an hour to years, besides, care needs to be taken when defining initial and boundary conditions at *j+1* th time step as the result of various hydrologic states and fluxes at *j* th time step. Details about the temporal coupling at an hourly time step, the temporal scale in this study, are given in Supplemental Information SI1.

2.1.1 Boussinesq Equation-Based Aquifer System Representation

The BE3S solves the non-linear Boussinesq equation, derived from the Dupuit-Forcheimer assumption, to represent catchment outflow (i.e., baseflow) as the phreatic aquifer's response to a drawdown of river stage [*Basha*, 2013; *Hornberger and Remson*, 1970; *Hornberger et al.*, 1970; *Lockington*, 1997]. The non-linear form of the Boussinesq equation was applied in a direction perpendicular to the flow direction of the river and incorporates the time-dependent river stage as the time-varying boundary condition at the discharge boundary (i.e., the interface between the river and the phreatic aquifer) (Figure 1 a). The outflow fluxes from the Boussinesq aquifer, therefore, were modeled based on the (time-varying) lateral hydraulic gradients between the river stage and adjacent groundwater level (GWL). The hybrid discretization scheme (i.e., finite volume and difference) in BE3S also enables an efficient node configuration for solving the head-based Boussinesq equation applied with river stage boundary conditions (Figure 1 d) (Equation 1, and 2).

$$f \times \left[\frac{h_i^{j+1} - h_i^j}{\Delta t}\right] = \frac{K_s}{\Delta x}\left[h_i^j\left(\frac{h_{i+1}^j - h_i^j}{\Delta x}\right) - h_{i-1}^j\left(\frac{h_i^j - h_r^j}{\Delta x}\right)\right] \qquad (i = 1) \qquad (1)$$

$$f \times \left[\frac{h_i^{j+1} - h_i^j}{\Delta t}\right] = \frac{K_s}{\Delta x}\left[h_i^j\left(\frac{h_{i+1}^j - h_i^j}{\Delta x}\right) - h_{i-1}^j\left(\frac{h_i^j - h_{i-1}^j}{\Delta x}\right)\right] \qquad (i = 2, \ldots, n_p) \qquad (2)$$

Where h_i^j is the hydraulic head of GWL at *j* th time step on *i* th node (L). h_r^j is the river stage at *j* th time step (L). *f* is effective porosity (-), and K_s is horizontal hydraulic conductivity (LT^{-1}). Δx is the size of a grid cell in the phreatic aquifer (L). n_p is the minimum number of aquifer grids from the river reach to the farthest aquifer grid, determined by the corresponding catchment's size. The BE3S incorporates the effects of the unsaturated soils on the groundwater storage in setting up the initial conditions of Equation 1 and 2 (Figure 1 c), through the interfacial boundary equation between the vadose zone and the phreatic aquifer (Equation 3).

$$D_i^{j+1} = h_i^j - \left(\frac{f_{vp}^{j+1} \times f}{A_v}\right) \quad (3)$$

Where D_i^{j+1} is the temporary groundwater level hydraulic head used to setup the initial condition (i.e., the horizontal profile of saturated aquifer thickness) (L). Again, h_i^j is the hydraulic head of groundwater level at j th time step on i th node from the river (L), and f is the effective porosity of the phreatic aquifer. A_v is the area of an aquifer grid cell (L^2). f_{vp}^{j+1} denotes the net exchange fluxes between the unsaturated zone and the phreatic aquifer during j+1 th time step. The value distributions of D were developed based on the distance from the river reach (Figure 1 b). For example, D_5^j is a set of D values at five grids away from the river at j th time step. Then, the expectation of D_i, denoted as $E[D_i]$, was computed using the probabilistic plot of each D_i (i=1,2,…n_p). The estimated $E[D_i]$ (i=1,2,…n_p) was used to set up the initial condition of the Boussinesq equation (Figure 1). The boundary forcing at the discharge interface h_r^j was decided by averaging the river stage profile from the most upstream node to the downstream node at the corresponding time step. The horizontal profile of groundwater level heads h_i^j

(i=1,2,…n_p) from the discharge interface to the edge of the catchment on *j* th time step was then finalized through Equation 1, and 2.

2.1.2. Interactions Based on Lateral Hydraulic Connections Between the River and the Aquifer

The amount of exchange fluxes between the hillslope and the river channel f_{pr} were computed based on the lateral hydraulic gradients, resulting from the (temporally dynamic) head differences between river stage and GWL (Equation 4).

$$f_{pr}^{j} = \begin{Bmatrix} K_s \times \left[\left(W_p^j \right) \times \left(\frac{h_r^j - h_{Mr}^j}{M_r} \right) \right] \times n_r \Delta t & \left(h_{M_r}^j > M_r \right) \\ K_s \times \left[\left(W_p^j \right) \times \left(\frac{h_r^j - M_r}{M_r} \right) \right] \times n_r \Delta t & \left(h_{M_r}^j \leq M_r \right) \end{Bmatrix} \tag{4}$$

Where f_{pr}^{j} is the cumulative exchange fluxes at *j* th time step (L^3), M_r is the thickness of river bottom sediment (L), $h_{M_r}^j$ is the hydraulic head of GWL at the distance of M_r from a river channel at *j* th time step (L). W_p^j is the bottom width of the river (L), which varies with the time-dependent river stage h_r^j. n_r is the number of the river grid cells from the inlet to the outlet of the catchment. For simplicity, the thickness of river bottom sediment M_r was considered equal to the size of one grid cell (i.e., 50 m) in the groundwater domain. The mass balance in groundwater storage at *j* th time step S_{gw}^j was defined by $S_{gw}^j = \sum_{i=1}^{n_p} A_p \times h_i^j$, where A_p is the area of an aquifer grid. The mass balance in the groundwater storage is estimated by Equation 5, and 6.

$$\epsilon_{gw}^{j+1} = \left| S_{gw}^{j+1} - \left(S_{gw}^{j} + \frac{dS_{gw}}{dt} \right) \right| \tag{5}$$

$$\left. \frac{dS_{gw}}{dt} \right|_{t=j} = \sum_{j=1}^{n_p} \nabla f_{vp,j}^{j} - \nabla f_{pr}^{j} \tag{6}$$

Where ϵ_{gw}^{j} denotes the mass balance in groundwater storage. $\left. \frac{dS_{gw}}{dt} \right|_{t=j}$ is considered as the result of net exchange fluxes between the vadose zone-the aquifer ∇f_{vp}^{j} and the aquifer-the river ∇f_{pr}^{j}. Errors in aquifer mass balance while simulating f_{pr} are kept below 0.01 % of the total volume of groundwater storage.

2.2 The National Water Model (NWM) Configuration

The core of the National Water Model (NWM) system is the National Center for Atmospheric Research (NCAR)-supported Weather Research and Forecasting Hydrologic (WRF-Hydro) model [*Gochis*, 2018]. The WRF-Hydro NWM configuration was developed to model land surface processes (e.g., water/energy balance) with the Noah-MP (Multi-Parameterization) model, utilizing NLDAS-2 hourly forcing and 1-km NRCS State Soil Geographic (STATSGO2) data for soil parameterization [*Salas et al.*, 2018; *Schwarz et al.*, 2018]. The NWM integrates separate routing options for representing subsurface flow for exfiltration calculation, overland diffusive flow, conceptual bucket baseflow, and open channel flow (i.e., Muskingum-Cunge routing). The NWM also provides capabilities to simulate (relatively simple) lake and reservoir surface routing, albeit not activated in this study. Here, we selectively focus on describing how the connectivity between the open channel and the aquifer is described using the conceptual storage-discharge bucket model to comparatively understand its physics differences against the Boussinesq aquifer flow assumption (presented in the BE3S).

2.2.1 Conceptual Storage-Discharge Bucket Model in the NWM

Each river reach has an associated storage-discharge bucket model. The groundwater discharge (i.e., baseflow), which contributes to the total streamflow, was calculated through the exponential functionality between the groundwater storage and discharge (Equation 7). Hereinafter the NWM's bucket model is referred to as Non-linear Single Reservoir, NLSR.

$$B_f^j = C \times (e^{exp \times (\frac{z^j}{z_{max}})} - 1) \tag{7}$$

Where B_f^j is the baseflow fluxes at *j* th time step (L^3T^{-1}), *C* (L^3T^{-1}) and *exp* (dimensionless) are calibration 'fitting' parameters. ${z^j}/{z_{max}}$ is the relative groundwater height as z^j is the groundwater height in the bucket, and z_{max} is the maximum bucket height, so that each NLSR has a specified volumetric capacity. The bottom drainage fluxes from the soil columns, included in the corresponding catchment, are aggregated at each time step (hourly) and accounted for as NLSR inflow (i.e., groundwater recharge) to determine time-dependent z^j (L).

As currently implemented, the NWM does not allow for losses from the channel because the baseflow is just a function of groundwater storage (i.e., NLSR height, *z*). The river reaches are thus always considered gaining reaches due to the structure allowing 'only' one-way exchanges from the NLSR to the river reach. This simplified baseflow estimation is also not a physically explicit representation of the aquifer system, so the fitting parameters must be introduced (i.e., *C* and *exp* in Equation 7) and empirically derived (for all the catchments). More importantly, as reported in other studies [*Karki et al.*, 2021], two things should be noted about the characteristics of the modeled baseflow in the NWM: (a) the lack of groundwater storage is found in most NLSR(s), meaning the total amount of NLSR inflow (i.e., soil bottom drainage) is almost identical to that of NLSR outflow (i.e., baseflow), (b) almost no time lag (mostly less than 1 hours) between the aquifer inflow and outflow is found. The (almost) linear relationship between the river discharge Q and groundwater storage S (i.e., $Q \cong S$) thus could be concluded from the current NLSR configuration as the result of the above two factors (a) and (b).

However, the baseflow estimation that relies on the relationship $Q \cong S$ inevitably should have a structural constraint that cannot depict the streamflow recessions showing high(er) nonlinearity (e.g., recession slope $b > 1.0$) [*Clark et al.*, 2009; *D Dralle et al.*, 2015; *Rupp and Selker*, 2006]. On the other hand, while the most recent NWM version 2.0 has implemented a river water infiltration scheme [*Lahmers et al.*, 2019], deactivating the channel percolation process was considered to meet the aim of our study. This is because: (1) only vertical channel infiltration was accounted for under the assumption that the stream and the aquifer are disconnected (e.g., semi-arid/arid), (2) lateral hydraulic connections (i.e., GW-SW) were still ignored so physically-explicit aquifer parameterization is not feasible, and river water infiltration due to river stage rising (e.g., storm events) cannot be represented [*Liang et al.*, 2018].

2.3 Representing the Dupuit-Boussinesq Aquifer System in the NWM Configuration

As an alternative representation of the aquifer system to the NLSR, the Boussinesq formulation is integrated to the NWM configuration. The coupled NWM-BE3S model was developed by implementing the parsimonious scheme for the Boussinesq aquifer representation in the BE3S into the framework of the WRF-Hydro configuration for the NWM. The represented Boussinesq aquifer(s) in the coupled NWM-BE3S model uses the soil bottom drainage modeled by the Noah-MP LSM and simulates catchment-scale lateral GWL profiles. Unlike the NLSR baseflow module, the river stage (simulated by the channel routing module) was explicitly incorporated into the (newly implemented) Boussinesq aquifer module to force the discharge boundary, and the exchange fluxes (between the stream-aquifer) are calculated bidirectionally according to lateral hydraulic gradients between the reach and adjacent aquifer (i.e., riparian zone). Since the framework of the Noah-MP LSM, as coupled with other routing schemes, was preserved in the NWM-BE3S model, both the NWM and the NWM-BE3S run on the same meteorological forcing (e.g., NLDAS-2) and soil/terrain routing parameters. The feature of the NHDPlus WBD that one reach corresponds to one catchment (i.e., connectivity from catchment to river reach) provided an appropriate structure for setting and solving the distance-based Boussinesq equation (Equation 1 and 2) in each catchment.

The coupled NWM-BE3S generated the bidirectional exchange fluxes (i.e., positive f_{pr} – baseflow, negative f_{pr} – river water infiltration) as the result of lateral hydraulic gradients (between the river and the riparian zone) on catchment-scale. The f_{pr}, along with the modeled overland flow, was ingested in the channel routing module, and the streamflow fluxes were predicted following the reach-based NHDPlus ver2.0 channel network. By incorporating physically-based aquifer representation, which stems from the Dupuit-Forcheimer assumption, importantly, we established the theoretical basis for effective parameterization of the catchment-scale aquifer (i.e., hydraulic groundwater theory) in the NWM-BE3S model. The effective aquifer properties K_s and f, explicitly used in calculating f_{pr} (Equation 4), can be inferred from the analysis of observed streamflow recessions [*Brutsaert and Nieber*, 1977; *Brutsaert and Lopez*, 1998; *Rupp and Selker*, 2006; *Troch et al.*, 1993; *Troch et al.*, 2013]. Furthermore, since the framework between the Boussinesq equation and power-law model [*Brutsaert and Nieber*, 1977; *Szilagyi et al.*, 1998; *Tallaksen*, 1995] provides the theoretical basis of various nonlinearity in the recessions (compared to the single (non-)linear reservoirs), the NWM-BE3S also provides a hydrologic structure for depicting streamflow with various recession nonlinearities. The differences in model physics configuration for the aquifer representation were summed up and presented in Table 1.

2.3 Streamflow Recession Analysis for Effective Parameterization of Catchment-Scale Aquifer

The mathematical relationship between the Boussinesq aquifer outflow and power-law model for streamflow recession (i.e., hydraulic groundwater theory) provides a unique physically explicit method, to date, for inferring catchment-scale effective aquifer properties based on average baseflow characteristics. In this section, we presented one approach to recession analysis for effective parameterization using daily streamflow observations, including 1) the recession extraction, 2) the recession parameter fitting, and 3) selected analytical solutions for each time domain (i.e., early and late time).

2.3.1 Recession Extraction Method

The criteria for identifying individual recession event (RE) include: 1) the onset of an individual recession event (RE) was defined as one day (24 hr) after the streamflow peak, following other studies, to exclude the effects of storm-related flow (e.g., overland flow/quick subsurface flow) on the streamflow [*Biswal and Marani*, 2010; *Shaw and Riha*, 2012], 2) each RE ends when the daily discharge is at the lowest based on the consecutive decline of discharge data (i.e., d*Q*/dt < 0, t=1d), and 3) the recession should last more than five days [*Biswal and Marani*, 2010; *Jachens et al.*, 2020; *Shaw and Riha*, 2012].

2.3.2 Estimation of Recession Parameters

Two recession parameters in the power-law model (Equation 8), intercept parameter *a* and slope parameter *b*, were estimated based on the aggregation of all observed recession data (i.e., point cloud).

$$\frac{dQ}{dt} = -aQ^b \tag{8}$$

Where *Q* is streamflow (L^3T^{-1}), and *t* is 1 day. While individual recession analysis could outperform in analyzing the variability in catchments' response to storm events with different magnitudes [*Jachens et al.*, 2020; *Karlsen et al.*, 2019; *Shaw and Riha*, 2012; *Szilagyi et al.*, 1998; *Tashie et al.*, 2020], we understood the point cloud data is still the only physically-explicit approach to determine 'average' characteristics of aquifer outflow. As presented by *Brutsaert and Nieber* [1977], we considered the lower envelope (LE) of the point cloud recession data under the assumption that small values of dQ/dt for a given Q represent the Boussinesq aquifer outflow [*Brutsaert and Lopez*, 1998; *Troch et al.*, 1993; *Vannier et al.*, 2014]. Thus, the recession parameter b was fixed to 3.0 and 1.5 for early time and late time domain, respectively [*Brutsaert and Nieber*, 1977], and a was determined such that 5 % of points were below the lower envelope [*Troch et al.*, 1993; *Wang*, 2011]. An alternative fitting method, wherein slope parameter b was decided as the best-fitted line to the point cloud, to understand the central tendency (CT) of the recession data (while addressing the undue weight of extreme data point) was also utilized [*Vogel and Kroll*, 1992]. The two fitting methods LE and CT were used to determine the catchment-scale aquifer properties (e.g., K_s and f) (LE method), and to represent the average catchments' response to storm events during the corresponding period (CT method).

2.3.3 Catchment-Scale Aquifer Parameterization

We selected analytical solutions for the respective early (i.e., high flow) and late time domain (i.e., low flow) to determine the catchment-scale aquifer properties under given recession parameters a and b (estimated through the LE method). The selective use of the recession parameters from the LE method reflects our effort to exclusively account for the low-flow conditions to infer the aquifer properties. The selected analytical solutions for early time [*Polubarinova-Koch*, 2015] (Equation 9) and late time domain [*Boussinesq*, 1904] (Equation 10) are described below.

$$\frac{dQ}{dt} = \frac{1.133}{K_s f {D_{ini}}^3 L^2} Q^3 \tag{9}$$

$$\frac{dQ}{dt} = \frac{4.804 {K_s}^{1/2} L}{f \alpha^{3/2}} Q^{3/2} \tag{10}$$

Where K_s is the horizontal saturated hydraulic conductivity (LT^{-1}), f is the catchment-scale effective porosity (-), D_{ini} is the initial saturated aquifer thickness (L), L is the channel length (L). α is the size of (effectively) contributing aquifer during the recessions. Since the size of contributing aquifer α and the initial saturated aquifer thickness D_{ini} are the two factors that affect the diffusivity between the stream and the aquifer, the range of α and D_{ini} should be adequately determined for realistic estimates of catchment-scale K_s and f. Specifically, the diffusivity K_s/f increases (non-linearly) with the increasing size of contributing aquifer and decreases (non-linearly) with increasing D_{ini} conditions. Thus, the upper and lower bound of D_{ini} and α were set up to meet these two criteria: 1) the effective porosity f should range from 0.1 % to 20.0 %, 2) the catchment-scale effective horizontal K_s should be less than 0.01 ms^{-1}. Once the range of K_s and f values was determined for each of the 40 catchments, the geometric mean of the respective range of K_s and f was calculated and considered as the representative value of effective K_s and f for the corresponding catchment.

2.4 Comparison Domain

The suitability of the newly represented Boussinesq formulation was evaluated based on the comparison between the respective streamflow outputs from the retrospective run of the NWM and the NWM-BE3S model. Both models were run for two years from 1/1/2016 – 12/31/2017, while the first year 2016 (365 d) was considered model spinning, and the modeled streamflow during the year 2017 (365 d) was used for the evaluation. Both models used the same hourly NLDAS-2 historic meteorological forcing and produced modeled outputs at hourly temporal resolution. Two statistical metrics of Pearson's correlation coefficient R and Root Mean Square Error ($RMSE$) were used to evaluate the temporal agreement of the respective modeled streamflow against corresponding observations (R) and the amount errors ($RMSE$).

3 Study Area and Data Description

3.1 Study Area

Three major basins in Texas, the Trinity River, the Brazos River, and the Colorado River basins, are selected as the study areas of this study (Figure 2). The combined area of the three basins accounts for 37.3 % of the entire Texas area, and the basins are essential sources of water for most major cities, meandering southeast. The drainage areas are 40,380 km^2 (1,140 km long), 116,000 km^2 (1,352 km long), and 103,000 km^2 (1,378 km long) in the Trinity River, Brazos River, and Colorado River Basin, respectively. According to the climate classification map, the northwestern regions of the Colorado River and the Brazos River belong to semi-arid climates. In contrast, the rest of the two Basins (i.e., southeastern) and the entire Trinity River Basin belong to the Humid Subtropical climate [*Kottek et al.*, 2006]. Lower annual precipitation is observed in areas closer to headwater (from the basin outlet). The average annual precipitation (over the last 30-year) in the Trinity River approximately ranges from 813 mm (i.e., headwater, -

98°53'31¨, 33°37'25¨) to 1,494 mm (i.e., outlet -98°31'58¨, 33°21'43¨), from 454 mm (i.e., headwater, -103°22'40¨, 34°28'30¨) to 1,338 (i.e., outlet -95°25'12¨, 28°59'38¨) in the Brazos River, and from 408 mm (i.e., headwater, -103°28'22¨, 33°17'27¨) to 1,106 mm (i.e., outlet -95°59'10¨, 28°43'05¨) in the Colorado River basin. Consistent with the precipitation gradient, which becomes more humid in areas from headwater to the outlet in each basin, the groundwater depths in the northwestern part of the study areas are deeper than 30 m. In contrast, the relatively shallower groundwater depths, ranging from 15 m to 3 m, are observed in the southeastern parts of the study areas.

3.1.1 Catchment Delineation

The study area consists of 73,436 catchments, delineated by the NHDPlus (Ver2.0) Watershed Boundary Dataset (WBD), following the current WRF-Hydro NWM configuration. As an integrated suite of application-ready geospatial datasets that incorporates many of the features of the National Hydrography Dataset (NHD) and the National Elevation Dataset (NED), the NHDPlus WBD includes a stream network based on the medium resolution 1:100,000 scale and elevation-derived catchments to enforce hydrologic divides (i.e., catchment drainage area). In this work, the groundwater divides were assumed to follow the boundaries of catchments since groundwater divides are likely to coincide with regional topographic highs [*Anderson et al.*, 2015]. The catchment boundaries were thus used to delineate the groundwater divides, which is assumed to be zero-flux BC (i.e., no water crosses a groundwater divide line).

3.1.2 Meteorological Forcing

For the two-year 2016 to 2017 retrospective simulation, the North American Land Data Assimilation System (NLDAS)-2 historical meteorological forcing data were applied to the original NWM and the coupled NWM-BE3S. As the aim of this study is to selectively evaluate the improvements in predictive performance by the structural change in the conceptual aquifer system, we applied the same NWM forcing to both modeling configurations (i.e., the NWM and the NWM-BE3S models). The NLDAS-2 were in 1/8th-degree grid spacing and range from 1/1/1979 to present on an hourly basis. To match the Noah-MP land surface model's spatial resolution, the NLDAS-2 forcing data were spatially downscaled (re-gridded) at 1-km using bilinear interpolation for the meteorological variables (e.g., precipitation rate, wind speed, temperature, and long/shortwave radiations) and used to force the upper boundary conditions of each 1-km land grid.

3.1.3 Soil Properties

The United States Geological Survey (USGS) developed the spatial dataset that represents soil texture attributes, as processed from STATSGO2 database, complied for the spatial component of the NHDPlus v2.0 data for the conterminous United States (CONUS) [*Schwarz et al.*, 2018; *Wieczorek and LaMotte*, 2010]. Sourcing the STATSGO2 soil data, the soil properties (e.g., permeability, percent soils, and bulk density) were estimated/provided as the minimum/maximum/average three values for the individual catchment. We considered that 1) the catchment-scale effective aquifer properties (e.g., (horizontal) saturated hydraulic conductivity) could have a significant relationship with catchment-average soil properties (Table 2), and thus 2) the catchment-scale aquifer properties in ungauged catchments could be predicted from the catchment-average soil attribute data.

3.2 Streamflow Observational Data

We used the observed daily averaged streamflow (L^3T^{-1}) from 40 USGS gauges, 15 gauges, 14 gauges, and 11 gauges in the Trinity River Basin, the Brazos River Basin, and the Colorado River Basin (Figure 2), respectively. The daily streamflow observations during the simulation period were used 1) to understand the seasonal/annual recession characteristics of the corresponding catchment (i.e., aquifer), especially the nonlinearity of recession curves, in each catchment, 2) to infer the catchment-scale effective aquifer properties K_s, f, D_{ini}, and α, and 3) to evaluate the improved predictive performance of the NWM-BE3S coupled model for streamflow (compared to the original NWM configuration). The major physical characteristics (necessitated for the recession analysis/effective parameterization) of the selected 40 catchments were listed up in Table 2.

4 Results and Discussion

4.1 Effective Parameterization of Catchment-Scale Aquifers

4.1.1 Recession Characteristics

As presented in Figure 3, the point cloud recession characteristics were investigated based on the identified transition of the hydraulic regimes (i.e., early time to late time domain) for the selected 40 USGS catchments. Parameter a and b were estimated to infer the decline rates (i.e., intercept log(a)) and the nonlinearity (i.e., slope b) of the recessions of the catchments. As addressed in several studies, we found that the different fitting methods resulted in distinct values of log(a) and b when applied to the daily streamflow observations [*D Dralle et al.*, 2015; *D N Dralle et al.*, 2017; *Jachens et al.*, 2020; *Stoelzle et al.*, 2013]. While the values of log(a) estimated from the CT method (i.e., CT, log(a) mean = -1.96) were generally found higher than those from the LE method (i.e., LE, log(a) mean = -2.56), the values of log(a) from the respective CT and LE methods are reasonably consistent with each other (i.e., R = 0.61). Also, the comparison of the estimated log(a) (from the two fitting methods) against the catchment-average permeability data revealed a significant linear relationship between them (i.e., log(a) from the CT method: R = 0.33, log(a) from the LE method: R = 0.46) as higher decline rates are expected in the catchments with higher permeability.

4.1.2 Effective Aquifer Parameterization for Catchment-Scale Aquifer

We determined the effective aquifer properties K_s and f for the 40 catchments by separating the hydraulic regimes from the early time to late time domain. Since the study areas (i.e., the three major basins) consist of 73,436 catchments (delineated by the NHDPlus v2.0 WBD dataset) and the Boussinesq aquifer flow is implemented for each catchment, the effective parameters K_s and f must be determined to represent distinct diffusivity conditions (between the stream and the aquifer) in each catchment. We identified a significant linear relationship between catchment-average permeability (LT^{-1}) and catchment-scale effective K_s (i.e., R = 0.56). Figure 4 a also shows that the catchment-scale effective K_s are well included in the 95 % band when predicted with catchment-average permeability (LT^{-1}), providing an empirical basis to determine the effective K_s in the ungauged catchments (i.e., 73,436 (total) – 40 (gauged) = 73,396 ungauged catchments). Since no significant ((non)-linear) relationship was found between catchment-average soil properties and effective f across the 40 studied catchments, moreover, we tried to to identify the probability distribution(s) of the diffusivities (i.e., Ks/f) in the 40 catchments to examine the value distribution pattern. Figure 4 b shows that the value distribution of diffusivity K_s/f are more properly represented by the log-normal (μ = -2.25, σ = 1.22) distribution than normal (μ = 0.19, σ = 0.23), exponential (scale = 0.19), and gamma (shape = 0.92, scale = 0.21) distributions. The arithmetic mean of the log-normally distributed diffusivity K_s/f was then considered as the representative diffusivity condition of the study area, yielding the diffusivity value of 0.022 ms^{-1}. The effective porosity f of each catchment was then determined according to the given diffusivity of 0.022 ms^{-1} and the effective K_s of the corresponding catchment. The catchment-scale effective aquifer properties including K_s, f, D_{ini}, and the size of contributing aquifer (determined following the presented effective parameterization scheme (section 2.3.3)) as well as the recession parameters a and b for each time domain are presented in Table 3.

4.2 Comparative Evaluation of Baseflow Estimates

4.2.1 Baseflow Module in the Original NWM and Its Structural Limitations

Before elucidating the similarities and differences between the respective baseflow estimates from the NWM and the coupled NWM-BE3S model, we first examined how the (one-way) baseflow fluxes were simulated under the original NWM configuration. It turned out that there was almost no time lag in the water entering (i.e., recharge) and leaving the NLSR, and almost all the groundwater recharge (into the NLSR) were discharged to the corresponding reach in the same time step (Figure 5 a, b, and c). The regression between the yearly cumulative NWM groundwater recharge (L^3) and yearly cumulative NWM baseflow (L^3) (during the simulation period) with the best-fit line yielded $y = 1.02x - 0.11$ ($R^2 = 0.998$). Considering the lack of groundwater storage (S) in the NLSR(s) (indicated by the small depths in the NLSR(s)), we could conclude that the NLSR simulates the (one-way) baseflow fluxes almost as a single linear reservoir while yielding the recession slope parameter b close to 1.0. In other words, the storage-discharge relationships in the original NWM physics configuration can be described as Q (discharge) $\cong S$ (storage) (consistent with previous studies), and thus the linear regression in the dQ/dt – Q bi-logarithmic space should yield the slope of 1.0. Consequently, the original NWM might have good predictive performance for recessions exhibiting linearity but also could show low performance for recessions with high nonlinearity (i.e., steeper/fast streamflow recessions) due to the structural limitations in its baseflow module.

4.2.2 Comparisons of the Respective Stream-Aquifer Exchange Fluxes According to Recession Slope Characteristics

Based on the understood behavioral characteristics of the baseflow outputs from the NWM, we compared the respective baseflow estimates from the NWM (i.e., B_f) and the NWM-BE3S model (i.e., f_{pr}) 1-year retrospective run (2017). This comparison was made for the 40 studied catchments that show distinct recession characteristics (i.e., from linear to highly non-linear) to investigate the similarities and differences between the respective baseflow estimates according to recession characteristics. The temporal agreements between the respective baseflow estimates (during the simulation period) were found higher as the recession slope b of the corresponding basin is closer to 1.0. That is, the Pearson's R value distribution (between the baseflow outputs from the respective NWM and NWM-BE3S) was higher (i.e., R average = 0.67) in the Colorado River basin, which showed the lowest b value distribution with the average b of 1.17. Likewise, the lowest correspondence of the baseflow estimates (i.e., R average = 0.43) was found in the Trinity River basin, where the recessions were highly non-linear (i.e., slope b = 1.52). Figure 6 a provides an insight that the outflow from the Boussinesq aquifer can exhibit similar temporal dynamics with the baseflow fluxes from the NLSR if one basin, as a linked hydrologic system, shows a linear relationship between groundwater storage Q and discharge S (i.e., b = 1.0) [*Clark et al.*, 2009]. The adaptable predictive capabilities of the Boussinesq aquifer were thus further supported as the linear recession characteristics could also be depicted by the Boussinesq aquifer. We found that the Pearson's R between the baseflow estimates ranged from 0.52 – 0.99 among the catchments where the absolute value (b – 1) (i.e., $|b\text{-}1|$) is less than 0.42. The temporal agreement quickly failed when the absolute value of (b – 1) is greater than 0.5, which means increasing nonlinearity of the recessions. Overall, whether the temporal dynamics of the respective baseflow outputs agree with each other (or not) was well predicted by the slope parameter b value with a non-linear fitting (i.e., R^2 = 0.48) (Figure 6 b).

4.2.3 Bidirectional Exchange Fluxes Estimated Based on Lateral Hydraulic Gradients in the NWM-BE3S model

Closer inspection revealed that, furthermore, the sign of the modeled f_{pr} fluxes, unlike the (one-way) baseflow fluxes in the NWM, could be either positive (i.e., groundwater discharge to the river) or negative (i.e., river infiltration to the aquifer) (Figure 7). The negative f_{pr} physically means that the rapid rises in the river stage (mostly) by storm events might lead to river water discharge into the phreatic aquifer [*Liang et al.*, 2018]. These temporary changes in the river states illustrate that the river states need to be understood as temporally dynamic following the lateral head differences between the river stage and adjacent groundwater level. As we consider that the groundwater storage is the combinative effects from the vadose zone and the river channel [*Hong et al.*, 2020], the structural enhancement in the NWM-BE3S to predict the bidirectional stream-aquifer exchange fluxes is not only capable of allowing channel losses but also improving the feasibility of the catchment-scale water budget closure.

4.3 Improved Streamflow Predictions

From the understood differences between the modeled outputs B_f and f_{pr}, we considered that the outflow from the implemented Boussinesq aquifer could yield 'more' pronounced improvements in streamflow predictions as the catchments function as a non-linear reservoir(s) between Q and S. To better evaluate the improvements in the streamflow predictions generated by the NWM-BE3S (compared to the original NWM), the evaluation of the streamflow outputs from the respective models against streamflow observations was carried out in consideration of the variability in the recession nonlinearity, which varies by region and time period.

1) Since the aquifers in each basin function differently in generating baseflow (i.e., high nonlinearity – Trinity River basin, low nonlinearity – Colorado River basin), the improvements in the streamflow predictions should be evaluated by the basin.
2) Even if it is the same basin, the recession nonlinearity appeared distinctly across time. We thus divided one year (2017) into four-month periods (i.e., JFMA (1-4), MJJA (5-8), SOND (9-12)), at which apparent differences in recession slope characteristics were identified.

Trinity River basin - The Trinity River basin exhibited the highest nonlinearity in the recession data among the three study basins (as addressed in section 4.2.2). For the selected 15 catchments in the Trinity River basin, the average b in each period gradually increased as 1.78 in JFMA, 2.19 in MJJA, and 2.48 in SOND. When it comes to temporal agreements of the respective streamflow predictions against the corresponding observations, we found significant improvements in R values from all three periods. The average R-value improves from 0.18 (NWM) to 0.39 (NWM-BE3S) in JFMA period, from 0.23 (NWM) to 0.42 (NWM-BE3S) in MJJA period, and from -0.02 (NWM) to 0.31 (NWM-BE3S) in SOND period. This is mainly because the recession nonlinearities (of the streamflow observations) from the Trinity River basin were maintained high throughout the year. We also found significant reductions in the *RMSE* values across the 15 catchments. The average *RMSE* reduced from 29.1 m^3hr^{-1} (NWM) to 19.8 m^3hr^{-1} (NWM-BE3S) in JFMA period, from 64.0 m^3hr^{-1} (NWM) to 46.7 m^3hr^{-1} (NWM-BE3S) in MJJA period, and from 32.5 m^3hr^{-1} (NWM) to 21.8 m^3hr^{-1} (NWM-BE3S) in SOND period.

Brazos River basin - The average *b* in each period was found lower than those in the Trinity River, yielding the average *b* of 1.43 in JFMA, 1.19 in MJJA, and 1.66 in SOND. The Brazos River basin was found to function closest to the linear reservoir during the MJJA (2017) period (i.e., average *b* of 1.19). The average *R* (for the selected 14 catchments) improved from 0.28 (NWM) to 0.43 (NWM-BE3S) in JFMA period, from 0.35 (NWM) to 0.45 (NWM-BE3S) in MJJA period, and from 0.02 (NWM) to 0.11 (NWM-BE3S) in SOND period. Consistent with the trend found in the *b* value distributions, we found minimal improvement in *R* during the MJJA period compared to other JFMA and SOND periods. We also identified pronounced reductions in *RMSE* values as the average *RMSE* reduced from 63.8 m^3hr^{-1} (NWM) to 45.9 m^3hr^{-1} (NWM-BE3S) in JFMA period, from 183.1 m^3hr^{-1} (NWM) to 75.7 m^3hr^{-1} (NWM-BE3S) in MJJA period, and from 129.9 m^3hr^{-1} (NWM) to 44.6 m^3hr^{-1} (NWM-BE3S) in SOND period.

Colorado River basin - The trends and value distributions of *b* (for the three periods) in the Colorado River basin were found similar to those of the Brazos River basin. In the Colorado River basin, the value distribution of *b* and its average were closest to 1.0 in MJJA period while the average *b* values during JFMA, and SOND were relatively high (i.e., the average *b* of 1.82 in JFMA, 1.03 in MJJA, and 1.69 in SOND period). Notably, the average *b* during MJJA in the Colorado River was 1.03, implying that the corresponding basin functioned almost as a linear reservoir during that time. The average *R* improved from -0.06 (NWM) to 0.31 (NWM-BE3S) in JFMA period, from 0.49 (NWM) to 0.50 (NWM-BE3S) in MJJA period, and -0.03 (NWM) to 0.11 (NWM-BE3S) in SOND period. As expected, the average and value distribution of *R* showed little improvement during the MJJA period, while significantly improved *R* (i.e., average and value distribution) was found in both JFMA and SOND periods. Like the other two basins, pronounced reductions in RMSE were identified as reducing from 31.5 m^3hr^{-1} (NWM) to 20.8

m^3hr^{-1} (NWM-BE3S) in JFMA period, from 74.5 m^3hr^{-1} (NWM) to 49.7 m^3hr^{-1} (NWM-BE3S) in MJJA period, and from 32.2 m^3hr^{-1} (NWM) to 9.97 m^3hr^{-1} (NWM-BE3S) in SOND period.

We overall identified that the degree of temporal agreement between the NWM-BE3S-derived streamflow outputs and corresponding observations was consistent with the trends in the *b* value distributions. For both Brazos and Colorado basins, the *R* improvements were little/minimal during MJJA when the basins behaved more like a linear reservoir (i.e., *b* close to 1.0) while significantly improved *R* was found in the periods JFMA and SOND. We note that the improved streamflow predictive skill in the NWM-BE3S (compared to the original NWM) was also ensured by the value distributions of *R* and *RMSE* (Figure 8). The improvements (i.e., reductions) in *RMSE* were found more pronounced during the low-flow conditions (periods) than high-flow conditions. Since river waters are considered sustained mainly by groundwater discharge during the low-flow conditions (e.g., low precipitation, extended dry period), the lower *RMSE* values during the low-flow conditions showed the suitability of the Boussinesq aquifer formulation for baseflow estimation (Figure 8). Figure 9 shows comparatively how the dynamics of the observed streamflow were simulated by the two model NWM and NWM-BE3S, respectively. Consequently, we identified significant improvements in the streamflow predictions from the NWM-BE3S in terms of both representing temporal dynamics as well as reducing amount errors.

To clarify the reasons of the improved model performance, we further estimated the errors in the modeled recession durations derived from both models by comparing them with the corresponding observed recession durations. Figure 10 showed that the duration of individual recession events was significantly better represented under the implemented Boussinesq aquifer in the NWM-BE3S. The recession events were generally predicted longer than the observed periods in the original NWM, as shown by that the most errors were biased negative. We understood that the reductions in the duration errors are the results of the adaptable predictive capabilities of the Boussinesq aquifer formulation (for the recession events with distinct recession characteristics). This is because the new Boussinesq aquifer module could account for (rapidly) decreasing baseflow fluxes due to the rapid reductions in the lateral head differences (between the stream-aquifer) as the recession progresses, which yields higher nonlinearity in the recessions. Unlike the NLSR model, to sum, the streamflow predictions by the Boussinesq aquifer were found able to represent the distinct recession characteristics observed in the corresponding (actual) catchment. Consequently, we argue that the newly represented lateral hydraulic connections (based on the Boussinesq equation) enabled the more accurate representation of the recessions which start/end based on the lateral hydraulic gradients between the river stage and water table in the actual catchment system, and thus the applicability of the Boussinesq aquifer, as an alternative to the current NLSR model, to represent hillslope-channel interactions is demonstrated.

4.4 Mapping the Examined River States

The NWM-BE3S model can explicitly simulate the bidirectional exchange fluxes f_{pr} while accounting for the temporal dynamics of the catchment-scale lateral hydraulic gradients. Here we examined the state of each of the 73,436 river reaches (distributed across the study basins) based on the yearly cumulative f_{pr} fluxes over the one-year evaluation period. The cumulative f_{pr} provided insight into whether the dominant process (between the stream and the aquifer) was river water infiltration (i.e., losing reach) or groundwater discharge to the river (i.e., gaining stream). Although the states of the channel reaches were always recognized as gaining reach due to one-way representation of GW-SW exchanges (i.e., B_f) in the (original) NWM, significant spatial variability/patterns were exhibited by the NWM-BE3S-derived f_{pr} outputs. The comparison between the spatial distributions from the cumulative B_f and f_{pr} clarified that the dominant processes between the stream and the aquifer in the 73,436 catchments were spatially heterogeneous (Figure 11 a, b). The yearly cumulative f_{pr} values were estimated to be ranging from -1,750 m^3 (negative) – 1,740 m^3 (positive), while most negative f_{pr} values (i.e., about 98.8 %) were simulated in the northwestern regions, where the average annual precipitation was only one-third compared to the southeastern parts. However, Figure 11 b showed that the cumulative B_f values modeled in the catchments belonging to semi-arid hydroclimate [*Kottek et al.*, 2006] were found (instead) positively more enormous, which is counter-intuitive, than other parts of the study area.

On the other hand, we also found some similarities between the respective value distributions of f_{pr} and B_f. The yearly cumulative exchange fluxes (i.e., net exchange fluxes) ranging from 0.0 m^3 – 1.0 m^3 were simulated by about half of the catchments from both the NWM and the NWM-BE3S (i.e., 52.1% in the NWM, and 47.5 % in the NWM-BE3S), and the number of the river reaches was found to decrease exponentially with increasing the (positive) value of both f_{pr} and B_f (Figure 11 c). These similar value distributions, albeit limited for positive values, were attributed to that the estimation of the exchange fluxes f_{pr} and B_f was dependent upon the size of corresponding catchment/reach (Equation 4) in both configurations of the NWM and the NWM-BE3S. According to the NWM-BE3S f_{pr} outputs, we found that river water infiltration was the dominant process among 10.1 % of the river reaches (i.e., 7,417 reaches) during the evaluation period 2017.

5 Conclusions and Future Work

Although the relevance of hydraulic groundwater theory in understanding the role of the aquifer(s) in the development of surface-subsurface hydrologic cycle has been addressed over the past few decades, the lack of an appropriate modeling structure was the primary constraint to applying the well-studied theory to large-scale hydrologic predictions [*Clark et al.*, 2015; *Fan et al.*, 2019; *Rupp and Selker*, 2006; *Troch et al.*, 2013]. With the aim of representing the theory-based depiction of GW-SW interactions (i.e., hillslope-stream connectivity) in an integrated hydrologic model, we established a novel hydrologic framework NWM-BE3S by integrating the BE3S's Boussinesq equation-based representation of (catchment-scale) aquifer into the NWM. The applicability of the Boussinesq formulation to large-scale hydrologic predictions was successfully demonstrated based on the improved predictive performance for streamflow identified in the NWM-BE3S model. To ensure the validity of groundwater flow based on the Dupuit-Forcheimer assumption in the large-scale integrated modeling, the primary premise of our comparative evaluation (between the Boussinesq aquifer and the NLSR) has been that the effects made by the newly implemented Boussinesq aquifer should be consistent with the theory (i.e., theory-based evaluation). In this context, the varying degree of improvement in streamflow predictions (by the NWM-BE3S) according to the recession nonlinearities (i.e., recession slope *b*) manifests (1) the applicability of the theory-based depiction of hillslope-channel interactions as well as (2) the technical enhancement of model structure.

For future work, we will mainly focus on the effects of fluvial system dynamics on horizontal groundwater redistribution and resultant impacts on land surface water/energy processes (e.g., atmospheric boundary layers (ABL) processes). While some studies attempted to understand the interactive relationship between (horizontal) subsurface flow and land surface processes, however, the groundwater flows as the response of fluvial system dynamics has not been adequately represented in most hydrologic models as well as land surface models (LSMs) / earth system models (ESMs) [*Bisht et al.*, 2017; *Fan et al.*, 2007; *Fan et al.*, 2019; *Gochis*, 2018; *Stefan J. Kollet and Maxwell*, 2008; *Lawrence et al.*, 2019; *Maxwell and Kollet*, 2008; *Wu et al.*, 2021]. By representing the bidirectional lateral hydraulic connections between the hillslopes and the open channel reaches (GW-SW interactions), the presented NWM-BE3S also provides one unique forward modeling capability to explicitly incorporate the spatiotemporally dynamic groundwater flow *as the results of bidirectional lateral hillslope-channel connectivity* into land surface schemes. Relying on understood effects of groundwater dynamics on vadose zone processes [*Hong et al.*, 2020], we thus aim to comprehensively understand the interaction patterns (over large areas) between the vadose zone-groundwater-channel that vary according to climatic/hydro(geo)logic conditions and its resulting impacts on the land surface water/energy budget (e.g., near-surface soil moisture, and ET) and thus ABL processes.

Acknowledgments, Samples, and Data

We acknowledge the funding support from the USDA grant (Project No. 2017-68007-26318) through the National Institute for Food and Agriculture's Agriculture and Food Research Initiative, Water for Agricultural Challenge Area, W-4188 multi-state HATCH project, and USGS TWRI Graduate Research Program. The dataset used in this study are available from the USGS website for the watersheds in the United States at https://maps.waterdata.usgs.gov/mapper/index.html

References

Alfieri, L., P. Burek, E. Dutra, B. Krzeminski, D. Muraro, J. Thielen, and F. Pappenberger (2013), Glofas – Global Ensemble Streamflow Forecasting and Flood Early Warning, *Hydrology and Earth System Sciences*, *17*(3), 1161-1175.

Anderson, M. P., W. W. Woessner, and R. J. Hunt (2015), *Applied Groundwater Modeling: Simulation of Flow and Advective Transport*, Academic press.

Baroni, G., B. Schalge, O. Rakovec, R. Kumar, L. Schüler, L. Samaniego, C. Simmer, and S. Attinger (2019), A Comprehensive Distributed Hydrological Modeling Intercomparison to Support Process Representation and Data Collection Strategies, *Water Resources Research*, *55*(2), 990-1010.

Barthel, R. (2014), Hess Opinions &Quot;Integration of Groundwater and Surface Water Research: An Interdisciplinary Problem?&Quot, *Hydrology and Earth System Sciences*, *18*(7), 2615-2628.

Basha, H. A. (2013), Traveling Wave Solution of the Boussinesq Equation for Groundwater Flow in Horizontal Aquifers, *Water Resources Research*, *49*(3), 1668-1679.

Berg, S. J., and W. A. Illman (2011), Capturing Aquifer Heterogeneity: Comparison of Approaches through Controlled Sandbox Experiments, *Water Resources Research*, *47*(9), n/a-n/a.

Binley, A., K. Beven, and J. Elgy (1989), A Physically Based Model of Heterogeneous Hillslopes: 2. Effective Hydraulic Conductivities, *Water Resources Research*, *25*(6), 1227-1233.

Bisht, G., M. Y. Huang, T. Zhou, X. Y. Chen, H. Dai, G. E. Hammond, W. J. Riley, J. L. Downs, Y. Liu, and J. M. Zachara (2017), Coupling a Three-Dimensional Subsurface Flow and Transport Model with a Land Surface Model to Simulate Stream-Aquifer-Land Interactions (Cp V1.0), *Geoscientific Model Development*, *10*(12), 4539-4562.

Biswal, B., and M. Marani (2010), Geomorphological Origin of Recession Curves, *Geophysical Research Letters*, *37*(24), n/a-n/a.

Boussinesq, J. (1904), Recherches the´Oriques Sur L'e´Coulement Des Nappes D'eau Infiltre´Es Dans Le Sol Et Sur De´Bit De Sources, *Journal de Mathématiques Pures et Appliquées*, *10*, 5-78.

Brutsaert, W., and J. L. Nieber (1977), Regionalized Drought Flow Hydrographs from a Mature Glaciated Plateau, *Water Resources Research*, *13*(3), 637-643.

Brutsaert, W., and J. P. Lopez (1998), Basin-Scale Geohydrologic Drought Flow Features of Riparian Aquifers in the Southern Great Plains, *Water Resources Research*, *34*(2), 233-240.

Clark, M. P., D. E. Rupp, R. A. Woods, H. J. Tromp-Van Meerveld, N. E. Peters, and J. E. Freer (2009), Consistency between Hydrological Models and Field Observations: Linking Processes at the Hillslope Scale to Hydrological Responses at the Watershed Scale, *Hydrological Processes*, *23*(2), 311-319.

Clark, M. P., et al. (2015), Improving the Representation of Hydrologic Processes in Earth System Models, *Water Resources Research*, *51*(8), 5929-5956.

Dewandel, B., Y. Caballero, J. Perrin, A. Boisson, F. Dazin, S. Ferrant, S. Chandra, and J.-C. Maréchal (2017), A Methodology for Regionalizing 3-D Effective Porosity at Watershed Scale in Crystalline Aquifers, *Hydrological Processes*, *31*(12), 2277-2295.

Dralle, D., N. Karst, and S. E. Thompson (2015), A, B Careful: The Challenge of Scale Invariance for Comparative Analyses in Power Law Models of the Streamflow Recession, *Geophysical Research Letters*, *42*(21), 9285-9293.

Dralle, D. N., N. J. Karst, K. Charalampous, A. Veenstra, and S. E. Thompson (2017), Event-Scale Power Law Recession Analysis: Quantifying Methodological Uncertainty, *Hydrology and Earth System Sciences*, *21*(1), 65-81.

Fallico, C., S. De Bartolo, M. Veltri, and G. Severino (2016), On the Dependence of the Saturated Hydraulic Conductivity Upon the Effective Porosity through a Power Law Model at Different Scales, *Hydrological Processes*, *30*(13), 2366-2372.

Fan, Y., G. Miguez-Macho, C. P. Weaver, R. Walko, and A. Robock (2007), Incorporating Water Table Dynamics in Climate Modeling: 1. Water Table Observations and Equilibrium Water Table Simulations, *Journal of Geophysical Research-Atmospheres*, *112*(D10).

Fan, Y., et al. (2019), Hillslope Hydrology in Global Change Research and Earth System Modeling, *Water Resources Research*, *55*(2), 1737-1772.

Fleckenstein, J. H., S. Krause, D. M. Hannah, and F. Boano (2010), Groundwater-Surface Water Interactions: New Methods and Models to Improve Understanding of Processes and Dynamics, *Advances in Water Resources*, *33*(11), 1291-1295.

Furman, A. (2008), Modeling Coupled Surface-Subsurface Flow Processes: A Review, *Vadose Zone Journal*, *7*(2), 741-756.

Gauthier, M. J., M. Camporese, C. Rivard, C. Paniconi, and M. Larocque (2009), A Modeling Study of Heterogeneity and Surface Water-Groundwater Interactions in the Thomas Brook Catchment, Annapolis Valley (Nova Scotia, Canada), *Hydrology and Earth System Sciences*, *13*(9), 1583-1596.

Gochis, D. J., M. Barlage, A. Dugger, K. FitzGerald, L. Karsten, M. McAllister, J. McCreight, J. Mills, A. RafieeiNasab, L. Read, K. Sampson, D. Yates, W. Yu, (2018), The Ncar Wrf-Hydro Modeling System Technical Description (Version 5.0), National Center for Atmospheric Research.

Gómez-Hernández, J. J., and S. M. Gorelick (1989), Effective Groundwater Model Parameter Values: Influence of Spatial Variability of Hydraulic Conductivity, Leakance, and Recharge, *Water Resources Research*, *25*(3), 405-419.

Hong, M., B. P. Mohanty, and Z. Sheng (2020), An Explicit Scheme to Represent the Bidirectional Hydrologic Exchanges between the Vadose Zone, Phreatic Aquifer, and River, *Water Resources Research*, *56*(9).

Hornberger, G. M., and I. Remson (1970), A Moving Boundary Model of a One-Dimensional Saturated-Unsaturated, Transient Porous Flow System, *Water Resources Research*, *6*(3), 898-+.

Hornberger, G. M., J. Ebert, and I. Remson (1970), Numerical Solution of Boussinesq Equation for Aquifer-Stream Interaction, *Water Resources Research*, *6*(2), 601-+.

Huntington, J. L., and R. G. Niswonger (2012), Role of Surface-Water and Groundwater Interactions on Projected Summertime Streamflow in Snow Dominated Regions: An Integrated Modeling Approach, *Water Resources Research*, *48*(11), n/a-n/a.

Jachens, E. R., D. E. Rupp, C. Roques, and J. S. Selker (2020), Recession Analysis Revisited: Impacts of Climate on Parameter Estimation, *Hydrology and Earth System Sciences*, *24*(3), 1159-1170.

Jasper, K., J. Gurtz, and H. Lang (2002), Advanced Flood Forecasting in Alpine Watersheds by Coupling Meteorological Observations and Forecasts with a Distributed Hydrological Model, *Journal of Hydrology*, *267*(1-2), 40-52.

Karki, R., J. M. Krienert, M. Hong, and D. R. Steward (2021), Evaluating Baseflow Simulation in the National Water Model: A Case Study in the Northern High Plains Region, USA, *JAWRA Journal of the American Water Resources Association*, *57*(2), 267-280.

Karlsen, R. H., K. Bishop, T. Grabs, M. Ottosson-Löfvenius, H. Laudon, and J. Seibert (2019), The Role of Landscape Properties, Storage and Evapotranspiration on Variability in Streamflow Recessions in a Boreal Catchment, *Journal of Hydrology*, *570*, 315-328.

Kim, N. W., I. M. Chung, Y. S. Won, and J. G. Arnold (2008), Development and Application of the Integrated Swat–Modflow Model, *Journal of Hydrology*, *356*(1-2), 1-16.

Kollet, S., et al. (2017), The Integrated Hydrologic Model Intercomparison Project, Ih-Mip2: A Second Set of Benchmark Results to Diagnose Integrated Hydrology and Feedbacks, *Water Resources Research*, *53*(1), 867-890.

Kollet, S. J., and R. M. Maxwell (2006), Integrated Surface-Groundwater Flow Modeling: A Free-Surface Overland Flow Boundary Condition in a Parallel Groundwater Flow Model, *Advances in Water Resources*, *29*(7), 945-958.
Kollet, S. J., and R. M. Maxwell (2008), Capturing the Influence of Groundwater Dynamics on Land Surface Processes Using an Integrated, Distributed Watershed Model, *Water Resources Research*, *44*(2), n/a-n/a.
Kottek, M., J. Grieser, C. Beck, B. Rudolf, and F. Rubel (2006), World Map of the Köppen-Geiger Climate Classification Updated, *Meteorologische Zeitschrift*, *15*(3), 259-263.
Lahmers, T. M., H. Gupta, C. L. Castro, D. J. Gochis, D. Yates, A. Dugger, D. Goodrich, and P. Hazenberg (2019), Enhancing the Structure of the Wrf-Hydro Hydrologic Model for Semiarid Environments, *J. Hydrometeorol.*, *20*(4), 691-714.
Lautz, L. K., and D. I. Siegel (2006), Modeling Surface and Ground Water Mixing in the Hyporheic Zone Using Modflow and Mt3d, *Advances in Water Resources*, *29*(11), 1618-1633.
Lawrence, D. M., et al. (2019), The Community Land Model Version 5: Description of New Features, Benchmarking, and Impact of Forcing Uncertainty, *Journal of Advances in Modeling Earth Systems*, *11*(12), 4245-4287.
Liang, X., H. Zhan, and K. Schilling (2018), Spatiotemporal Responses of Groundwater Flow and Aquifer-River Exchanges to Flood Events, *Water Resources Research*, *54*(3), 1513-1532.
Lockington, D. A. (1997), Response of Unconfined Aquifer to Sudden Change in Boundary Head, *Journal of Irrigation and Drainage Engineering*, *123*(1), 24-27.
Maurer, E. P. (2005), Uncertainty in Projections of Streamflow Changes Due to Climate Change in California, *Geophysical Research Letters*, *32*(3).
Maxwell, R. M., and S. J. Kollet (2008), Interdependence of Groundwater Dynamics and Land-Energy Feedbacks under Climate Change, *Nature Geoscience*, *1*(10), 665-669.
Maxwell, R. M., and L. E. Condon (2016), Connections between Groundwater Flow and Transpiration Partitioning, *353*(6297), 377-380.
Maxwell, R. M., L. E. Condon, and S. J. Kollet (2015), A High-Resolution Simulation of Groundwater and Surface Water over Most of the Continental Us with the Integrated Hydrologic Model Parflow V3, *Geoscientific Model Development*, *8*(3), 923-937.
Maxwell, R. M., et al. (2014), Surface-Subsurface Model Intercomparison: A First Set of Benchmark Results to Diagnose Integrated Hydrology and Feedbacks, *Water Resources Research*, *50*(2), 1531-1549.
McDonnell, J. J., et al. (2007), Moving Beyond Heterogeneity and Process Complexity: A New Vision for Watershed Hydrology, *Water Resources Research*, *43*(7), n/a-n/a.
Nijssen, B., D. P. Lettenmaier, X. Liang, S. W. Wetzel, and E. F. Wood (1997), Streamflow Simulation for Continental-Scale River Basins, *Water Resources Research*, *33*(4), 711-724.
Paniconi, C., P. A. Troch, E. E. Van Loon, and A. G. J. Hilberts (2003), Hillslope-Storage Boussinesq Model for Subsurface Flow and Variable Source Areas Along Complex Hillslopes: 2. Intercomparison with a Three-Dimensional Richards Equation Model, *Water Resources Research*, *39*(11), n/a-n/a.
Polubarinova-Koch, P. I. (2015), *Theory of Ground Water Movement*, Princeton University Press.

Rinaudo, J.-D. (2015), Long-Term Water Demand Forecasting, edited, pp. 239-268, Springer Netherlands.

Rorabaugh, M. (1964), Estimating Changes in Bank Storage and Ground-Water Contribution to Streamflow: Extract of Publication No. 63 of the Iash Symposium Surface Waters.

Rupp, D. E., and J. S. Selker (2005), Drainage of a Horizontal Boussinesq Aquifer with a Power Law Hydraulic Conductivity Profile, *Water Resources Research*, *41*(11), n/a-n/a.

Rupp, D. E., and J. S. Selker (2006), On the Use of the Boussinesq Equation for Interpreting Recession Hydrographs from Sloping Aquifers, *Water Resources Research*, *42*(12), n/a-n/a.

Salas, F. R., M. A. Somos-Valenzuela, A. Dugger, D. R. Maidment, D. J. Gochis, C. H. David, W. Yu, D. Ding, E. P. Clark, and N. Noman (2018), Towards Real-Time Continental Scale Streamflow Simulation in Continuous and Discrete Space, *JAWRA Journal of the American Water Resources Association*, *54*(1), 7-27.

Savenije, H. H., and P. Van der Zaag (2008), Integrated Water Resources Management: Concepts and Issues, *Physics and Chemistry of the Earth, Parts A/B/C*, *33*(5), 290-297.

Schwarz, G. E., S. E. Jackson, and M. E. o. Wieczorek (2018), Select Attributes for Nhdplus Version 2.1 Reach Catchments and Modified Network Routed Upstream Watersheds for the Conterminous United States, edited by U. S. G. Survey.

Seo, Y. S., J. Šimůnek, and E. P. Poeter (2007), Documentation of the Hydrus Package for Modflow-2000, the U.S. Geological Survey Modular Ground-Water Model, Ground Water Modeling Center.

Shaw, S. B., and S. J. Riha (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*, *434-435*, 46-54.

Shen, C., J. Niu, and M. S. Phanikumar (2013), Evaluating Controls on Coupled Hydrologic and Vegetation Dynamics in a Humid Continental Climate Watershed Using a Subsurface-Land Surface Processes Model, *Water Resources Research*, *49*(5), 2552-2572.

Shen, C. P., and M. S. Phanikumar (2010), A Process-Based, Distributed Hydrologic Model Based on a Large-Scale Method for Surface-Subsurface Coupling, *Advances in Water Resources*, *33*(12), 1524-1541.

Shin, Y., B. P. Mohanty, and A. V. M. Ines (2013), Estimating Effective Soil Hydraulic Properties Using Spatially Distributed Soil Moisture and Evapotranspiration, *Vadose Zone Journal*, *12*(3), vzj2012.0094.

Stoelzle, M., K. Stahl, and M. Weiler (2013), Are Streamflow Recession Characteristics Really Characteristic?, *Hydrology and Earth System Sciences*, *17*(2), 817-828.

Szilagyi, J., M. B. Parlange, and J. D. Albertson (1998), Recession Flow Analysis for Aquifer Parameter Determination, *Water Resources Research*, *34*(7), 1851-1857.

Tallaksen, L. M. (1995), A Review of Baseflow Recession Analysis, *Journal of Hydrology*, *165*(1-4), 349-370.

Tashie, A., T. Pavelsky, and L. E. Band (2020), An Empirical Reevaluation of Streamflow Recession Analysis at the Continental Scale, *Water Resources Research*, *56*(1).

Tesfa, T. K., H. Y. Li, L. R. Leung, M. Huang, Y. Ke, Y. Sun, and Y. Liu (2014), A Subbasin-Based Framework to Represent Land Surface Processes in an Earth System Model, *Geoscientific Model Development*, *7*(3), 947-963.

Tijerina, D., L. Condon, K. Fitzgerald, A. Dugger, M. M. O'Neill, K. Sampson, D. Gochis, and R. Maxwell (2021), Continental Hydrologic Intercomparison Project, Phase 1: A Large-Scale Hydrologic Model Comparison over the Continental United States, *Water Resources Research*, *57*(7).

Troch, P. A., F. P. De Troch, and W. Brutsaert (1993), Effective Water Table Depth to Describe Initial Conditions Prior to Storm Rainfall in Humid Regions, *Water Resources Research*, *29*(2), 427-434.

Troch, P. A., C. Paniconi, and E. Emiel Van Loon (2003), Hillslope-Storage Boussinesq Model for Subsurface Flow and Variable Source Areas Along Complex Hillslopes: 1. Formulation and Characteristic Response, *Water Resources Research*, *39*(11), n/a-n/a.

Troch, P. A., 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.

Vannier, O., I. Braud, and S. Anquetin (2014), Regional Estimation of Catchment-Scale Soil Properties by Means of Streamflow Recession Analysis for Use in Distributed Hydrological Models, *Hydrological Processes*, *28*(26), 6276-6291.

Viterbo, F., K. Mahoney, L. Read, F. Salas, B. Bates, J. Elliott, B. Cosgrove, A. Dugger, D. Gochis, and R. Cifelli (2020), A Multiscale, Hydrometeorological Forecast Evaluation of National Water Model Forecasts of the May 2018 Ellicott City, Maryland, Flood, *J. Hydrometeorol.*, *21*(3), 475-499.

Vogel, R. M., and C. N. Kroll (1992), Regional Geohydrologic-Geomorphic Relationships for the Estimation of Low-Flow Statistics, *Water Resources Research*, *28*(9), 2451-2458.

Wagener, T., M. Sivapalan, P. Troch, and R. Woods (2007), Catchment Classification and Hydrologic Similarity, *Geography Compass*, *1*(4), 901-931.

Wang, D. (2011), On the Base Flow Recession at the Panola Mountain Research Watershed, Georgia, United States, *Water Resources Research*, *47*(3), n/a-n/a.

Wieczorek, M., and A. E. LaMotte (2010), Attributes for Nhdplus Catchments (Version 1.1) for the Conterminous United States: Statsgo Soil Characteristics, *Report Rep. 490-26*, Reston, VA.

Wu, R., X. Chen, G. Hammond, G. Bisht, X. Song, M. Huang, G.-Y. Niu, and T. Ferre (2021), Coupling Surface Flow with High-Performance Subsurface Reactive Flow and Transport Code Pflotran, *Environmental Modelling & Software*, *137*, 104959.

Zhao, T., X. Cai, and D. Yang (2011), Effect of Streamflow Forecast Uncertainty on Real-Time Reservoir Operation, *Advances in Water Resources*, *34*(4), 495-504.

Zhu, J., and B. P. Mohanty (2003), Effective Hydraulic Parameters for Steady State Vertical Flow in Heterogeneous Soils, *Water Resources Research*, *39*(8), n/a-n/a.

List of Tables

Table 1. Physics Differences in the Representation Scheme for the Aquifer (Subsurface) System between the Original NWM (i.e., NLSR Module) and the Presented NWM-BE3S model (i.e., Boussinesq Aquifer Module). Except for the Aquifer Representation Scheme, All Other Physics Configurations were Preserved in the Newly Developed NWM-BE3S.

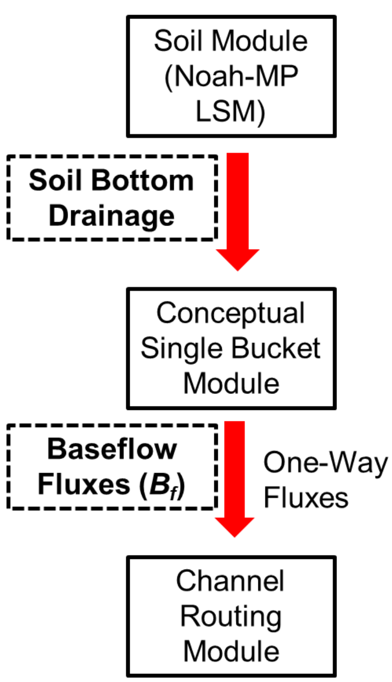
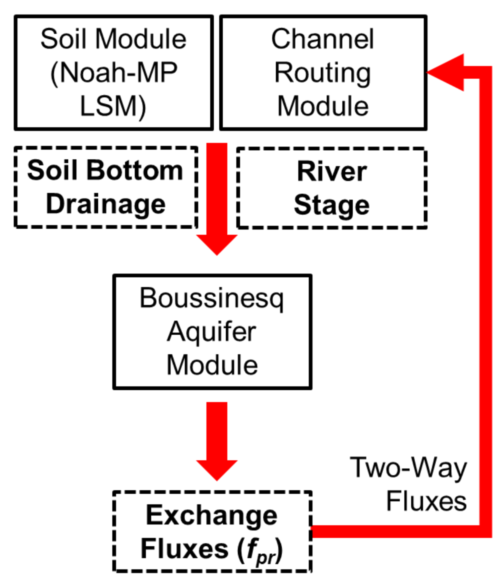
<table>
<tr><th>Representation Scheme for The Aquifer(s) and GW-SW Exchanges</th><th>The NWM</th><th>The NWM-BE3S</th></tr>
<tr><td>Model Physics</td><td>Exponential Functionality between Bucket Height (z) and Baseflow Fluxes B_f (Equation 7)</td><td>The Boussinesq Equation-Based Catchment-Scale Lateral Groundwater Flow (Equation 1, and 2)</td></tr>
<tr><td>Parameters</td><td>Empirical Parameters
-Two Fitting Parameters C and Exp
-Z_{max} to Define the Storage Capacity of the Conceptual Bucket</td><td>Physical Parameters
-Horizontal Saturated Hydraulic Conductivity (K_s)
-Effective (Drainable) Porosity (f)</td></tr>
<tr><td>Resolution</td><td>Aggregated on Catchment-Scale</td><td>Lateral Resolution 50-m</td></tr>
<tr><td>Module Input & Output</td><td>
</td><td>
</td></tr>
</table>

Table 2. Primary Physical Characteristics of the Selected 40 Catchments (Distributed over the Studied Three Major Basins). (* data derived from the NWM input, ** from the STATSGO2)

Basin	USGS Gage Number	Area (km^2)	Total Stream Length (km)	Stream Order	Channel Bottom Width (m)*	Catchment Average Permeability (ms^{-1})**
Trinity River Basin	08048543	0.52	0.91	5	29.59	0.7
	084956950	11.18	5.39	2	9.17	3.31
	08064100	16.57	8.89	4	19.70	0.55
	08065800	1.59	1.39	5	14.54	1.57
	08063562	12.59	15.58	4	15.33	0.51
	08062500	10.07	18.17	6	43.19	0.5
	08062700	70.14	25.73	6	43.90	0.74
	08055560	39.46	7.42	6	29.12	1.44
	08045550	33.27	10.94	5	27.16	1.3
	08048000	3.81	3.70	5	29.33	0.9
	08049300	8.20	4.92	5	30.67	0.9
	08049500	8.51	6.31	5	30.96	0.87
	08057000	3.95	5.70	6	39.17	0.51
	08065000	3.90	2.37	6	50.49	0.57
	08065350	1.34	2.69	6	51.82	0.23
Brazos River Basin	08082000	0.75	0.37	5	34.02	1.36
	08091000	13.52	6.80	7	59.10	1.96
	08093100	36.90	7.23	7	60.47	2.16
	08162000	4.69	3.99	6	72.78	0.81
	08108700	12.60	1.33	7	70.04	1.48
	08111500	4.16	3.61	7	73.29	0.29
	08111850	19.17	7.04	7	73.76	0.53
	08114000	8.41	7.96	7	73.99	1.23
	08095300	26.08	4.34	3	12.07	0.71
	08104300	8.17	3.27	3	10.81	0.49
	08099382	10.36	6.87	3	11.04	1.05
	08090800	2.97	1.69	7	58.51	2.14
	08088610	9.42	3.90	7	56.83	1.36
	08116650	2.53	3.71	7	74.19	1.24
Colorado River Basin	08159200	1.95	2.06	6	71.42	1.31
	08159500	1.03	1.41	6	71.75	1.3
	08161000	8.89	5.34	6	72.54	1.92
	08143600	0.47	1.12	5	27.14	0.42
	08123850	3.36	2.76	6	47.97	2.12
	08136700	14.35	6.09	6	58.57	1.7
	08109700	14.42	6.09	4	13.05	1.39
	08158380	26.42	9.83	2	3.75	1.55
	08160800	10.21	5.41	2	5.50	5.45
	08155300	3.65	3.14	3	10.22	0.82
	08158970	4.28	4.73	4	6.34	0.74

Table 3. Recession Parameters for Each Time Domain (i.e., Early and Late Time) and Resulting Determined Effective Aquifer Properties on Catchment-Scale for the Selected 40 Catchments.

Basin	*Log(a)* (Early) (Fixed *b* 3.0)	*Log(a)* (Late) (Fixed *b* 1.5)	K_s (ms^{-1})	*f* (%)	Initial Aquifer Thickness D_{ini} (m)	The size of Contributing Aquifer (km^2)
Trinity River Basin	-3.36	-1.78	4.80E-04	2.18E-01	45.3	1.77E-02
	-3.77	-1.89	8.82E-03	4.01E+00	0.6	9.06E-01
	-6.86	-2.58	4.81E-03	2.18E+00	7.4	1.52E+00
	-4.15	-1.97	4.42E-03	2.01E+00	3.8	1.30E-01
	-5.19	-2.39	4.12E-03	1.87E+00	2.1	3.53E-01
	-7.61	-3.26	2.38E-04	1.08E-01	60.9	6.75E-01
	-7.55	-3.22	4.28E-03	1.95E+00	8	4.35E+00
	-5.89	-2.49	9.99E-03	4.54E+00	2.3	2.45E+00
	-3.32	-2.16	8.70E-03	3.95E+00	0.3	9.32E-01
	-4.56	-2.78	2.73E-03	1.24E+00	7.5	3.20E-01
	-6.61	-3.15	2.80E-03	1.27E+00	29.5	3.69E-01
	-6.61	-3.05	3.01E-03	1.37E+00	24.4	4.34E-01
	-7.15	-3.13	2.66E-04	1.21E-01	89.2	3.67E-01
	-7.81	-3.25	8.24E-04	3.75E-01	182.1	1.17E-01
	-8.02	-3.46	1.01E-05	4.61E-03	848	9.65E-02
Brazos River Basin	-4.2	-1.75	5.93E-03	2.70E+00	5.9	1.88E-02
	-4.3	-2.15	8.27E-03	3.76E+00	0.7	1.24E+00
	-5.32	-2.61	9.67E-03	4.39E+00	1.4	3.32E-01
	-8.1	-3.2	6.41E-04	2.91E-01	174.8	2.81E-02
	-7.68	-3.29	4.58E-03	2.08E+00	54.1	1.26E-02
	-8.65	-3.47	1.57E-04	7.13E-02	455	4.16E-03
	-8.81	-3.3	3.84E-03	1.75E+00	86	1.55E+00
	-9	-3.25	7.42E-04	3.37E-01	209.6	7.57E-01
	-5.95	-2.67	9.79E-03	4.45E+00	2.9	2.14E+00
	-5.08	-2.69	4.37E-03	1.99E+00	4.5	2.61E-01
	-3.43	-2.21	5.60E-03	2.54E+00	0.5	9.22E-01
	-6.47	-2.83	3.78E-03	1.72E+00	45.1	2.23E-01
	-6.4	-3.3	3.49E-03	1.58E+00	27.3	6.22E-01
	-8.9	-3.3	1.49E-04	6.78E-02	472	1.14E-01
Colorado River Basin	-7.98	-3.29	1.14E-04	5.16E-02	442	1.31E-01
	-7.75	-3.42	1.93E-05	8.76E-03	854	3.50E-02
	-5.01	-2.6	4.19E-03	1.91E+00	3.5	1.51E-01
	-4.65	-2.27	5.96E-04	2.71E-01	29.7	1.69E-02
	-3.95	-2.07	4.69E-03	2.13E+00	1.8	3.12E-01
	-3.62	-2.16	9.54E-03	4.34E+00	0.4	1.15E-01
	-4.48	-2.44	6.07E-03	2.76E+00	1.1	7.21E-01
	-1.65	-1.62	5.99E-03	2.72E+00	0.2	5.28E-02
	-1.89	-1.64	1.04E-02	4.71E+00	0.14	5.92E-01
	-5.62	-2.61	3.82E-03	1.74E+00	13.6	2.96E-01
	-2.37	-1.68	5.72E-03	2.60E+00	0.3	1.11E-01

List of Figures

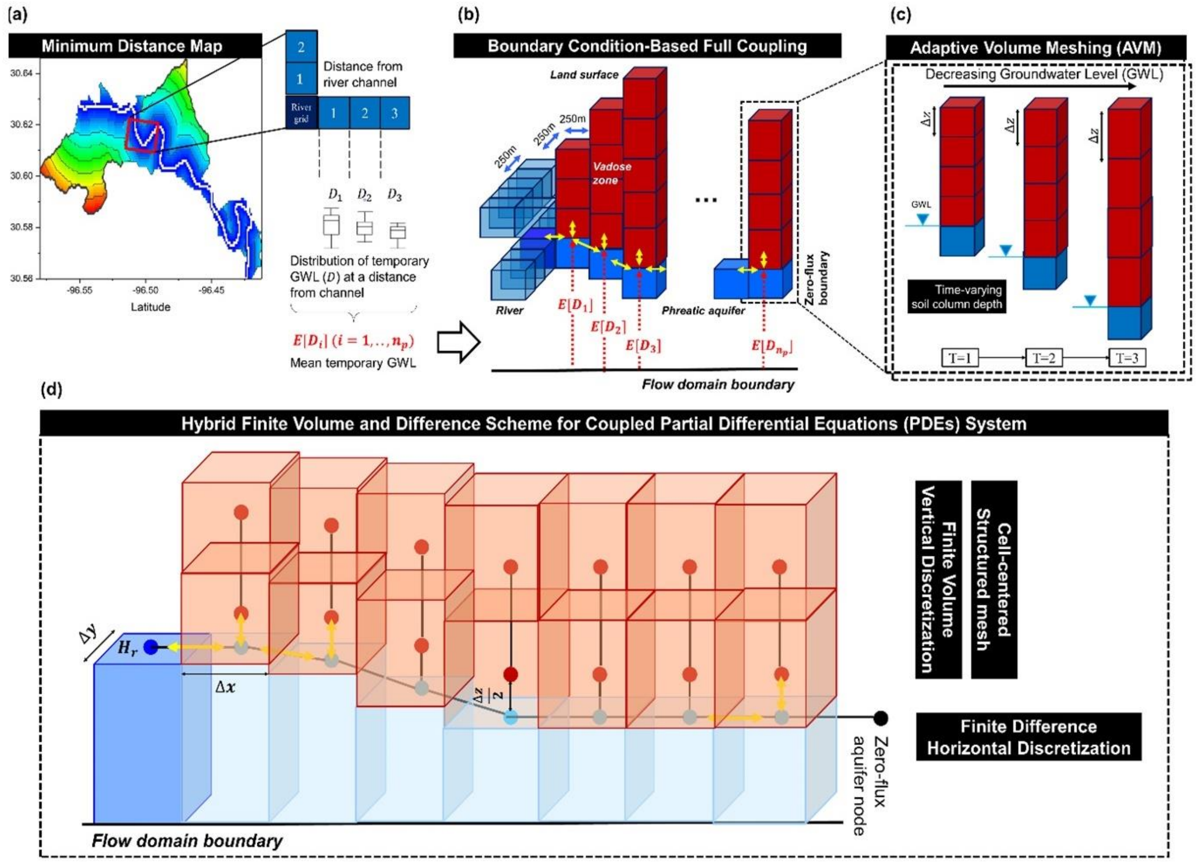


Figure 1. A Schematic of Boundary Condition-based Coupling Between River, Phreatic aquifer, and Vadose zone. (a) Euclidean Minimum Distance Map to Set Up Horizontal Profile of Temporary Aquifer Thickness *E[D*$_i$*]*, (b) Boundary Condition-Based Coupling Approach, Especially for Numerical Representation of Boussinesq Solutions under Transient River Stage Conditions during the Recession, (c) Soil Grid According to Time-Dependent Groundwater Level (GWL) using the Adaptive Volume Meshing (AVM) Technique, and (d) Hybrid Node-based Finite Volume and Difference Scheme. Adapted from "An Explicit Scheme to Represent the Bidirectional Hydrologic Exchanges Between the Vadose Zone, Phreatic Aquifer, and River" by *Hong et al.* [2020], Water Resources Research, by the American Geophysical Union.

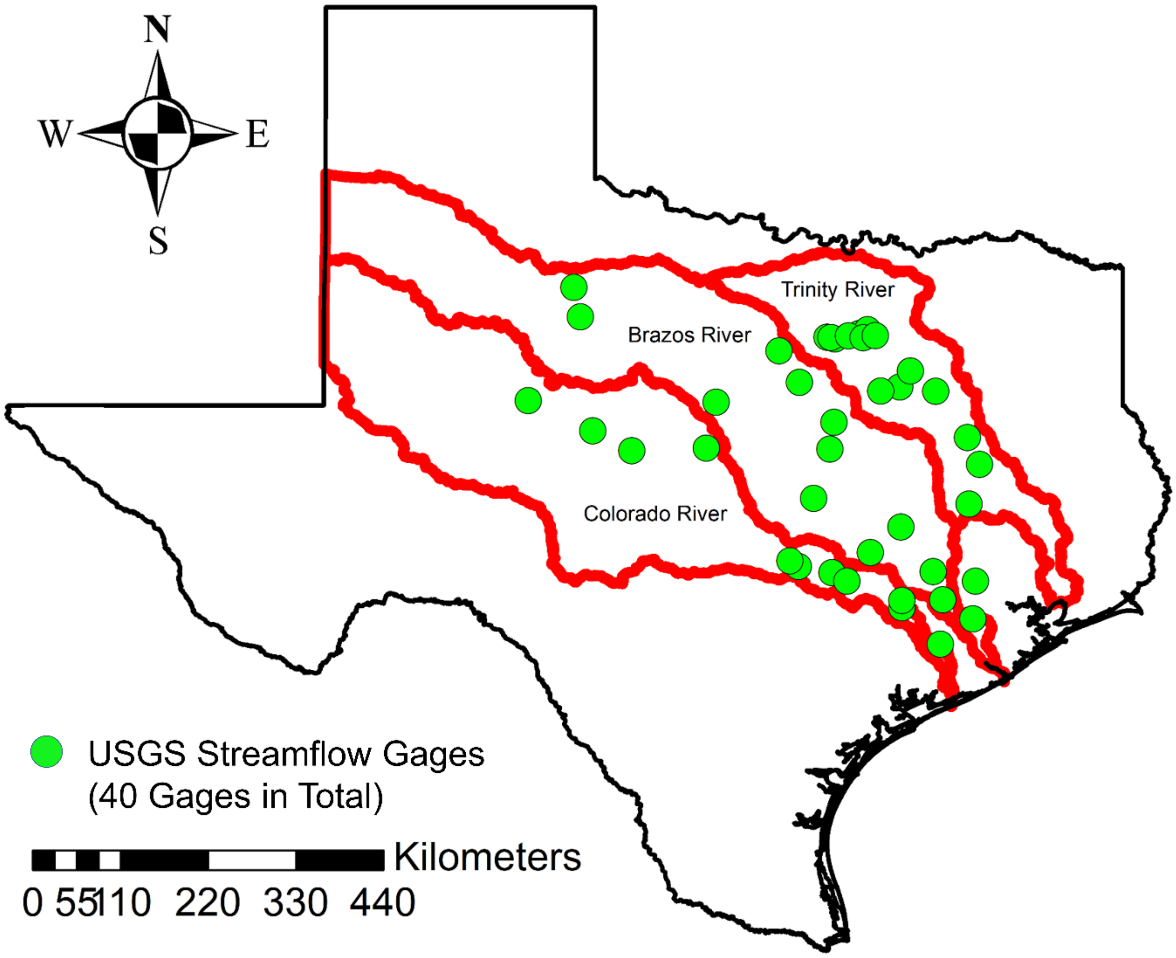


Figure 2. Study Area Including the Entire Spatial Extent of the Trinity River Basin, the Brazos River Basin, and the Colorado River Basin Within Texas. 40 USGS Streamflow Gauges (i.e., Trinity River Basin - 15 Gauges, Brazos River Basin - 14 Gauges, and Colorado River Basin - 11 gauges) Were Selected to Assess the Improved Predictive Performance for Streamflow Due to the Implemented Boussinesq Aquifer Module in the NWM-BE3S Model.

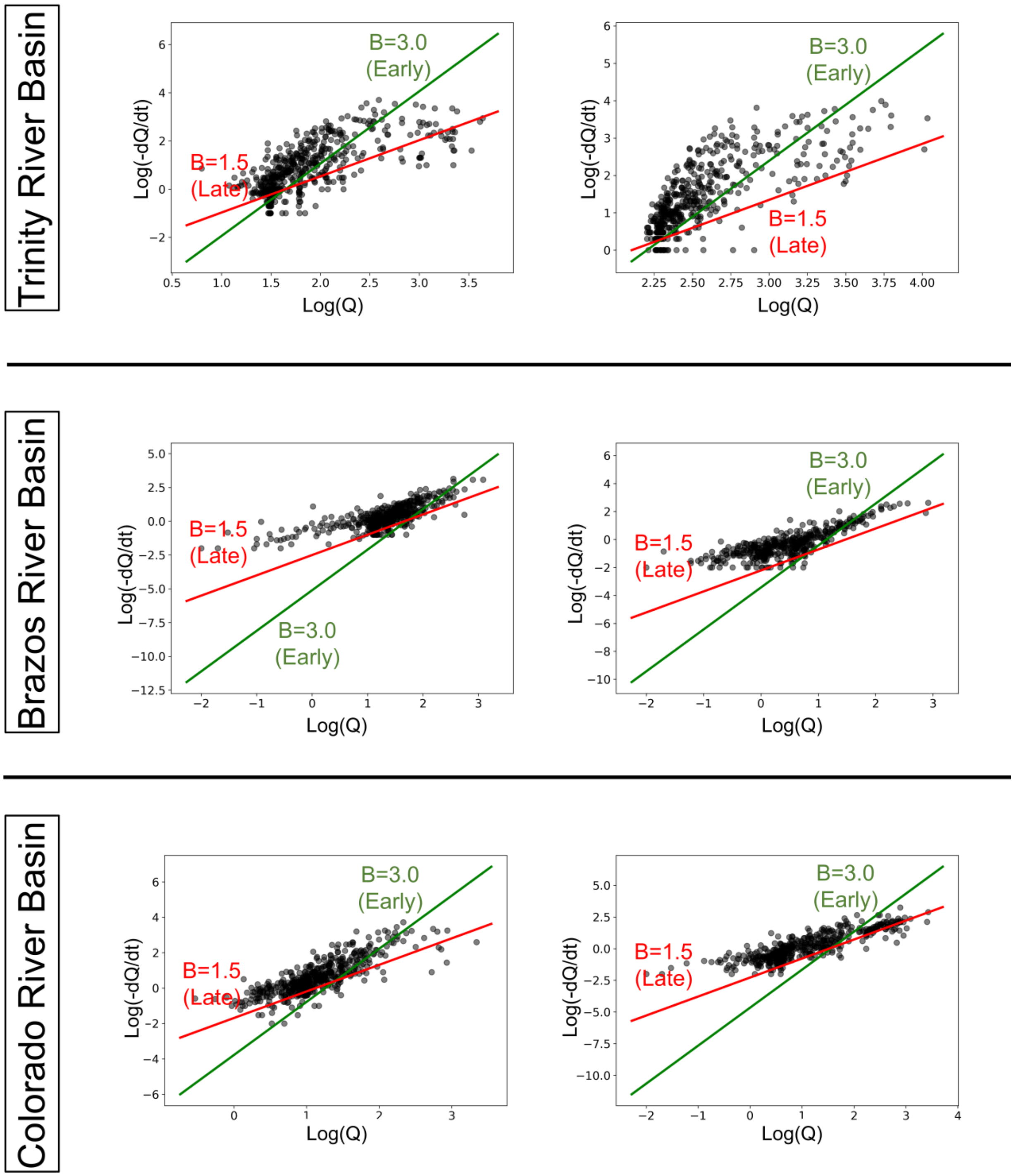


Figure 3. Recession Analysis Plots in Bi-logarithmic Space for the Two Selected Gauges in Each Studied Basin. The Examples Illustrate How the Recession Parameters *a* and *b* were Estimated from Point Cloud data Using Lower Envelope (LE) Method. The Recession Parameters *a* and *b* Determined for the Respective Early and Late Time Domains were Used for Inferring Effective Aquifer Properties on Catchment-Scale.

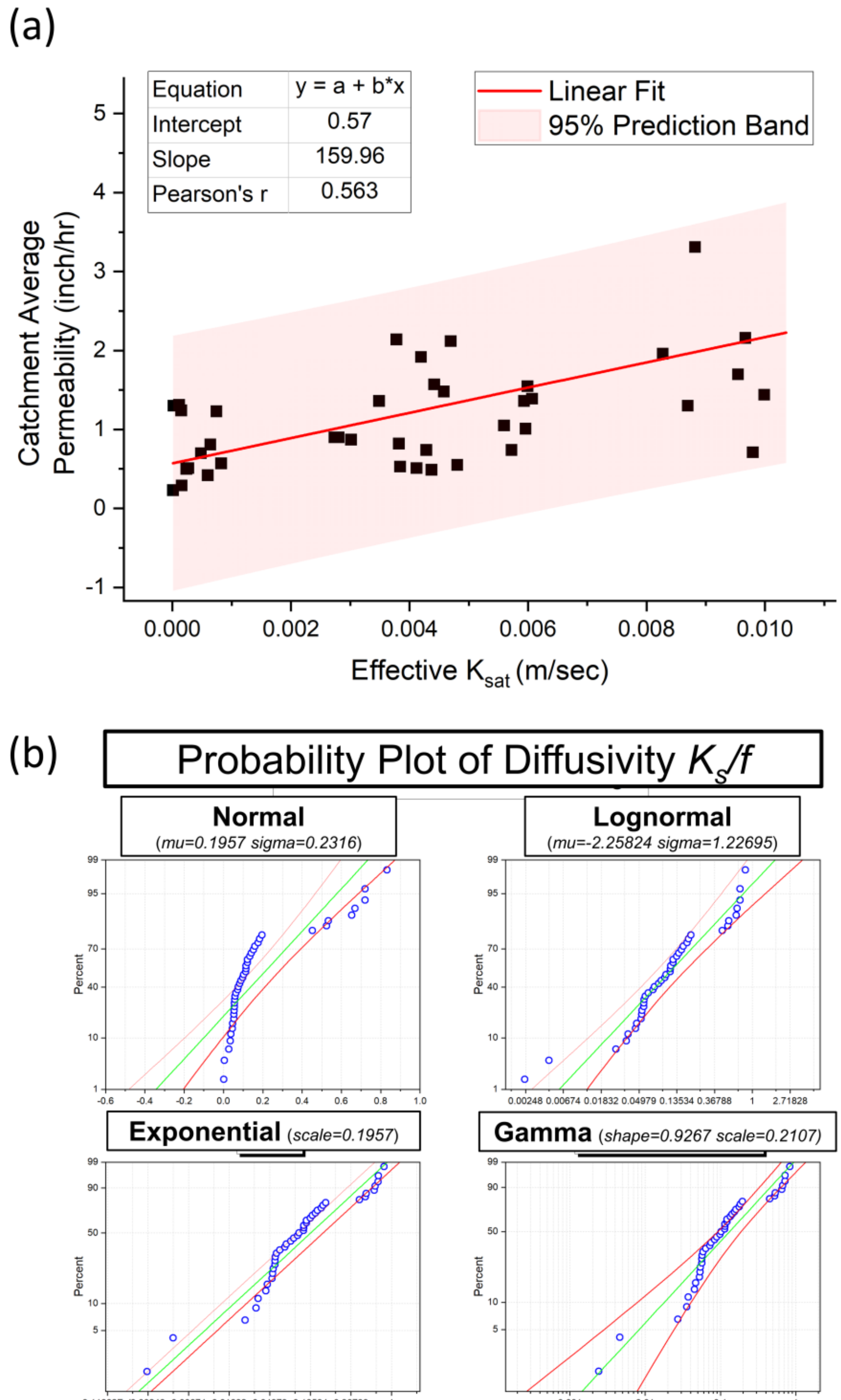


Figure 4. (a) The Identified Linear Relationship Between Catchment Average Permeability derived from STATSGO2 dataset and the Catchment-Scale Effective Horizontal Saturated Hydraulic Conductivity (K_s) (Estimated From Recession Analysis), (b) Probability Plots of Diffusivity K_s/f. The Lognormal Distribution was Considered As the Most Appropriate Probabilistic Distribution of K_s/f Derived From the 40 Selected Catchments/Gauges.

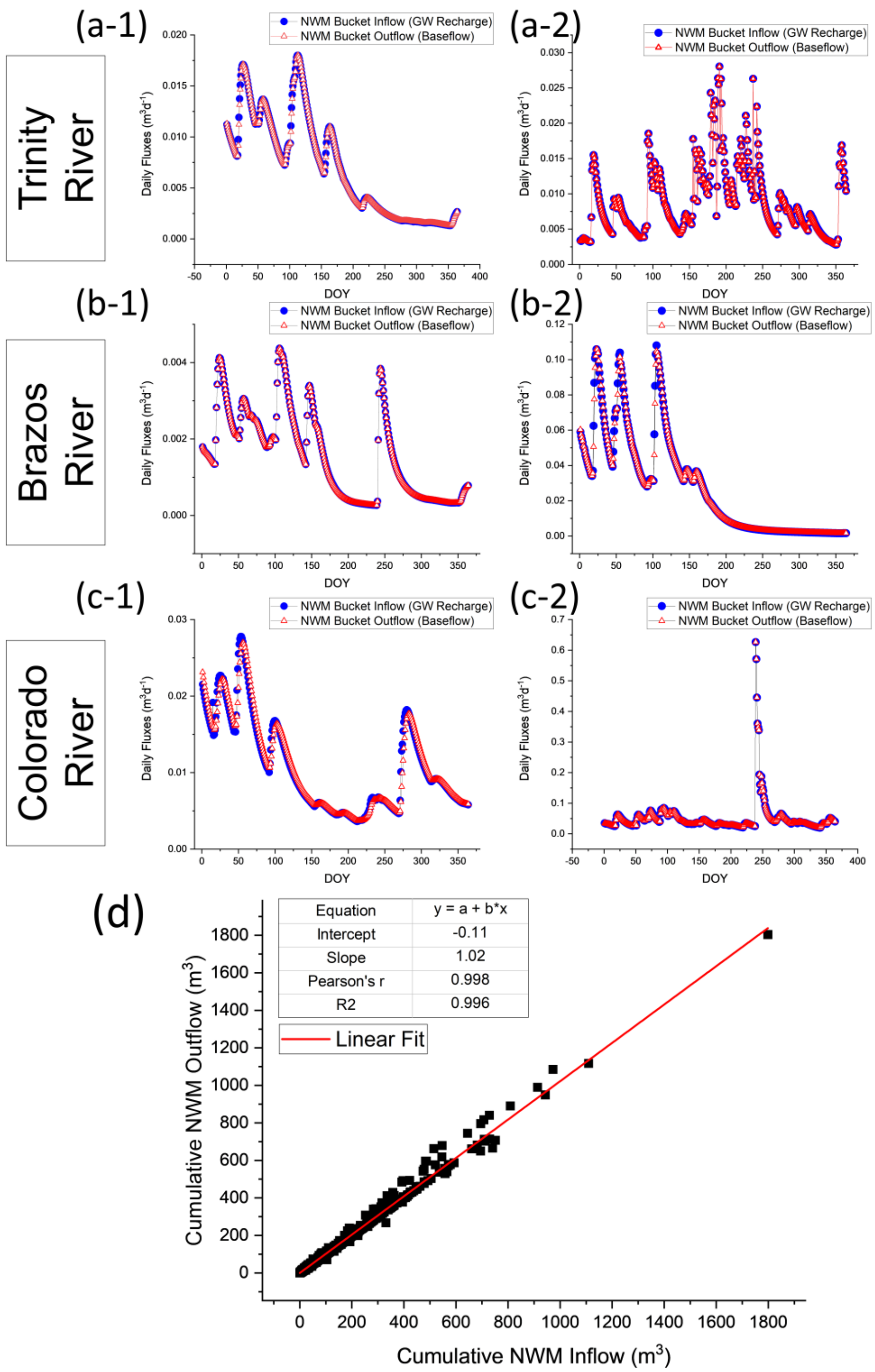


Figure 5. Almost No Time Lags were Found Between the Soil Bottom Drainages Fluxes (Modeled by the Noah-MP LSM) and Baseflow Fluxes in the Current NLSR Configuration From the Selected 40 Catchments (a-1, a-2, b-1, b-2, c-1, c-2), (d) The (Almost) Identical Estimates Between Cumulative Soil Drainage Fluxes and Baseflow Fluxes Exhibit the Lack of Groundwater Storage in the NLSR(s).

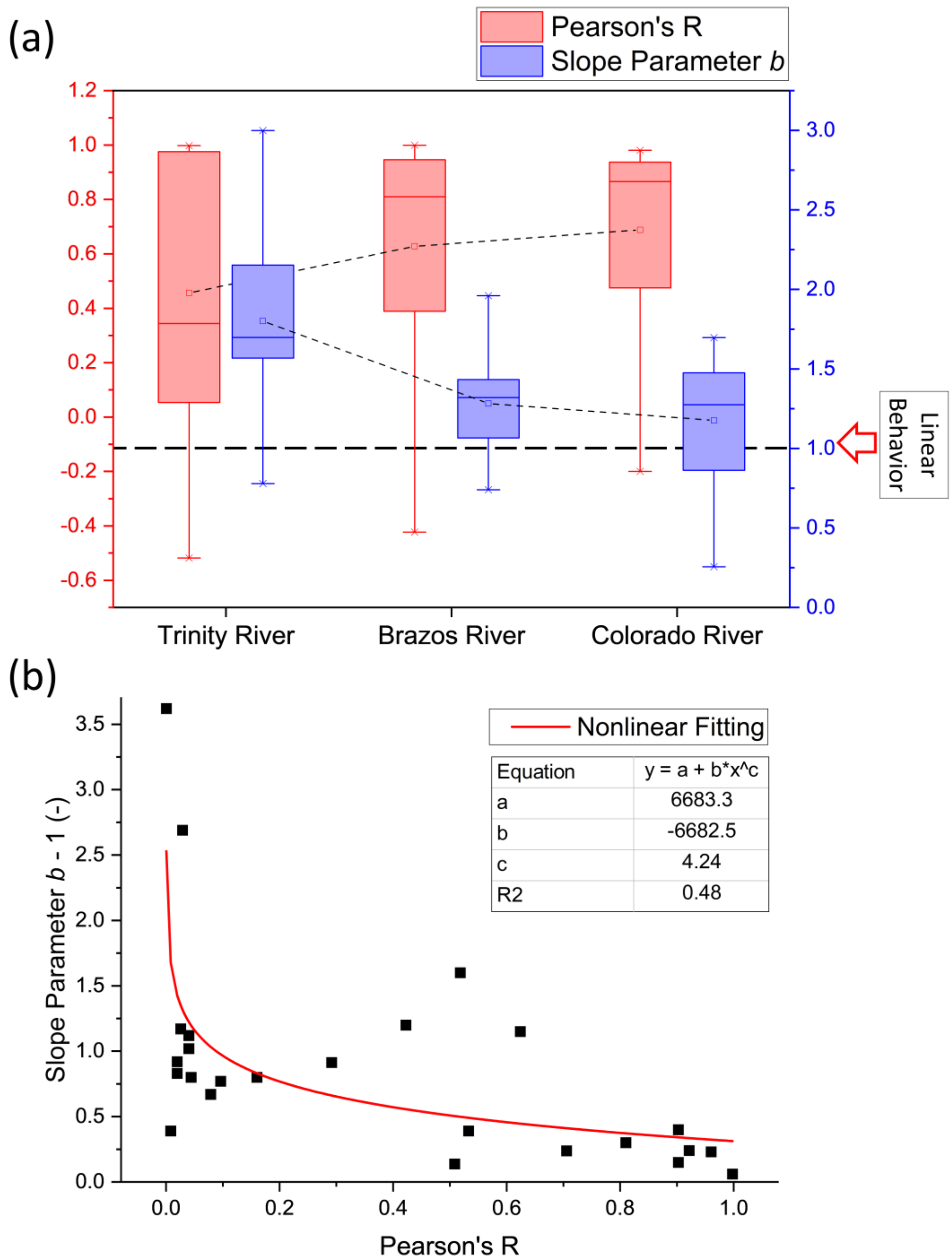


Figure 6. (a) The Temporal Agreement Between the Respective Baseflow Estimates from the NWM and the NWM-BE3S was Found to be Related to the Linearity of the Streamflow Recessions Observed in Corresponding Basin. Higher Agreements Between B_f and f_{pr} were Identified as the Basin Functions More Like a Linear Reservoir, (b) The Relationship Between the Temporal Correspondence Between B_f and f_{pr} (i.e., Pearson's R) and Recession Non-linearity was Also Captured by a Non-linear Fitting (Power).

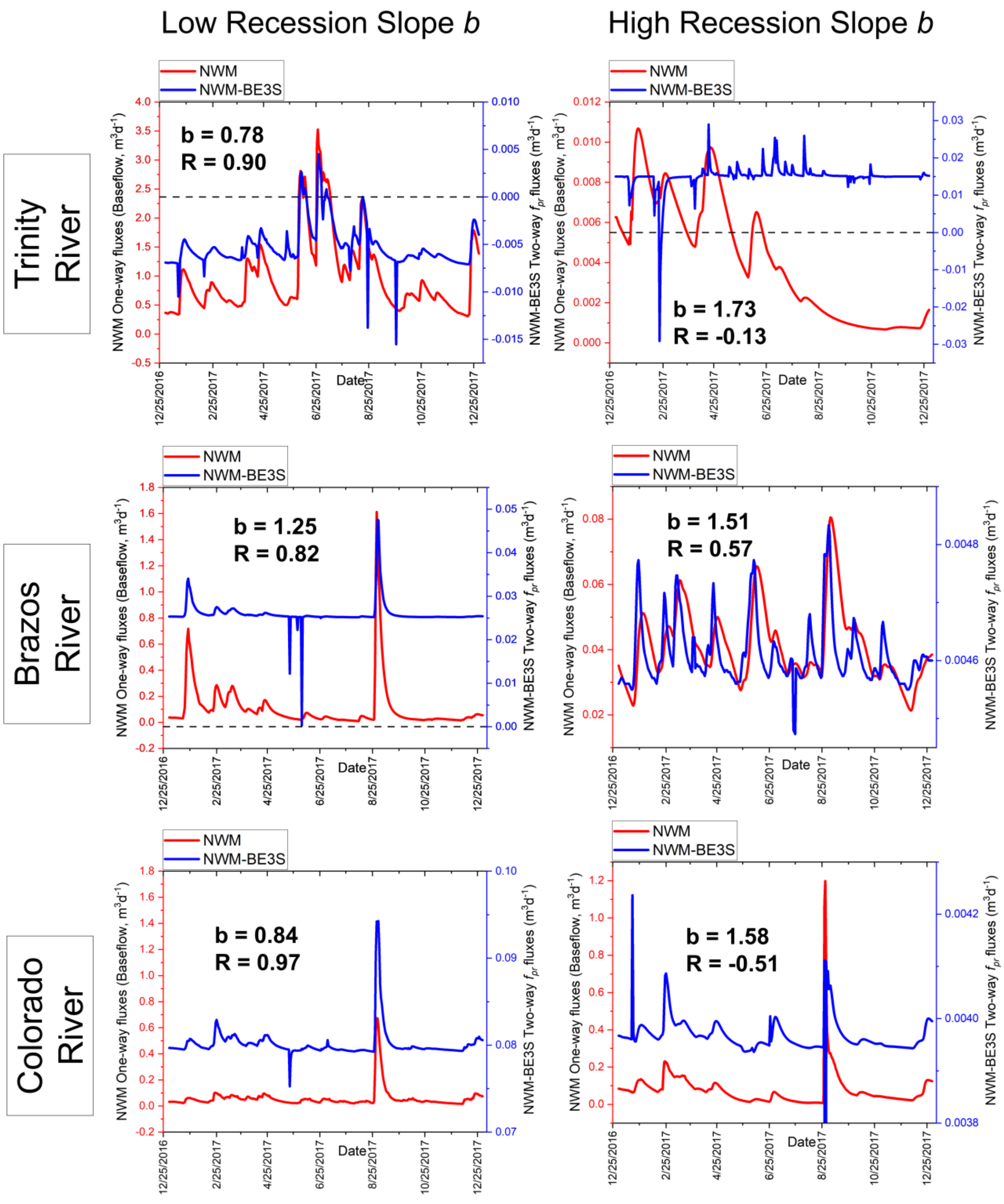


Figure 7. Comparative Time Series of the B_f and f_{pr} Estimates for Two Catchments Showing Contrasting Recession Non-linearity (i.e., Low or High) in the Streamflow Observations. The Modeled Outflow Fluxes From the Boussinesq Aquifer (f_{pr}) were Identified Able to Better Represent the Baseflow Recessions with High Non-linearity than the NLSR model.

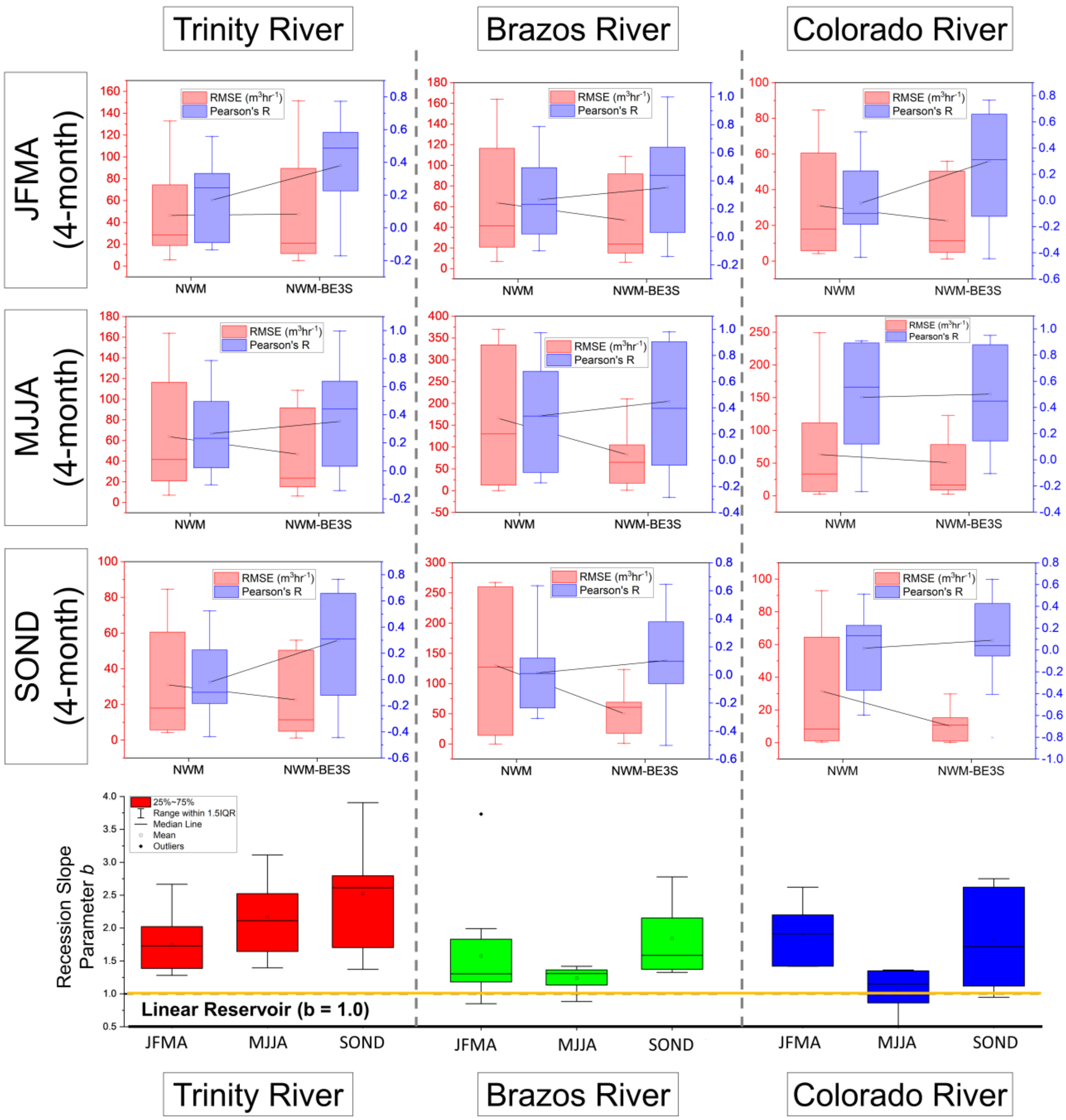


Figure 8. The Improvements in the Streamflow Predictions Found in the NWM-BE3S model. While Significant Improvements in *R* and Reductions in *RMSE* From All the Four-month Interval Periods (i.e., JFMA, MJJA, SOND), More Pronounced Improvements in *R* and *RMSE* were Found As the Corresponding Basin Exhibited Low Non-linearities (i.e., Linearity, Parameter *b* Close to 1.0). The Comparison of the Average and Value Distributions of *R* and *RMSE* between the NWM and the NWM-BE3S Manifests the Technical Enhancement of Model Structure.

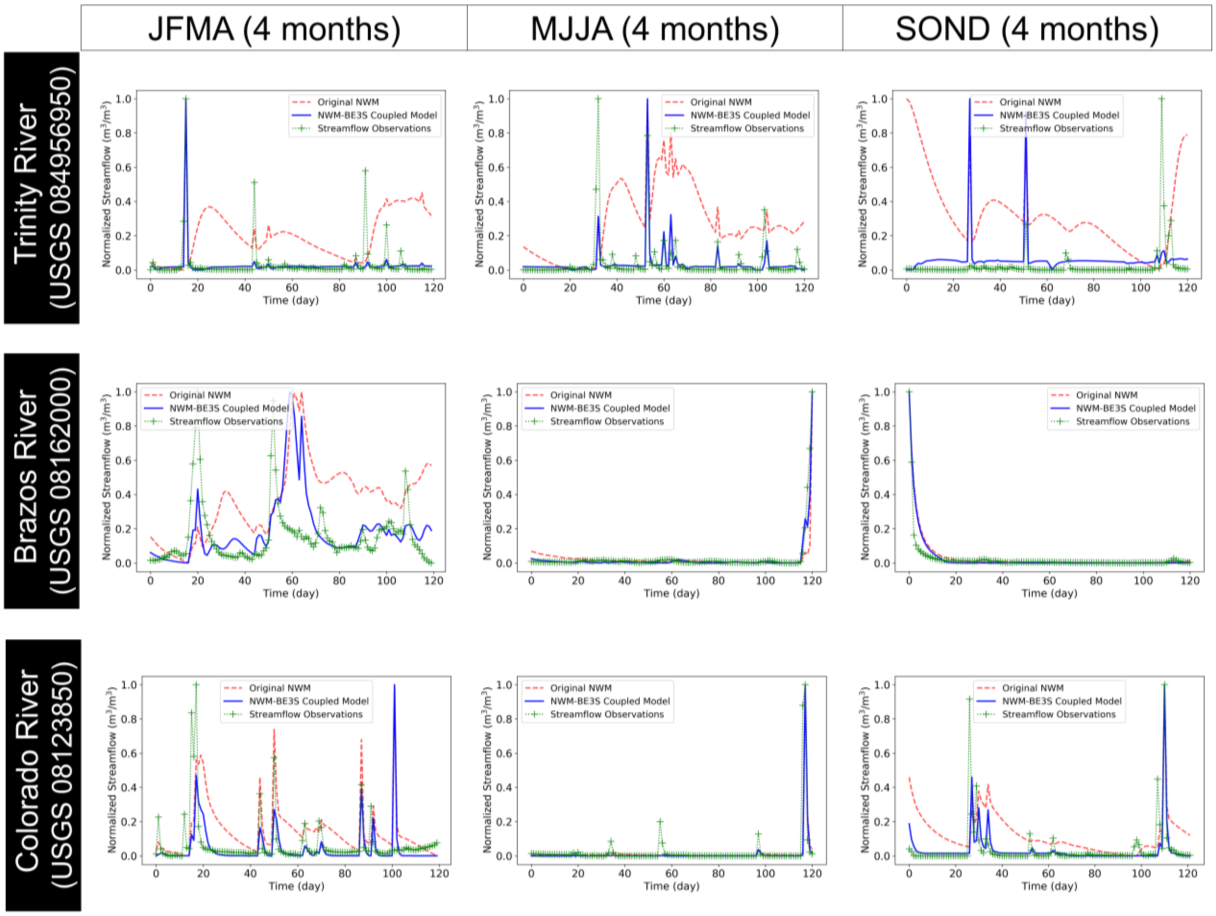


Figure 9. Comparative Time Series of Normalized Streamflow in Each Four-month Interval Period and Evaluation against the Corresponding Streamflow Observations. As Expected from the Improved Metrics *R* and *RMSE* in the NWM-BE3S, the Recession Durations and the Timings of Peak Discharge were Better Predicted in the NWM-BE3S model Compared to the Original NWM.

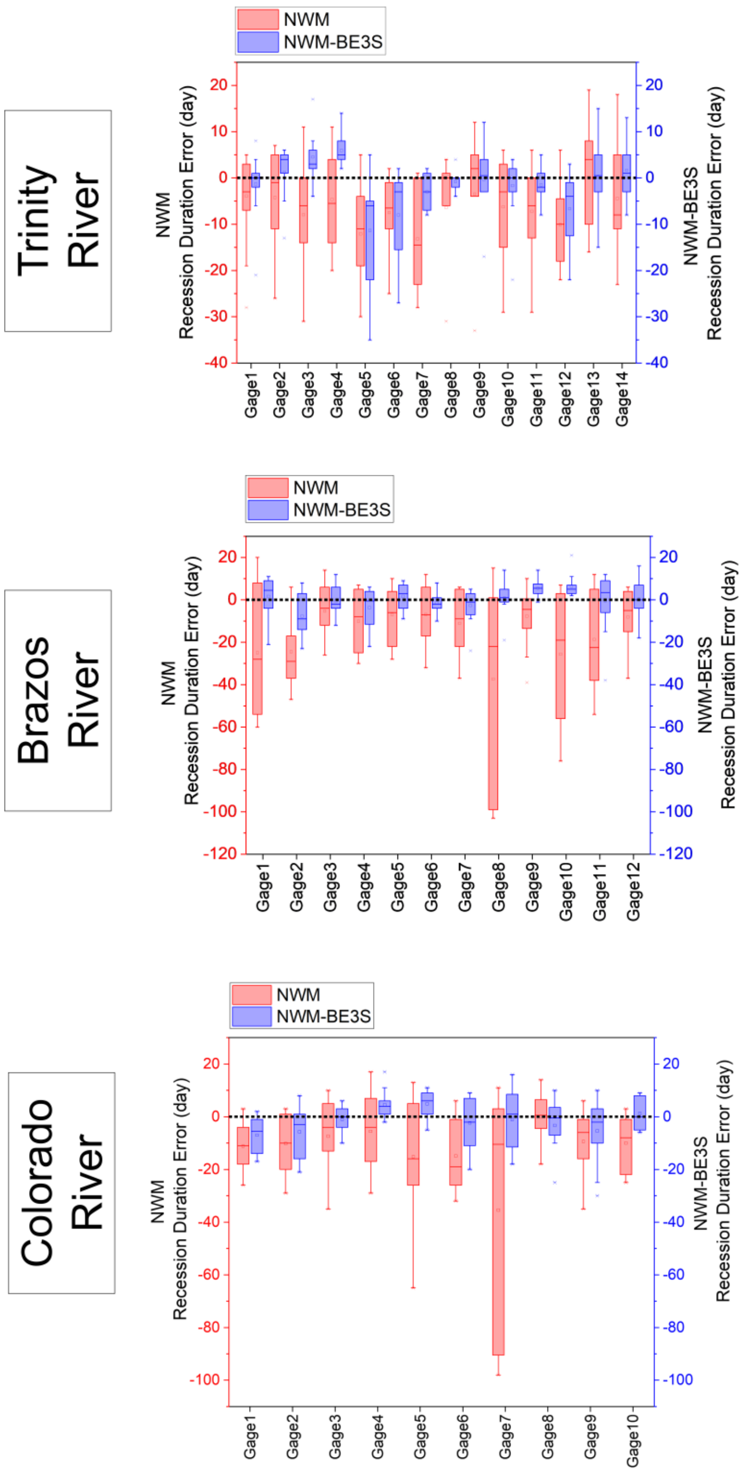


Figure 10. Comparative Evaluation of Recession Duration Errors in the Studied Catchments. The NWM Identified That the Modeled Recession Duration were generally longer than the (Corresponding) Actual Duration, and the Significant Reductions in the Duration Error were Identified in the NWM-BE3S.

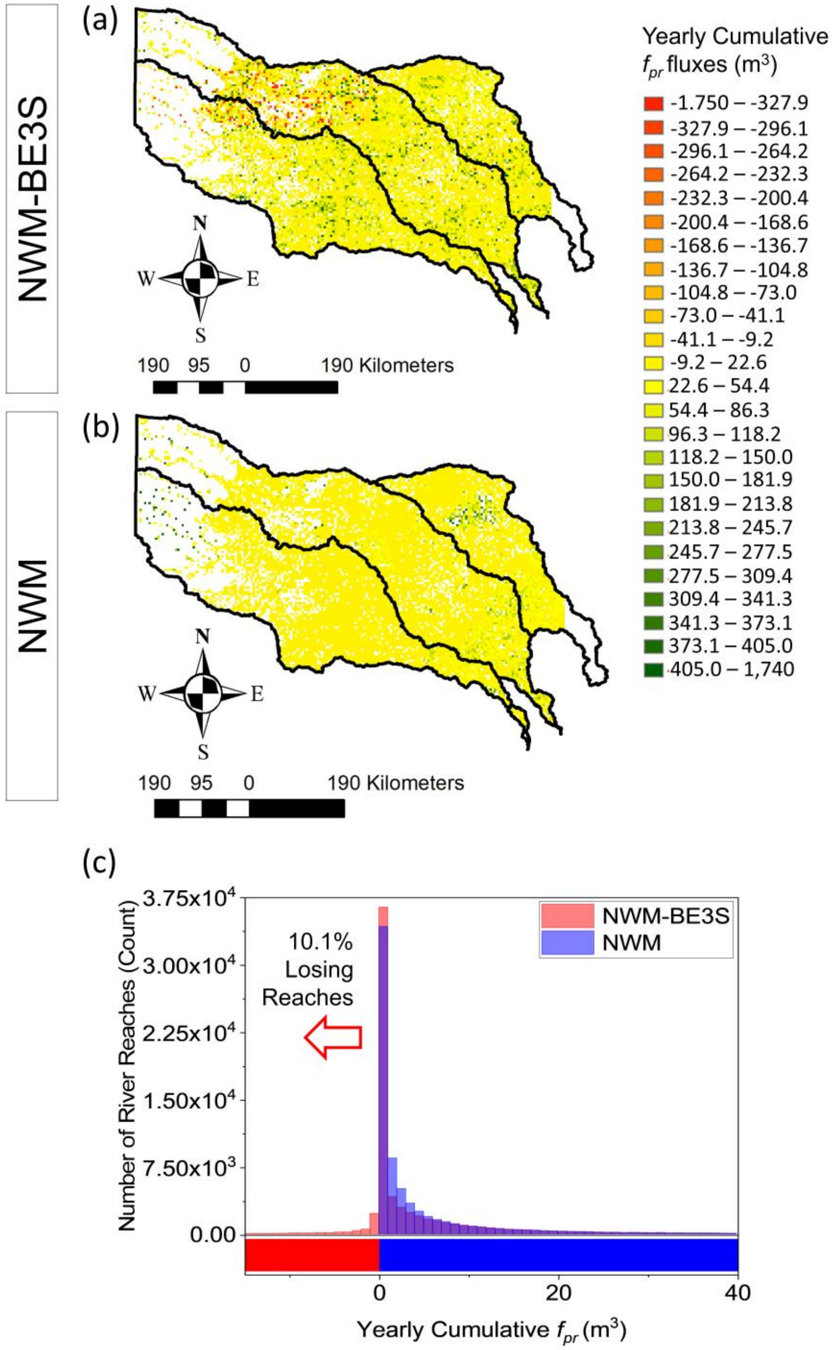


Figure 11. Comparison Between (a) and (b) Shows: (1) Structural Limitation of the Current NLSR model in the NWM that Cannot Depict Losing Conditions in the River Reaches, (2) The Dominant Processes Between the River and the Aquifer were Spatially Heterogeneous Roughly Following the Climatic Gradients over the Study Area. (c) About 10.1 % of the Total River Reaches were Diagnosed as Losing Reaches While Showing the New Predictive Capabilities in the NWM-BE3S to Represent Both River States (i.e., Either Gaining or Losing) Based on the Bidirectional Lateral Hydraulic Connections Between the Stream and the Aquifer.