Water Resources Research

RESEARCH ARTICLE
10.1029/2024WR037821

Quantifying Dynamic Linkages Between Precipitation, Groundwater Recharge, and Streamflow Using Ensemble Rainfall-Runoff Analysis

Huibin Gao[1,2], Qin Ju[1,3], Dawei Zhang[4], Zhenlong Wang[5], Zhenchun Hao[1,3], and James W. Kirchner[2,6,7]

[1]National Key Laboratory of Water Disaster Prevention, Hohai University, Nanjing, China, [2]Department of Environmental Systems Science, ETH Zurich, Zurich, Switzerland, [3]China Meteorological Administration Hydro-Meteorology Key Laboratory, Nanjing, China, [4]China Institute of Water Resources and Hydropower Research, Beijing, China, [5]Anhui and Huaihe River Institute of Hydraulic Research, Bengbu, China, [6]Swiss Federal Research Institute WSL, Birmensdorf, Switzerland, [7]Department of Earth and Planetary Science, University of California, Berkeley, CA, USA

Key Points:

- Nonlinear and nonstationary behaviors of streamflow and groundwater recharge are quantified using Ensemble Rainfall-Runoff Analysis (ERRA)
- Subsurface processes, rather than overland flow, generate episodic streamflow in an intensively farmed catchment
- ERRA can help evaluate model realism by comparing the characteristic impulse responses of models with those of real-world systems

Correspondence to:
J. W. Kirchner and Q. Ju,
kirchner@env.ethz.ch;
juqin@hhu.edu.cn

Citation:
Gao, H., Ju, Q., Zhang, D., Wang, Z., Hao, Z., & Kirchner, J. W. (2025). Quantifying dynamic linkages between precipitation, groundwater recharge, and streamflow using ensemble rainfall-runoff analysis. *Water Resources Research*, *61*, e2024WR037821. https://doi.org/10.1029/2024WR037821

Received 24 APR 2024
Accepted 9 NOV 2024

Author Contributions:
Conceptualization: Huibin Gao, James W. Kirchner
Data curation: Huibin Gao
Formal analysis: Huibin Gao, James W. Kirchner
Funding acquisition: Qin Ju, Zhenchun Hao
Investigation: Huibin Gao, Qin Ju, Dawei Zhang, Zhenlong Wang, Zhenchun Hao
Methodology: Huibin Gao, James W. Kirchner
Resources: Qin Ju, Dawei Zhang, Zhenlong Wang, Zhenchun Hao
Software: Huibin Gao, James W. Kirchner
Supervision: Zhenchun Hao, James W. Kirchner
Validation: Huibin Gao, James W. Kirchner
Visualization: Huibin Gao

Abstract Understanding streamflow generation at the catchment scale requires quantifying how different components of the system are linked, and how they respond to meteorological forcing. Here we present a proof-of-concept study characterizing and quantifying dynamic linkages between precipitation, groundwater recharge, and streamflow using a data-driven nonlinear deconvolution and demixing approach, Ensemble Rainfall-Runoff Analysis (ERRA). Streamflow in our mesoscale, intensively farmed test catchment is flashy, but occurs at time lags that are too long to be plausibly attributed to overland flow. Instead, ERRA's estimates of the impulse responses of groundwater recharge to precipitation, and of streamflow to groundwater recharge, imply that this intermittent streamflow is primarily driven by precipitation infiltrating to recharge groundwater, followed by discharge of groundwater to streamflow. ERRA reveals that streamflow increases nonlinearly with increasing precipitation intensity or groundwater recharge, and exhibits almost no response to precipitation or recharge rates of less than 10 mm d^{-1}. Groundwater recharge is both nonlinear, increasing more-than-proportionally with precipitation intensity, and nonstationary, increasing with antecedent wetness. Simulations with the infiltration model Hydrus-1D can reproduce the observed water table time series reasonably well (NSE = 0.70). However, ERRA shows that the model's impulse response is inconsistent with the real-world impulse response estimated from measured precipitation and groundwater recharge, illustrating that conventional goodness-of-fit statistics can be weak tests of model realism. Thus, our proof-of-concept study demonstrates how impulse responses estimated by ERRA can help clarify linkages between precipitation and streamflow at the catchment scale, quantify nonlinearity and nonstationarity in hydrologic processes, and critically evaluate simulation models.

1. Introduction

Understanding hydrologic response at the catchment scale is a longstanding problem, and one with important practical implications. Diverse methods have been used to quantify responses of various hydrologic cycle components to forcing, such as groundwater response to rainfall (e.g., Gootman & Hubbart, 2021; Russo & Lall, 2017; Tashie et al., 2016), soil moisture response to rainfall (e.g., Kim, 2009; N. K. Singh et al., 2021), and particularly runoff response to rainfall, which is influenced by diverse factors including antecedent conditions (Nippgen et al., 2016; Pathiraja et al., 2012), topography (Appels et al., 2016; L. Chen et al., 2013), and canopy interception (Oda et al., 2021; Tao et al., 2020).

Many approaches to exploring rainfall-runoff response have benefited in recent years from advances in computer technology and data availability, but still face a range of challenges. Physically based models have been extensively used for predicting future changes and ungauged catchments, but can suffer from over-parameterization, calibration uncertainties, questionable upscaling of governing equations from the microscale to the macroscale, and difficulty in transferring parameters from one catchment to another (Beven & Freer, 2001; Beven & Smith, 2015; Dooge, 1986; Kirchner, 2009; Wagener & Gupta, 2005). Empirical methods, such as unit hydrographs, are often strongly founded in mathematics and statistics, but typically assume a linear and stationary system (Dooge, 1959; James & Johanson, 1999) whereas real-world hydrological systems are often nonlinear and nonstationary (Singh & Woolhiser, 2002). "Black box" techniques like machine learning and deep learning can yield impressive predictions, given sufficient training data, and as long as they do not extrapolate beyond those

Writing – original draft: Huibin Gao
Writing – review & editing: Huibin Gao, Qin Ju, Dawei Zhang, Zhenlong Wang, Zhenchun Hao, James W. Kirchner

training conditions (Best et al., 2015; Chen et al., 2020; Kratzert et al., 2019), but inferring physical processes from them remains difficult (although combining process understanding with machine learning is an active area of research; Nearing et al., 2021; Pelissier et al., 2020).

Recently, Kirchner (2022) developed methods for estimating impulse response functions for nonlinear, nonstationary, and heterogeneous systems using a combination of nonlinear nonparametric deconvolution and demixing techniques. A derivative of these methods, Ensemble Rainfall-Runoff Analysis (ERRA), enables data-driven, model-independent estimation of the impulse response of hydrologic systems (Kirchner, 2024a). The approach facilitates estimating timescales and magnitudes of hydrologic response to multiple overlapping inputs, and quantifying the nonlinearity and nonstationarity of hydrologic response directly from data, without the need to postulate, or calibrate, a process model.

Here we apply ERRA to a mesoscale intensively farmed catchment, with the goals of (a) characterizing how groundwater recharge and streamflow respond to precipitation, and how streamflow responds to groundwater recharge, over time; (b) quantifying these relationships' nonlinear dependence on precipitation intensity and their nonstationary dependence on antecedent wetness conditions; and (c) exploring whether comparisons between response distributions derived from real-world data and those derived from model simulations can provide useful tests of model behavior, using Hydrus-1D as an example. Our case study in an agricultural plain with shallow groundwater depth shows that ERRA can quantify magnitudes and timescales of groundwater recharge and runoff response, and that these relationships imply that streamflow at this site is primarily generated by infiltration of precipitation to groundwater, followed by linear-reservoir drainage of the groundwater system. Our analysis also illustrates how ERRA can be used to diagnose model limitations that are not apparent from conventional goodness-of-fit statistics.

2. Data and Methods

2.1. Study Site and Data

Our study site is the 265 km^2 Yanglou catchment (Figure 1a) in the Huaibei Plain, an alluvial plain in eastern China. Average precipitation is ∼782 mm yr^{-1}, and the average annual air temperature is ∼15°C (2004−2018). Meteorological conditions vary seasonally due to the monsoon climate, with ∼70% of precipitation falling between June and September, but they are spatially rather uniform due to the very flat catchment topography (∼21 m total relief).

Catchment soils are developed from unconsolidated Quaternary deposits (200−300 m deep) of the ancestral Yellow River and its associated lakes. Their specific yield varies between 0.035 and 0.055 in the upper 5 m, and their bulk density varies between 1.397 and 1.483 g cm^{-3} in the upper 2 m. Near-surface soils are highly permeable; hydraulic conductivities at W9 (Figure 1a) are 22.2 mm hr^{-1} at 0−30 cm depth, 19.6 mm hr^{-1} at 30−100 cm depth, and 18.5 mm hr^{-1} at 100−200 cm depth (Zhu, 2013). Land use in the catchment is predominantly agricultural, and shallow groundwater is the major supply for irrigation and rural residential use. Groundwater pumping records are not available for the Yanglou catchment, but annual groundwater withdrawals in a 233 km^2 district that partly overlaps with the catchment averaged 52 mm yr^{-1} in 2005–2006. Streamflow is intermittent, flowing primarily during the summer monsoon, and no streamflow is observed at the outlet over 90% of the time.

Precipitation and groundwater elevations (relative to the Yellow Sea datum) were measured daily from 1 January 2004 to 31 December 2018 at 6 rain gauges and 9 monitoring wells (average well depth 20 m; range 8–28 m) distributed throughout the catchment (Figure 1a). Streamflow over the same period was measured only at the outlet (W9), and averaged 32.3 mm yr^{-1}, or only 4% of average precipitation. Non-zero streamflow occurred during only 7.6% of the study period. Catchment average precipitation was calculated by the Thiessen method; other meteorological data including wind speed, sun hours, etc. were arithmetically averaged from the two nearest weather stations, Xuzhou and Dangshan (see inset in Figure 1a), outside the catchment.

The water table depth (WTD) of each well was calculated as the difference between the ground surface elevation and water table elevation (Figure 1b), and then arithmetically averaged each day to obtain the catchment-averaged WTD. Average water table depth during the study period was 2.5 m, ranging from 1.3 to 5.1 m among the 9 wells. Water table data are missing at well W4 for 151 days in the 5479-day study period. For those days, the remaining 8

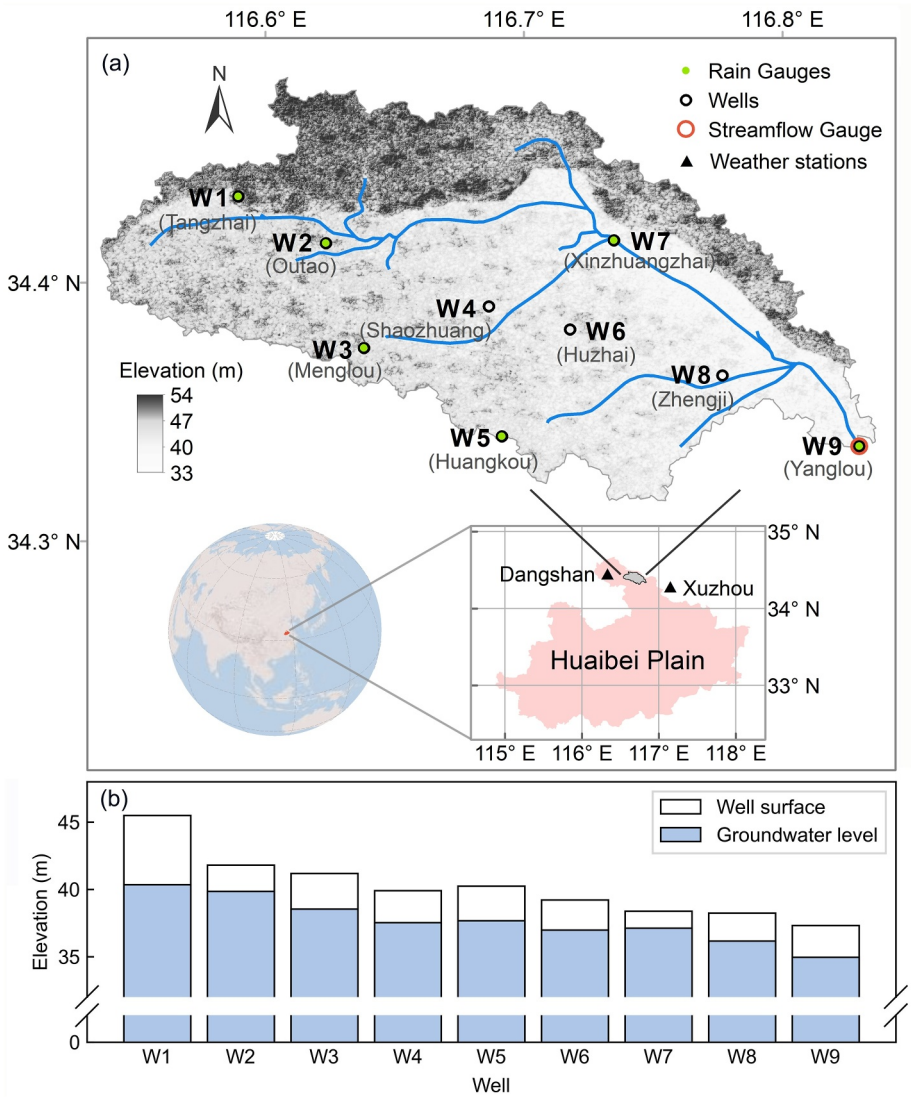

Figure 1. Study site description. (a) Location of Yanglou catchment, rain gauges and wells inside the catchment, and weather stations outside the catchment. (b) The elevation of the ground surface and mean groundwater level (2004−2018) for each well. Water table depth was calculated as the elevation difference. Elevation in (a) is relative to the SRTM DEM datum; elevation in (b) is relative to the Yellow Sea datum.

wells were averaged; the missing data have negligible influence on the catchment-averaged WTD because well W4 typically has values close to the average of the other wells.

Groundwater recharge (GR) was calculated by multiplying a mean specific yield of 0.045 by the rate of decrease of WTD. Most increases in WTD (i.e., declines in groundwater levels) are due to groundwater pumping; to minimize the effects of pumping on our analysis we set GR to 0 on days with groundwater declines (when GR would otherwise be negative). Our approach implicitly assumes that when water tables are rising (and GR is positive), the effects of pumping are small relative to the effects of infiltration. This assumption is supported by the observation that water table rises are briefer, and much steeper, than water table declines.

2.2. Ensemble Rainfall-Runoff Analysis

Here, we analyze the coupling between precipitation, groundwater recharge, and streamflow, using Ensemble Rainfall-Runoff Analysis (ERRA (Kirchner, 2024a)). Based on recently developed methods for quantifying the impulse response of nonlinear, nonstationary, and heterogeneous systems (Kirchner, 2022), ERRA is a data-

driven, model-independent approach to quantifying hydrological response at the catchment scale. In its simplest form, ERRA is conceptually similar to unit hydrograph approaches that have been used in hydrology for decades, but conventional unit hydrographs assume linearity (runoff response is proportional to effective precipitation) and stationarity (runoff response to a given amount of rainfall is identical, regardless of when it falls). ERRA relaxes both of these assumptions through a combination of nonlinear deconvolution and demixing methods. Not only can this approach quantify the impulse response of streamflow to precipitation, it can also quantify how this impulse response changes with rainfall intensity (nonlinearity) and how it varies with catchment wetness (nonstationarity), even when these signals are all overprinted on one another at the catchment outlet. For detailed documentation, benchmark tests, and proof-of-concept demonstrations, the reader is referred to Kirchner (2022, 2024a).

Two main outputs of ERRA are presented here. The first is the Runoff Response Distribution (RRD), which quantifies the impulse response of the system to one or more inputs as a function of lag time. The RRD is analogous to a conventional impulse response function and, as the name implies, quantifies how the system's response is distributed over time, per unit of input. Thus the RRD has dimensions of time^{-1}, if the input and response are measured in the same units. If the system were linear and time-invariant (which is unlikely in real-world hydrological processes), the RRD would be a complete description of its behavior: when convolved with the input time series, it would reproduce the output time series. In the more likely case of a system that is linear but time-variant (i.e., nonstationary), the RRD represents the ensemble average response of the system to the input time series.

The second main output of ERRA, the Nonlinear Response Function (NRF), quantifies the impulse response of nonlinear systems. The NRF replaces the lag coefficients of the RRD with broken-stick piecewise-linear functions of input intensity. Thus it allows for the possibility that the impulse response is nonlinear in different ways at different lag times. Rather than expressing the system response per unit of input (as in the RRD), the NRF expresses the system response to one time step of input at a given intensity (or over a given range of intensities); framed in this way, it can be measured in the same units as the system output. To characterize the impulse response of nonlinear systems, ERRA provides response curves for user-defined ranges of input intensity (i.e., the NRF); it also provides tables showing how the peak impulse response, and the total runoff volume, change as functions of input intensity.

3. Results

3.1. Streamflow and Groundwater Recharge Response to Unit Precipitation Input

Figure 2 presents time series of precipitation (P), water table depth (WTD), groundwater recharge (GR), and streamflow, with days of non-zero flow highlighted using darker symbols. Streamflow is associated with periods of rapid groundwater rise instead of periods with high groundwater levels, and during runoff periods, rapid groundwater recharge continues beneath an unsaturated vadose zone. These observations, combined with the high permeability of the soils, suggest that infiltration-excess and saturation-excess overland flow are not widespread at this site. Streamflow is not a simple function of catchment storage; runoff peaks are not synchronized with peak groundwater levels, but instead follow shortly after days with rapid groundwater recharge, when water tables are rapidly rising (Figures 2b and 2c). Streamflow recession is also faster than groundwater recession, further supporting the inference that streamflow is not directly linked to groundwater levels.

The impulse response of the coupled groundwater-surface water system can be summarized using response distributions calculated from ERRA (Figure 3). The runoff response distribution calculated with precipitation as its input (RRD_P; Figure 3a) expresses how the runoff response to one unit of precipitation is distributed over time. Figure 3a shows that the runoff response lasts about 12 days and peaks approximately 2 days after precipitation falls (the same-day runoff response corresponds to the first point in the RRD_P).

To quantify the coupling between precipitation and groundwater recharge that is visually evident in Figure 2, we repeated the ERRA calculations that we used to estimate the RRD_P, but this time with groundwater recharge (GR, Figure 2) as the output (this is possible because groundwater recharge is expressed in the same units as precipitation and streamflow). This yields the groundwater recharge response distribution driven by precipitation ($GRRD_P$; Figure 3b), which expresses how groundwater recharge responds, over time, to one unit of precipitation. As Figure 3b shows, groundwater recharge is quick and brief; the $GRRD_P$ peaks one day after precipitation falls and then rapidly declines, becoming nearly zero within about 6–10 days. Compared to the runoff response to

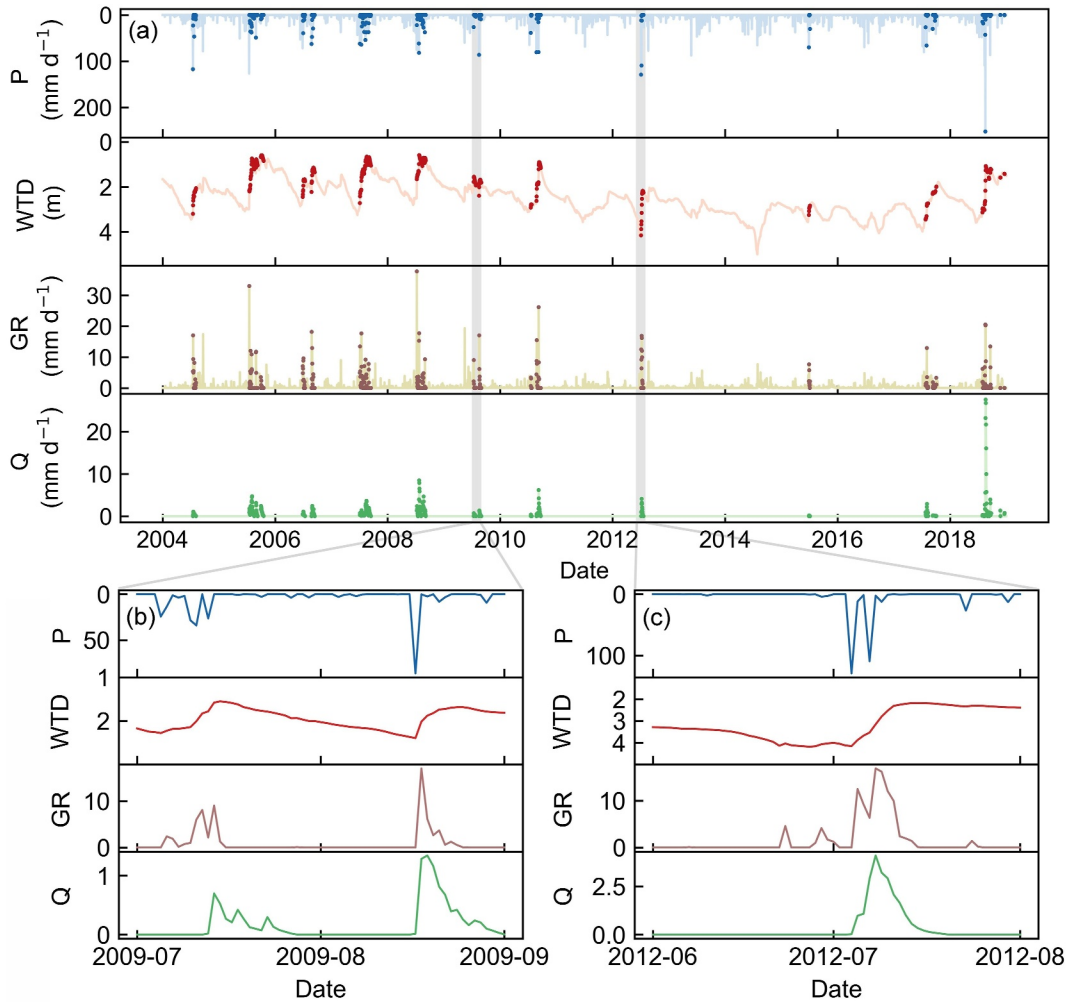

Figure 2. Overview of measured time series of precipitation (P), water table depth (WTD), groundwater recharge (GR) and streamflow (Q) during the study period 2004–2018 (a); darker dots show days with non-zero flow (accounting for 7.6% of the whole time series). Panels (b) and (c) show selected 2-month periods to more clearly illustrate the temporal dynamics of several precipitation events. Streamflow was closely linked to groundwater recharge (GR), but peaked earlier and receded faster than the water table itself (WTD).

precipitation (RRD_P), the peak in the groundwater recharge response distribution ($GRRD_P$) is nearly 5 times higher (0.12 vs. 0.025), and 3 times narrower (width at half maximum of 1.6 vs. 4.8 days). The integral under the $GRRD_P$ curve (0.214 ± 0.007, estimate ± standard error) is about 1.7 times bigger than the integral under the RRD_P curve (i.e., runoff coefficient, 0.125 ± 0.004), implying that precipitation triggers more groundwater recharge than streamflow over the 20-day time scales shown in Figure 3.

In Figure 2, one can see a close association between peaks in groundwater recharge and peaks in streamflow shortly thereafter. To quantify this coupling between groundwater recharge and streamflow, we repeated the same ERRA calculations that we used to estimate the RRD_P, but this time with groundwater recharge rates as the input. This approach recognizes that groundwater recharge is the input to the saturated zone, just as precipitation is the input to the ground surface; it then estimates the coupling between this input and streamflow. The resulting runoff response distribution driven by groundwater recharge, or RRD_{GR} (Figure 3c), quantifies the groundwater system's impulse response as it transmits recharge signals to streamflow. As Figure 3c shows, the RRD_{GR} is very nearly exponential; this is consistent with conventional "linear reservoir" assumptions in groundwater systems. It is also consistent with the geological setting in the catchment: a flat landscape with deep unconsolidated deposits having approximately constant thickness and presumably little vertical permeability gradient. The exponential decay constant of 0.38 d^{-1} in Figure 3c implies a characteristic response time of 2.6 days; that is, groundwater-driven

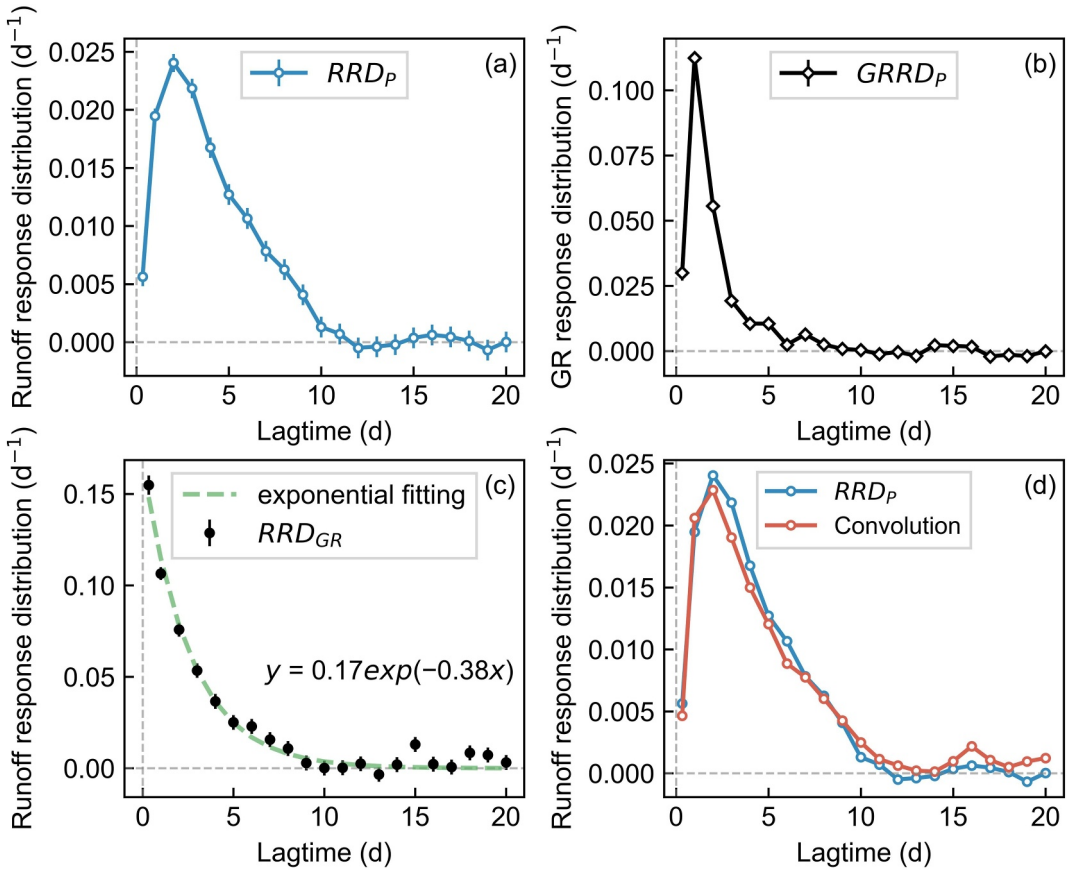

Figure 3. Response distributions to unit input. (a) Runoff response distribution driven by precipitation (RRD_P). The runoff response peaks (at 0.025 d^{-1}, or 2.5% of precipitation per day) on the second day after precipitation input. (b) Groundwater recharge response distribution driven by precipitation ($GRRD_P$). The groundwater recharge response to precipitation was quicker and larger than the runoff response to precipitation, peaking at 0.118 d^{-1}, or 11.8% of precipitation per day. (c) Runoff response distribution driven by groundwater recharge (RRD_{GR}). The response of streamflow to GR closely approximates an exponential function, consistent with linear reservoir behavior in the groundwater system. (d) Good match between RRD_P (blue) and the convolution $GRRD_P \otimes RRD_{GR}$ (red), demonstrating that groundwater recharge is a key component of the causal chain linking precipitation with streamflow.

streamflow can be expected to decline by a factor of e in 2.6 days, all else equal. The integral under the RRD_{GR} curve is 0.445 ± 0.018, implying that about half of groundwater recharge is eventually reflected in streamflow during the 20-day interval plotted in Figure 3c. The rest presumably is removed through pumping, through uptake by deep-rooted vegetation, or through infiltration into deeper aquifers.

The evidence presented above suggests that groundwater recharge plays a key role in the rainfall-runoff relationship in this catchment. We can further test this conceptual picture as follows. Each of the relationships in Figure 3a–3c is a convolution kernel, formed by deconvolving streamflow by precipitation (RRD_P, Figure 3a), deconvolving groundwater recharge by precipitation ($GRRD_P$, Figure 3b) and deconvolving streamflow by groundwater recharge (RRD_{GR}, Figure 3c). If groundwater recharge is an intermediary link in the chain rainfall→groundwater→streamflow, we would expect that the impulse response of groundwater to precipitation (Figure 3b), convolved with the impulse response of streamflow to groundwater (Figure 3c), should closely approximate the impulse response of streamflow to precipitation (Figure 3a). And this is indeed the case. The convolution of $GRRD_P$ and RRD_{GR}, shown in red in Figure 3d, very closely matches the runoff response distribution linking precipitation and streamflow (RRD_P, shown in blue in Figure 3d). Rescaling the red curve in Figure 3d by a ratio of 1.05 yields a nearly exact match to the blue curve, suggesting that about 95% of streamflow was initially driven by vertical infiltration and groundwater recharge, and then by "linear reservoir"-type groundwater discharge to the stream. The quicker and larger response of $GRRD_P$ than RRD_P in Figures 3a and 3b

is consistent with a conceptual picture in which streamflow is primarily driven by groundwater, and thus comes later in the rainfall-groundwater-streamflow chain.

It should be emphasized that the good agreement shown in Figure 3d is not mathematically inevitable (see also Section 4.1 below). Without causal linkages between precipitation and groundwater recharge, and between groundwater recharge and streamflow, there is no theoretical expectation that their convolution kernels, when convolved together (the red curve in Figure 3d), should closely approximate the convolution kernel of the relationship between precipitation and streamflow (the blue curve in Figure 3d).

3.2. Quantifying the Nonlinear Response to Different Input Rates

In many hydrological systems, outputs may respond nonlinearly to inputs; for example, streamflow may respond more-than-proportionally to changes in precipitation rates. ERRA allows these nonlinear relationships to be quantified directly from hydrological time series, by expressing the response to precipitation at each lag time using a piecewise-linear function of precipitation intensity. The resulting response curves, termed Nonlinear Response Functions (NRFs), quantify how the system responds over time to lagged inputs at a given intensity. Whereas response distributions like the runoff response distribution driven by precipitation (RRD_P) express how the system responds to a unit of precipitation (and thus have units of d^{-1} here), NRFs express how the system responds to one time step of precipitation at a given intensity (and thus have units of mm d^{-1} here).

Figure 4a presents the nonlinear response function for groundwater recharge driven by precipitation (NRF_GR_P). The peak values of this nonlinear response function (which are found, in each case, at a one-day lag) increase nonlinearly with precipitation intensity, particularly above precipitation rates of roughly 30 mm d^{-1} (Figure 4b). The recharge volume (the area under each curve in Figure 4a) also increases more than proportionally with precipitation intensity, particularly above precipitation rates of roughly 40 mm d^{-1} (Figure 4c).

The NRF of streamflow driven by precipitation (Figure 4d) is noisier than the NRF of groundwater recharge (Figure 4a), in part because our data set includes 1,679 days with non-zero groundwater recharge, but only 416 days with non-zero streamflow. The NRF of streamflow driven by precipitation (NRF_R_P) is also noisier at higher P intensity (Figure 4d) because days with high-intensity precipitation are rare. Both peak runoff response (Figure 4e) and runoff volume (Figure 4f) appear to increase more steeply above precipitation rates of about 30 mm d^{-1}.

The NRF of streamflow driven by groundwater recharge NRF_R_{GR} (Figure 4g) is nearly zero for groundwater recharge rates less than about 10 mm d^{-1}. The peak response (Figure 4h) rises more steeply above groundwater recharge rates of 10–20 mm d^{-1}, but the runoff volume (Figure 4i) appears to increase almost linearly with groundwater recharge rate. There is almost no runoff response at either precipitation intensities <10 mm d^{-1} or groundwater recharge rates <10 mm d^{-1} (Figures 4d and 4g).

One could also ask how differences in antecedent wetness shape the nonlinear responses shown in Figure 4 above. ERRA allows us to explore this question by separating the data set into specified ranges of antecedent wetness, and then jointly estimating nonlinear response functions for all of these subsets (whose effects potentially overlap through time). Here, we use the sum of precipitation over the previous 3 days as a proxy for antecedent wetness, and estimate the NRF of groundwater recharge driven by precipitation (NRF_GR_P) for different ranges of this antecedent wetness proxy. We do not perform a similar calculation for the response of streamflow to precipitation, because non-zero streamflows are too rare in this catchment (comprising only 7.6% of our data set) to yield well-constrained results.

Figure 5 shows how the peak of groundwater recharge's nonlinear response function (NRF_GR_P) increases with both precipitation intensity and antecedent wetness. Peak groundwater recharge rates increase nonlinearly as a function of precipitation intensity (Figure 5a), with the vertical displacements between the different curves demonstrating a tendency toward greater groundwater recharge (for the same precipitation rates) when antecedent wetness is higher.

Figure 5b presents perpendicular slices through the same data. As Figure 5b shows, groundwater recharge rates increase systematically with antecedent wetness, for each range of precipitation intensity. The vertical offsets between the curves also show that groundwater recharge rates are higher (for the same antecedent wetness) when precipitation rates are higher. Figures 5a and 5b both suggest that there may be an interaction between the effects

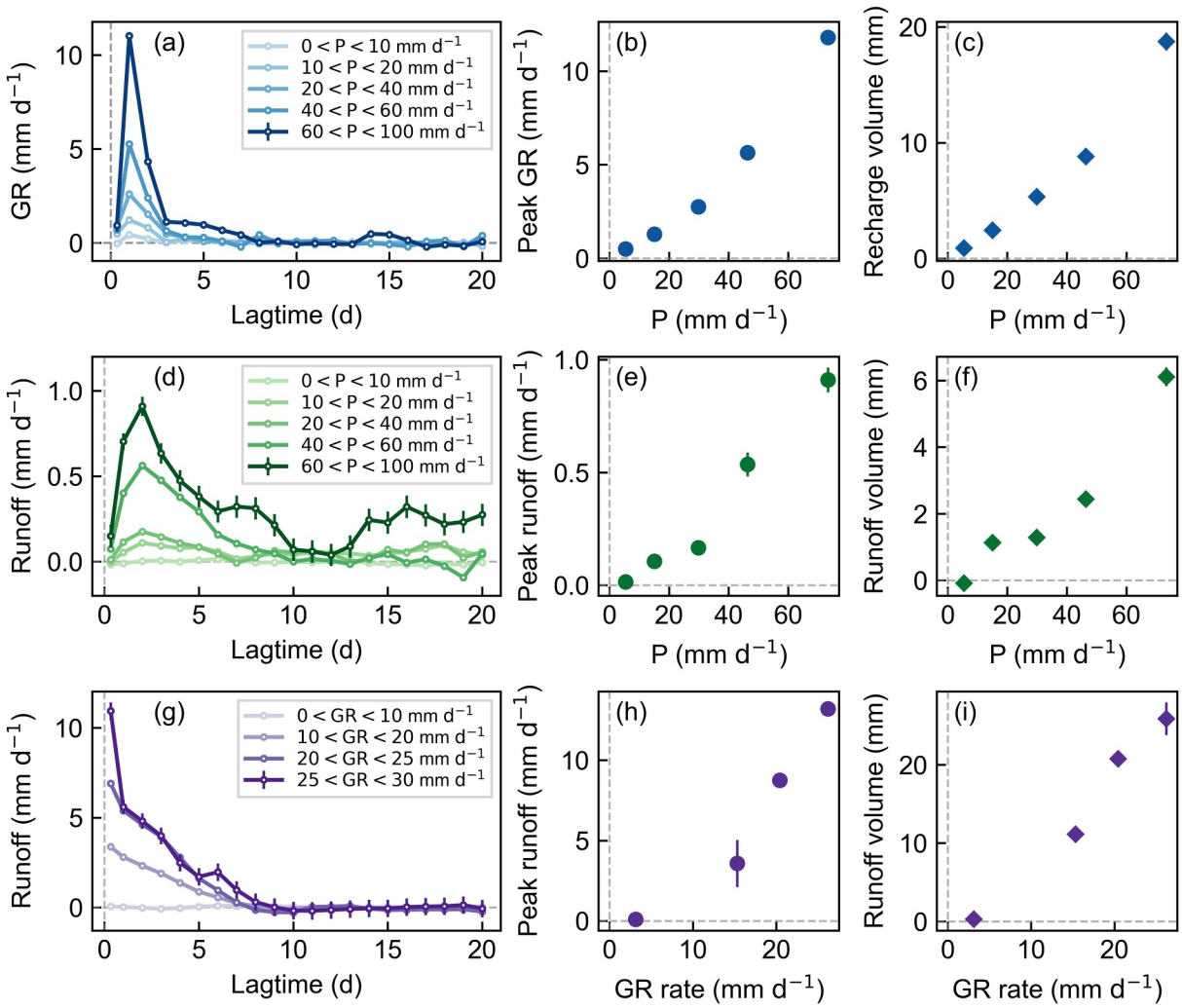

Figure 4. Nonlinear response behaviors at different input rates quantified by nonlinear response functions (NRF). (a) NRF of groundwater recharge driven by precipitation intensity (NRF_GR_P). (b) Peak groundwater recharge response (i.e., the peaks of the curves in panel (a), as a function of precipitation intensity. (c) Groundwater recharge volume under different precipitation intensities. (d) NRF of runoff driven by precipitation intensity (NRF_R_P). (e) Peak runoff response as a function of precipitation intensity. (f) Runoff volume under different precipitation intensities. (g) NRF of runoff driven by groundwater recharge (NRF_R_{GR}). (h) Peak runoff response as a function of GR rate. (i) Runoff volume under different GR rates. Peak NRF_GR_P and NRF_R_P increased almost linearly when P intensity was less than approximately 30 mm d^{-1}, and then increased more quickly (b), (e); peak groundwater recharge and runoff volumes increased disproportionately at P intensities exceeding roughly 40 mm d^{-1} (c), (f). Clear offsets from zero were observed in streamflow response to both precipitation and groundwater recharge, with near-zero responses observed for substantial non-zero inputs (e), (f), (h), (i).

of precipitation intensity and antecedent wetness (i.e., groundwater recharge is more sensitive to precipitation intensity under wetter antecedent conditions, and also more sensitive to antecedent wetness when precipitation is more intense). This visual impression can be confirmed statistically; a linear model of peak groundwater recharge using both precipitation intensity and antecedent wetness as explanatory variables shows that the interaction between them is statistically significant ($p = 0.03$), but weak (it accounts for roughly 2% of the variance).

3.3. Response Distribution Estimation Based on Modeling Results

The ERRA analyses presented above outline a data-driven exploration of the coupling between precipitation, groundwater, and streamflow in the Yanglou catchment. Simulation modeling provides a different, and potentially complementary, strategy for exploring these coupled processes. In addition, data-driven methods like ERRA may facilitate model-data comparisons, by extracting fingerprints of both model behavior and real-world behavior that can then be compared (Kirchner et al., 1996). Such comparisons may facilitate process-relevant

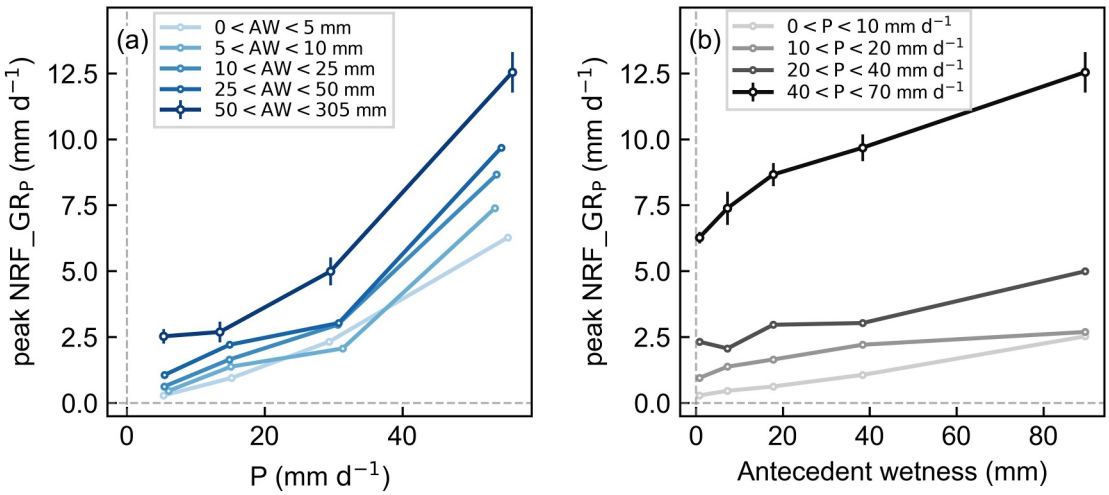

Figure 5. Nonlinear peak groundwater recharge (GR) response under different antecedent wetness conditions. (a) Peak groundwater recharge rates as a function of precipitation intensity under different antecedent wetness conditions. (b) Peak groundwater recharge rates as a function of antecedent wetness (using the sum of previous 3 days precipitation as a proxy) under different P intensity ranges. Peak groundwater recharge was higher under wetter antecedent conditions for a given precipitation rate (a), and higher for greater P intensity for a given antecedent wetness (b).

assessments of model behavior, to complement conventional performance indicators such as Nash-Sutcliffe Efficiency (NSE) or Root Mean Squared Error (RMSE). Here we explore these questions by applying a widely used hydrological model to our study catchment.

3.3.1. Numerical Simulation of Water Table Depth Using Hydrus-1D

Hydrus-1D simulates vadose zone water movement by solving the modified Richards equation using a Galerkin-type linear finite element scheme, implementing the van Genuchten-Mualem model as one of several possible soil hydraulic functions to describe the nonlinear dependence of volumetric water content and unsaturated hydraulic conductivity on water pressure head (Twarakavi et al., 2008). Several key equations can be found in appendix A; more complete documentation can be found in the Hydrus-1D manual and the corresponding literature (van Genuchten, 1980; Šimůnek & van Genuchten, 1996). The model equations contain six soil hydraulic parameters that must be calibrated: residual soil water content θ_r [L^3L^{-3}], saturated soil water content θ_s [L^3L^{-3}], saturated hydraulic conductivity K_s [LT^{-1}], and the van Genuchten-Mualem parameters α [L^{-1}], n, and l. The pore-connectivity parameter l was fixed at 0.5, which is the original optimal value based on 45 soil tests (Mualem, 1976) and has been commonly used as a default value (Fields et al., 2020; Glass et al., 2020; Schaap & Leij, 2000).

We simulated water table depth (WTD) by describing the catchment as a one-dimensional soil column without groundwater pumping and root water uptake. The soil column depth was set to 800 cm with an even spatial discretization of 10 cm. The upper boundary was specified as Atmospheric BC with Surface Runoff; ponding and interception were assumed to be zero. The lower boundary condition was specified as Deep Drainage because the groundwater levels are shallow and the water table never goes below the soil profile. Daily time series of Thiessen-averaged precipitation and arithmetically averaged meteorological data were used as input to the model. To facilitate direct comparisons between the modeling results and our ERRA analyses of the real-world measurements, and to give the model every possible advantage, we used the entire 2004−2018 time series for calibration, and no validation was performed against independent data.

The Nash-Sutcliffe efficiency coefficient (NSE) and root mean square error (RMSE), which indicate better model performance and higher reliability by being closer to 1 and 0 respectively, were used for parameter optimization and performance evaluation. Five soil hydraulic characteristics (θ_r, θ_s, K_s, α, n) were optimized by running parameter combinations in batch mode in sequential steps by narrowing ranges and intervals, with the parameter combination that resulted in the highest NSE between the simulated and measured WTD being selected as the best-fitting parameters in each step. When the NSE improvement became less than 0.5%, the parameters were assumed to be optimal, yielding values of $\theta_r = 0.015$, $\theta_s = 0.415$, $K_s = 156.7$ cm d^{-1}, $\alpha = 0.0325$ cm^{-1}, and

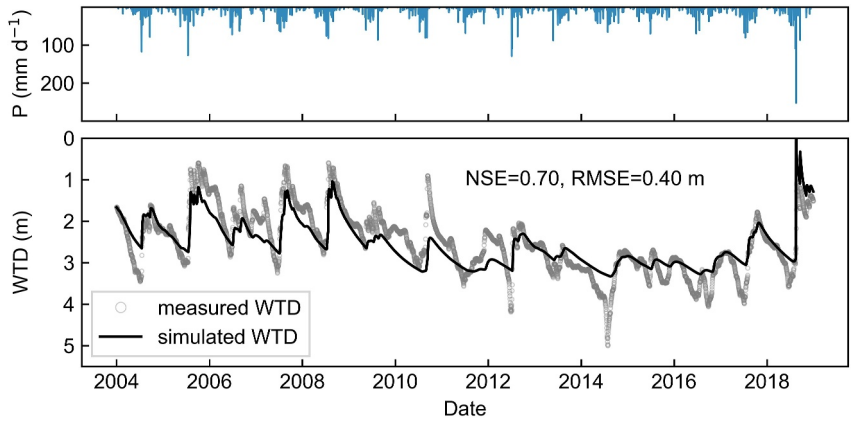

Figure 6. Comparison of the measured and simulated water table depth (WTD) time series (top) Catchment average daily precipitation (bottom) Hollow dots are catchment-averaged WTD, and the solid line is the simulated WTD from Hydrus-1D. Hydrus-1D exhibited reasonable simulation performance (NSE = 0.70, RMSE = 0.40 m) capturing the seasonal patterns and long-term evolution of WTD but missing short-term peaks and drawdowns.

n = 1.30. The empirical parameters A_{qh} (=−0.1568 cm d^{-1}) and B_{qh} (=−0.0087 cm^{-1}) required by the Deep Drainage boundary condition were fitted based on water table depths during the 416 days of non-zero streamflow.

Hydrus-1D simulated the WTD time series reasonably well (NSE = 0.70, RMSE = 0.40 m; Figure 6), particularly in view of the fact that the simulations ignore interception, root water uptake, and groundwater pumping, which may account for the sharp drawdowns that occur in some summers (unfortunately pumping data are unavailable at our site).

3.3.2. Response Distributions in the Real World and Model World

To characterize the water table response to precipitation in the calibrated model and to compare it to water table response in the real world, we used ERRA to estimate groundwater recharge response distributions driven by precipitation ($GRRD_P$) using both the observed water table depths (real world) and those simulated by Hydrus-1D (model world). The process of estimating these response distributions was identical between the model world and the real world, and used the same (measured) precipitation time series; only the WTD time series differed between the model world and the real world. We compared model-world and real-world response distributions for the whole 15-year period (2004−2018), as well as three 5-year periods with contrasting model performance (2004−2008, 2009−2013, 2014−2018). Results for each period are shown in separate columns in Figure 7.

Figure 7 reveals substantial differences between the model world and the real world. Real-world groundwater recharge response to precipitation (blue symbols) peaks on the first day after precipitation falls, and the peak

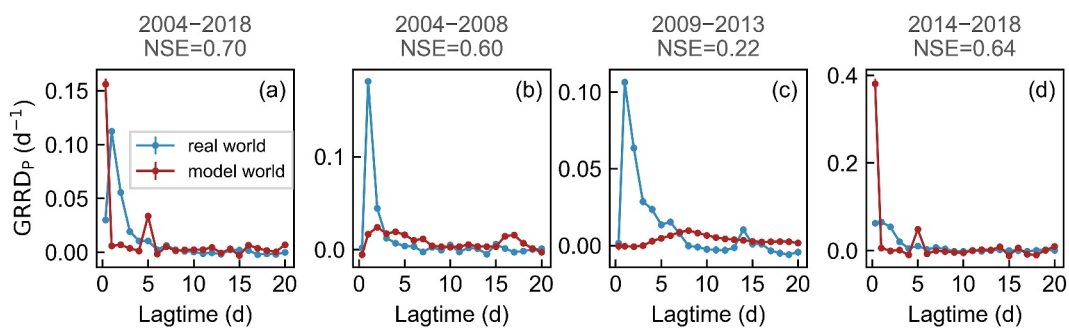

Figure 7. Groundwater recharge response distribution driven by precipitation ($GRRD_P$) calculated from real-world WTD (blue symbols) and Hydrus-1D simulated WTD (red symbols). The four columns show response distributions for different periods, with contrasting simulation model performance as indicated by the NSE values. The model response distributions vary dramatically between the different periods, and differ markedly from the responses observed in the real-world data, even during periods when model NSE values are relatively high.

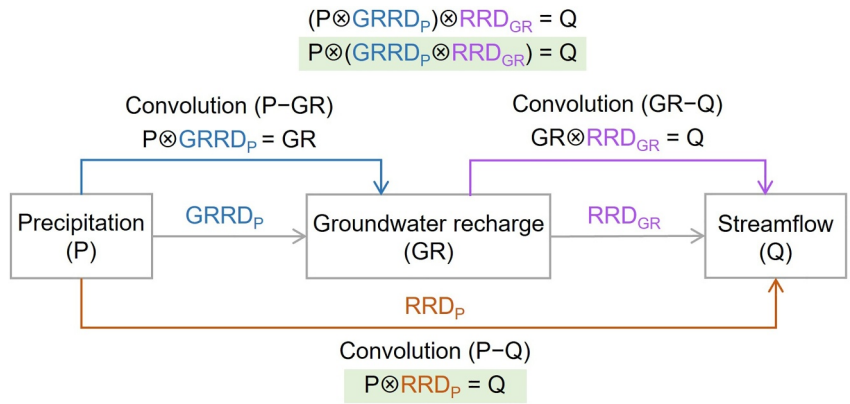

Figure 8. Diagram illustrating potential convolution relationships between precipitation, groundwater recharge, and streamflow. If precipitation is linked to groundwater recharge by a convolution (blue arrow) with kernel $GRRD_P$, and groundwater recharge is linked to streamflow by a convolution (purple arrow) with kernel RRD_{GR}, then precipitation and streamflow are also linked by a convolution (red arrow) whose kernel, RRD_P, is itself the convolution of $GRRD_P$ and RRD_{GR}.

height is consistent within a factor of 2.7 across the four time intervals in Figure 7 (note that the axis scales differ substantially). In the model world (red symbols), by contrast, groundwater recharge response to precipitation may peak on the same day as precipitation, or up to 8 days later, with peak heights varying by roughly a factor of 40. In no case does the groundwater recharge response in the model resemble the groundwater recharge response in the real world.

Although the model's overall performance appears reasonably good (NSE = 0.70), Figure 7 reveals significant deviations from the real-world behavior. Groundwater recharge response in the model is not meaningfully better during periods when the NSE is relatively high than when it is relatively low, which may seem counterintuitive. This behavior is explained by the fact that, as Figure 6 shows, the NSE will mainly be determined by how well the model matches the seasonal and long-term evolution of the water table, whereas the groundwater recharge response in Figure 7 mostly takes place on time scales of just a few days. This example illustrates how model performance metrics like NSE may not reflect whether key mechanisms within the model behave realistically or not. In such cases, greater diagnostic insight may be gained by statistically extracting key fingerprints of interest from the model behavior and the real-world data, and comparing these fingerprints directly, as we have done here (see also Kirchner et al., 1996).

4. Discussion

4.1. Causation Chain Linking Rainfall to Groundwater Recharge and Streamflow

The results presented above demonstrate that groundwater recharge is a key link between rainfall and streamflow in our intensively farmed catchment with intermittent streamflow. The streamflow response to groundwater recharge (Figure 3c) implies that ~45% of groundwater recharge becomes streamflow, and when it is convolved with the groundwater recharge response to precipitation, it implies that ~95% of streamflow results from the rainfall-groundwater-streamflow chain (Figure 3d).

Mathematically, the response distributions (responses of output to unit input) estimated by ERRA are convolution kernels between inputs and outputs over a range of lag times. That is, the response of groundwater recharge to 1 mm precipitation ($GRRD_P$; Figure 3b) is the convolution kernel of the rainfall-groundwater system, assuming this system can be represented as a convolution, denoted as $P \otimes GRRD_P = GR$. Similarly, the response of streamflow to groundwater recharge (RRD_{GR}; Figure 3c) is the kernel of the convolution linking groundwater and streamflow ($GR \otimes RRD_{GR} = Q$), and the runoff response to precipitation (RRD_P, Figure 3a) is the kernel of the convolution linking precipitation and streamflow ($P \otimes RRD_P = Q$).

The convolution experiment in Figure 3d tests the hypothesis (see Figure 8) that if streamflow fluctuations are primarily generated by the lagged effects of groundwater recharge, which in turn is primarily generated by the lagged effects of precipitation, the runoff response distribution RRD_P should be the convolution of the

groundwater recharge response distribution driven by precipitation ($GRRD_P$) and the runoff response distribution driven by groundwater recharge (RRD_{GR}). Mathematically, the convolution experiment in Figure 3d tests whether $RRD_P = GRRD_P \otimes RRD_{GR}$. If this is the case, then it implies a convolution chain linking precipitation to groundwater recharge to streamflow: by the associativity property of convolution, $Q = GR \otimes RRD_{GR} = (P \otimes GRRD_P) \otimes RRD_{GR} = P \otimes (GRRD_P \otimes RRD_{GR}) = P \otimes RRD_P = Q$. But this mathematical relationship only holds if the underlying mechanistic relationships actually behave like convolutions. Therefore the good match between $GRRD_P \otimes RRD_{GR}$ and RRD_P (shown by the red and blue symbols in Figure 3d) strongly suggests that the linkages between precipitation, groundwater recharge and streamflow are well approximated by convolutions, and that they form a causal chain between precipitation and streamflow with groundwater recharge as an intermediary. Hydrologically, this good match between $GRRD_P \otimes RRD_{GR}$ and RRD_P implies that the observed precipitation, groundwater and streamflow dynamics are consistent with vertical infiltration recharging the groundwater followed by "linear reservoir" discharge from the groundwater system to streamflow in our study catchment.

The results of this convolution experiment are also consistent with a much simpler thought experiment. The flashy response of streamflow to precipitation, in a catchment with no streamflow over 90% of the time, would intuitively seem to indicate that streamflow at Yanglou is generated mostly by overland flow processes. But note that the runoff response distribution shows that streamflow peaks a full 2 days after precipitation falls, in a catchment of only 265 km^2, with typical flow distances to the outlet on the order of 15 km (Figure 1). The median flood wave celerity at global scale is 1.6 m s^{-1} (Allen et al., 2018), which would travel 15 km in 2.6 hr, and flood celerities in natural streams typically range from 0.25 to 10 m s^{-1} (Allen et al., 2018; Brakenridge et al., 1998; Dykstra & Dzwonkowski, 2020; Meyer et al., 2018; Sriwongsitanon et al., 1998; Turner-Gillespie et al., 2003), implying time lags of between 0.4 and 17 hr within the channel network. Thus the time for water to reach the channel must be over 30 hr, which is much too slow to plausibly describe typical overland flow processes. Instead, our analysis shows that the observed impulse responses connecting precipitation to recharge, and recharge to streamflow, are quantitatively sufficient—both in magnitude and in lag time—to explain the observed impulse response between precipitation and streamflow.

Both our ERRA analysis and the Hydrus-1D simulations average over spatial variations in hydrological processes and catchment properties. An obvious consequence is that we have no insight into, for example, whether runoff generation is strongly localized near channels or more broadly distributed across the landscape. Such mechanistic insights would be best gained from smaller-scale observational studies. Our analysis, instead, is a proof-of-concept for what can be learned from impulse responses linking precipitation, groundwater dynamics, and streamflow at the catchment scale. Our analysis does not assume that hydrological processes and properties are spatially uniform in our study catchment; instead, it averages over whatever heterogeneity is present. Readers should understand, of course, that any results from this analysis are descriptions of catchment-scale behavior, which do not necessarily apply at each point on the landscape.

4.2. Nonlinear and Nonstationary Dependence of Groundwater Recharge on Precipitation Intensity and Antecedent Wetness

In some groundwater systems, recharge may occur only after rainfall exceeds a threshold, such that high-intensity events dominate total recharge, whereas in others, infiltration rates may be limited, such that lower-intensity, more frequent events may dominate total recharge (Asoka et al., 2018; Jasechko et al., 2014; Taylor et al., 2013). The results shown in Figures 4a–4c suggest that high-intensity precipitation contributes more-than-proportionally to groundwater recharge. Figure 4c implies that at precipitation intensities below about 50 mm d^{-1}, recharge is approximately 16–19% of precipitation, but at precipitation rates of roughly 75 mm d^{-1}, recharge is about 26% of precipitation. Of course, the total contribution to groundwater recharge depends not only on how much recharge occurs from a given precipitation event, but also on how frequently precipitation of similar intensity occurs. At Yanglou, high-intensity events contribute more-than-proportionally to recharge, but they are rare enough that they do not dominate total recharge on an annual time scale.

Our results also indicate that groundwater recharge at Yanglou is higher, for any given precipitation intensity, when antecedent wetness (here inferred from 3-day antecedent precipitation) is also higher (Figure 5). The curves shown in Figure 5 have significant uncertainty, particularly at the highest precipitation intensities and antecedent wetness levels (see error bars in Figure 5), because such conditions are rare and thus the available data describing those conditions are scarce. Nonetheless there is some suggestion that groundwater recharge may respond more

nonlinearly to precipitation intensity under wetter conditions (Figure 5a), and that for any given precipitation intensity, groundwater recharge may asymptotically approach a plateau at high levels of antecedent wetness (Figure 5b). Thus, under wetter conditions, groundwater recharge may be more sensitive to increases in precipitation and less sensitive to increases in wetness. However, these conjectures would need to be confirmed with a longer data set that captures more events with high antecedent wetness or high precipitation intensity.

4.3. Model Performance Versus Model Realism

Hydrological models have evolved greatly in recent decades, but in many cases their performance is still mostly evaluated based on the statistical goodness-of-fit between time series of model projections and measured data. The advantages, disadvantages, applicability and effectiveness of single-index and multi-objective frameworks have been extensively discussed in the literature (e.g., Beldring, 2002; Gupta et al., 2009; Legates & McCabe, 1999; Manikanta & Vema, 2022; Ritter & Muñoz-Carpena, 2013). The question remains open whether (or how much) a good model fit implies realism in the underlying model mechanisms. Critics have argued that conventional goodness-of-fit tests are insufficient for evaluating model realism (e.g., Beven, 2002; Kirchner et al., 1996; Wagener, 2003) because of the equifinality of parameter sets (Beven, 2006) and uncertainties introduced by deficiencies in the model structure and calibration data (Wagener & Gupta, 2005). Thus it is often unclear whether models are getting the right answers for the right reasons (Kirchner, 2006). It has been suggested that more incisive model-data comparisons could be achieved by extracting characteristic relationships from both the model and the data, and then comparing these with one another (rather than comparing the raw time series themselves).

The simple analysis outlined in Section 3.3 tests the effectiveness of this approach, using a Hydrus-1D model of groundwater recharge at Yanglou. The results demonstrate significant deviations between how the model and the real world transmit precipitation to groundwater (Figure 7), even though the model achieves reasonable performance as measured by conventional goodness-of-fit statistics. The modeled response to precipitation varies greatly between different time periods, with poor agreement between the modeled and observed response distributions shown in Figure 7, both when the conventional goodness-of-fit statistics are relatively good and when they are relatively poor. The groundwater recharge response distribution asks the question: how much does recharge respond to precipitation over different time lags? The real-world data gave broadly similar answers to this question in the different periods shown in Figure 7. For these same periods, the model's answers were inconsistent among themselves and deviated substantially from the data, suggesting that the modeled mechanisms do not capture essential features of the groundwater recharge process. Fingerprints of behavior, such as those provided by ERRA, can not only show that models and observations are different; they can also quantify how they differ. Thus they can aid in improving our understanding of the system, and in refining the simulation models that encapsulate that understanding.

5. Conclusion

This proof-of-concept study has explored how impulse responses estimated by nonparametric deconvolution and demixing (Ensemble Rainfall-Runoff Analysis, or ERRA) can help to characterize and quantify linkages between precipitation, groundwater dynamics, and streamflow at the catchment scale. Runoff response distributions estimated by ERRA suggest that runoff generation at the Yanglou catchment occurs primarily via infiltration of precipitation to the water table, which in turn triggers the release of groundwater to streamflow. The response distribution of groundwater recharge to precipitation ($GRRD_P$, Figure 3b) peaks the day after precipitation falls, reflecting rapid infiltration to depths of several meters (Figure 2). The response distribution of streamflow to groundwater recharge (RRD_{GR}, Figure 3c) is nearly exponential, suggesting linear-reservoir behavior of the groundwater system at the catchment scale. The combined effects of groundwater recharge and groundwater discharge (as inferred by convolving these two response distributions) explain roughly 95% of the response distribution of streamflow to precipitation (RRD_P, Figures 3a and 3d) suggesting that the great majority of streamflow is driven by the groundwater system. The runoff response distribution peaks 2 days after precipitation falls, and one day after the peak in groundwater recharge, further suggesting that groundwater recharge is the intermediate link in streamflow generation at Yanglou. The response distributions suggest that roughly 21% of precipitation eventually becomes groundwater recharge, and roughly 45% of groundwater recharge eventually becomes streamflow.

Our proof-of-concept study illustrates how ERRA can quantify nonlinearity and nonstationarity in hydrological response at the catchment scale. At Yanglou, the response distributions of streamflow exhibit nonlinear dependence on both precipitation and groundwater recharge (Figures 4d–4i), with little or no streamflow response to precipitation or recharge rates of less than 10 mm d^{-1}. Groundwater recharge also exhibits nonlinear dependence on precipitation intensity (Figures 4a–4c), and nonstationary dependence on catchment wetness (as inferred from antecedent precipitation; Figure 5). These nonlinear and nonstationary behaviors also interact: groundwater recharge is more sensitive to precipitation intensity under wet conditions, and more sensitive to wetness when precipitation intensity is high, although this interaction explains only a small fraction of the variation in recharge rates.

We have also presented a proof-of-concept for the use of ERRA to critically evaluate simulation models. A one-dimensional infiltration model (Hydrus-1D) fits the water table time series reasonably well (NSE = 0.70; Figure 6). However, groundwater recharge response distributions estimated from the model behavior for different time periods (red points, Figure 7) are inconsistent with each other, and with response distributions estimated from real-world data (blue points, Figure 7), suggesting significant problems with the model's representation of the infiltration process. This example illustrates how signatures of system behavior, such as the response distributions estimated by ERRA, can provide effective diagnostic tests of simulation models of natural systems.

Appendix A: Governing Equations of Hydrus-1D Model

Hydrus-1D simulates vadose zone water movement by solving the modified Richards equation using a Galerkin-type linear finite element scheme (Twarakavi et al., 2008):

$$\frac{\partial \theta}{\partial t} = \frac{\partial}{\partial x}\left[K\left(\frac{\partial h}{\partial x} + \cos\alpha\right)\right] - S \tag{A1}$$

where t is time [T], x is the spatial coordinate [L] (positive upward), S is the sink term, α is the angle between the flow direction and the vertical axis, h is the water pressure head [L], θ is the volumetric water content, and K is the unsaturated hydraulic conductivity.

The nonlinear dependence of θ and K on h is commonly described by the van Genuchten-Mualem equations,

$$\theta(h) = \theta_r + \frac{\theta_s - \theta_r}{[1 + |\alpha h|^n]^m} \quad (h < 0) \tag{A2}$$

$$K(h) = K_s S_e^l \left[1 - \left(1 - S_e^{1/m}\right)^m\right]^2 \tag{A3}$$

where θ_r and θ_s are residual soil water content and saturated soil water content [L^3L^{-3}], respectively; K_s is hydraulic conductivity [LT^{-1}], α [L^{-1}] and n are empirical shape parameters, and l is a pore-connectivity parameter.

For shallow groundwater conditions, the vertical drainage $q(h)$ is approximated by a flux which depends on the position of the groundwater level:

$$q(h) = A_{qh} \exp(B_{qh}|h - GWL0L|) \tag{A4}$$

where A_{qh} and B_{qh} are empirical parameters, $GWL0L$ represents the reference position of the groundwater level.

Conflict of Interest

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

Data Availability Statement

The Ensemble Rainfall-Runoff Analysis (ERRA) source code, along with introductory documentation for users, is available at https://doi.org/10.16904/envidat.529 (Kirchner, 2024b). The data that support the findings of this study are available at https://doi.org/10.5281/zenodo.14197671 (Gao et al., 2024).

Acknowledgments
This work was supported by the National Natural Science Foundation of China (Grant 52179013), Fundamental Research Funds for the Central Universities of China (B240203007), and the IWHR Research and Development Support Program (JZ110145B0022023, WH0145B022021). HG thanks the Chinese Scholarship Council for financial support (Grant 202206710058). We thank all the people in the field and lab for the data support, and we thank the editor and two anonymous reviewers for their thoughtful comments.

References

Allen, G. H., David, C. H., Andreadis, K. M., Hossain, F., & Famiglietti, J. S. (2018). Global estimates of river flow wave travel times and implications for low-latency satellite data. *Geophysical Research Letters*, *45*(15), 7551–7560. https://doi.org/10.1029/2018GL077914

Appels, W. M., Bogaart, P. W., & Van Der Zee, S. E. A. T. M. (2016). Surface runoff in flat terrain: How field topography and runoff generating processes control hydrological connectivity. *Journal of Hydrology*, *534*, 493–504. https://doi.org/10.1016/j.jhydrol.2016.01.021

Asoka, A., Wada, Y., Fishman, R., & Mishra, V. (2018). Strong linkage between precipitation intensity and monsoon season groundwater recharge in India. *Geophysical Research Letters*, *45*(11), 5536–5544. https://doi.org/10.1029/2018GL078466

Beldring, S. (2002). Multi-criteria validation of a precipitation–runoff model. *Journal of Hydrology*, *257*(1), 189–211. https://doi.org/10.1016/S0022-1694(01)00541-8

Best, M. J., Abramowitz, G., Johnson, H. R., Pitman, A. J., Balsamo, G., Boone, A., et al. (2015). The plumbing of land surface models: Benchmarking model performance. *Journal of Hydrometeorology*, *16*(3), 1425–1442. https://doi.org/10.1175/JHM-D-14-0158.1

Beven, K. (2002). Towards a coherent philosophy for modelling the environment. *Proceedings: Mathematical, Physical and Engineering Sciences*, *458*(2026), 2465–2484. https://doi.org/10.1098/rspa.2002.0986

Beven, K. (2006). A manifesto for the equifinality thesis. *Journal of Hydrology*, *320*(1), 18–36. https://doi.org/10.1016/j.jhydrol.2005.07.007

Beven, K., & Freer, J. (2001). Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. *Journal of Hydrology*, *249*(1), 11–29. https://doi.org/10.1016/S0022-1694(01)00421-8

Beven, K., & Smith, P. (2015). Concepts of information content and likelihood in parameter calibration for hydrological simulation models. *Journal of Hydrologic Engineering*, *20*(1), 1–15. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000991

Brakenridge, G. R., Tracy, B. T., & Knox, J. C. (1998). Orbital SAR remote sensing of a river flood wave. *International Journal of Remote Sensing*, *19*(7), 1439–1445. https://doi.org/10.1080/014311698215559

Chen, C., He, W., Zhou, H., Xue, Y., & Zhu, M. (2020). A comparative study among machine learning and numerical models for simulating groundwater dynamics in the Heihe River Basin, northwestern China. *Scientific Reports*, *10*(1), 3904. https://doi.org/10.1038/s41598-020-60698-9

Chen, L., Sela, S., Svoray, T., & Assouline, S. (2013). The role of soil-surface sealing, microtopography, and vegetation patches in rainfall-runoff processes in semiarid areas. *Water Resources Research*, *49*(9), 5585–5599. https://doi.org/10.1002/wrcr.20360

Dooge, J. C. I. (1959). A general theory of the unit hydrograph. *Journal of Geophysical Research (1896-1977)*, *64*(2), 241–256. https://doi.org/10.1029/JZ064i002p00241

Dooge, J. C. I. (1986). Looking for hydrologic laws. *Water Resources Research*, *22*(9S), 46S–58S. https://doi.org/10.1029/WR022i09Sp0046S

Dykstra, S. L., & Dzwonkowski, B. (2020). The propagation of fluvial flood waves through a backwater-estuarine environment. *Water Resources Research*, *56*(2), e2019WR025743. https://doi.org/10.1029/2019WR025743

Fields, J. S., Owen, J. S., Stewart, R. D., Heitman, J. L., & Caron, J. (2020). Modeling water fluxes through containerized soilless substrates using HYDRUS. *Vadose Zone Journal*, *19*(1), e20031. https://doi.org/10.1002/vzj2.20031

Gao, H., Ju, Q., Zhang, D., Wang, Z., Hao, Z., & Kirchner, J. (2024). Data - Quantifying dynamic linkages between precipitation, groundwater recharge, and streamflow using ensemble rainfall-runoff analysis [Dataset]. *Zenodo*. https://doi.org/10.5281/zenodo.14197671

Glass, J., Šimůnek, J., & Stefan, C. (2020). Scaling factors in HYDRUS to simulate a reduction in hydraulic conductivity during infiltration from recharge wells and infiltration basins. *Vadose Zone Journal*, *19*(1), e20027. https://doi.org/10.1002/vzj2.20027

Gootman, K. S., & Hubbart, J. A. (2021). Rainfall, runoff and shallow groundwater response in a mixed-use, agro-forested watershed of the northeast, USA. *Hydrological Processes*, *35*(8), 1–14. https://doi.org/10.1002/hyp.14312

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, *377*(1–2), 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003

James, W., & Johanson, R. C. (1999). A note on an inherent difficulty with the unit hydrograph method. *Journal of Water Management Modeling*, 1–10. https://doi.org/10.14796/JWMM.R204-01

Jasechko, S., Birks, S. J., Gleeson, T., Wada, Y., Fawcett, P. J., Sharp, Z. D., et al. (2014). The pronounced seasonality of global groundwater recharge. *Water Resources Research*, *50*(11), 8845–8867. https://doi.org/10.1002/2014WR015809

Kim, S. (2009). Characterization of soil moisture responses on a hillslope to sequential rainfall events during late autumn and spring. *Water Resources Research*, *45*(9), 1–15. https://doi.org/10.1029/2008WR007239

Kirchner, J. W. (2006). Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. *Water Resources Research*, *42*(3). https://doi.org/10.1029/2005WR004362

Kirchner, J. W. (2009). Catchments as simple dynamical systems: Catchment characterization, rainfall-runoff modeling, and doing hydrology backward. *Water Resources Research*, *45*(2). https://doi.org/10.1029/2008WR006912

Kirchner, J. W. (2022). Impulse response functions for nonlinear, nonstationary, and heterogeneous systems, estimated by deconvolution and demixing of noisy time series. *Sensors*, *22*(9), 3291. https://doi.org/10.3390/s22093291

Kirchner, J. W. (2024a). Characterizing nonlinear, nonstationary, and heterogeneous hydrologic behavior using ensemble rainfall-runoff analysis (ERRA): Proof of concept. *Hydrology and Earth System Sciences*, *28*, 4427–4454. https://doi.org/10.5194/hess-28-4427-2024

Kirchner, J. (2024b). ERRA -- an R script for Ensemble Rainfall-Runoff Analysis [Dataset]. *EnviDat*. https://doi.org/10.16904/envidat.529

Kirchner, J. W., Hooper, R. P., Kendall, C., Neal, C., & Leavesley, G. (1996). Testing and validating environmental models. *Science of the Total Environment*, *183*(1), 33–47. https://doi.org/10.1016/0048-9697(95)04971-1

Kratzert, F., Klotz, D., Herrnegger, M., Sampson, A. K., Hochreiter, S., & Nearing, G. S. (2019). Toward improved predictions in ungauged basins: Exploiting the power of machine learning. *Water Resources Research*, *55*(12), 11344–11354. https://doi.org/10.1029/2019WR026065

Legates, D. R., & McCabe, G. J. (1999). Evaluating the use of "goodness-of-fit" Measures in hydrologic and hydroclimatic model validation. *Water Resources Research*, *35*(1), 233–241. https://doi.org/10.1029/1998WR900018

Manikanta, V., & Vema, V. K. (2022). Formulation of wavelet based multi-scale multi-objective performance evaluation (WMMPE) metric for improved calibration of hydrological models. *Water Resources Research*, *58*(7), e2020WR029355. https://doi.org/10.1029/2020WR029355

Meyer, A., Fleischmann, A. S., Collischonn, W., Paiva, R., & Jardim, P. (2018). Empirical assessment of flood wave celerity–discharge relationships at local and reach scales. *Hydrological Sciences Journal*, *63*(15–16), 2035–2047. https://doi.org/10.1080/02626667.2018.1557336

Mualem, Y. (1976). A new model for predicting the hydraulic conductivity of unsaturated porous media. *Water Resources Research*, *12*(3), 513–522. https://doi.org/10.1029/WR012i003p00513

Nearing, G. S., Kratzert, F., Sampson, A. K., Pelissier, C. S., Klotz, D., Frame, J. M., et al. (2021). What role does hydrological science play in the age of machine learning? *Water Resources Research*, *57*(3), 1–15. https://doi.org/10.1029/2020WR028091

Nippgen, F., McGlynn, B. L., Emanuel, R. E., & Vose, J. M. (2016). Watershed memory at the Coweeta Hydrologic Laboratory: The effect of past precipitation and storage on hydrologic response. *Water Resources Research*, *52*(3), 1673–1695. https://doi.org/10.1002/2015WR018196

Oda, T., Egusa, T., Ohte, N., Hotta, N., Tanaka, N., Green, M. B., & Suzuki, M. (2021). Effects of changes in canopy interception on stream runoff response and recovery following clear-cutting of a Japanese coniferous forest in Fukuroyamasawa Experimental Watershed in Japan. *Hydrological Processes*, *35*(5), 1–14. https://doi.org/10.1002/hyp.14177

Pathiraja, S., Westra, S., & Sharma, A. (2012). Why continuous simulation? The role of antecedent moisture in design flood estimation. *Water Resources Research*, *48*(6), 1–15. https://doi.org/10.1029/2011WR010997

Pelissier, C., Frame, J., & Nearing, G. (2020). Combining parametric land surface models with machine learning. In *IEEE international geoscience and remote sensing symposium* (pp. 3668–3671). https://doi.org/10.1109/IGARSS39084.2020.9324607

Ritter, A., & Muñoz-Carpena, R. (2013). Performance evaluation of hydrological models: Statistical significance for reducing subjectivity in goodness-of-fit assessments. *Journal of Hydrology*, *480*, 33–45. https://doi.org/10.1016/j.jhydrol.2012.12.004

Russo, T. A., & Lall, U. (2017). Depletion and response of deep groundwater to climate-induced pumping variability. *Nature Geoscience*, *10*(2), 105–108. https://doi.org/10.1038/ngeo2883

Schaap, M. G., & Leij, F. J. (2000). Improved prediction of unsaturated hydraulic conductivity with the mualem-van Genuchten model. *Soil Science Society of America Journal*, *64*(3), 843–851. https://doi.org/10.2136/sssaj2000.643843x

Šimůnek, J., & van Genuchten, M. T. (1996). Estimating unsaturated soil hydraulic properties from tension disc infiltrometer data by numerical inversion. *Water Resources Research*, *32*(9), 2683–2696. https://doi.org/10.1029/96WR01525

Singh, N. K., Emanuel, R. E., McGlynn, B. L., & Miniat, C. F. (2021). Soil moisture responses to rainfall: Implications for runoff generation. *Water Resources Research*, *57*(9), 1–17. https://doi.org/10.1029/2020WR028827

Singh, V. P., & Woolhiser, D. A. (2002). Mathematical modeling of watershed hydrology. *Journal of Hydrologic Engineering*, *7*(4), 270–292. https://doi.org/10.1061/(ASCE)1084-0699(2002)7:4(270)

Sriwongsitanon, N., Ball, J. E., & Cordery, I. (1998). An investigation of the relationship between the flood wave speed and parameters in runoff-routing models. *Hydrological Sciences Journal*, *43*(2), 197–213. https://doi.org/10.1080/02626669809492118

Tao, W., Wang, Q., Guo, L., Lin, H., Chen, X., Sun, Y., & Ning, S. (2020). An enhanced rainfall–runoff model with coupled canopy interception. *Hydrological Processes*, *34*(8), 1837–1853. https://doi.org/10.1002/hyp.13696

Tashie, A. M., Mirus, B. B., & Pavelsky, T. M. (2016). Identifying long-term empirical relationships between storm characteristics and episodic groundwater recharge. *Water Resources Research*, *52*(1), 21–35. https://doi.org/10.1002/2015WR017876

Taylor, R. G., Todd, M. C., Kongola, L., Maurice, L., Nahozya, E., Sanga, H., & MacDonald, A. M. (2013). Evidence of the dependence of groundwater resources on extreme rainfall in East Africa. *Nature Climate Change*, *3*(4), 374–378. https://doi.org/10.1038/nclimate1731

Turner-Gillespie, D. F., Smith, J. A., & Bates, P. D. (2003). Attenuating reaches and the regional flood response of an urbanizing drainage basin. *Advances in Water Resources*, *26*(6), 673–684. https://doi.org/10.1016/S0309-1708(03)00017-4

Twarakavi, N. K. C., Šimůnek, J., & Seo, S. (2008). Evaluating interactions between groundwater and vadose zone using the HYDRUS-based flow package for MODFLOW. *Vadose Zone Journal*, *7*(2), 757–768. https://doi.org/10.2136/vzj2007.0082

van Genuchten, M. T. (1980). A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Science Society of America Journal*, *44*(5), 892–898. https://doi.org/10.2136/sssaj1980.03615995004400050002x

Wagener, T. (2003). Evaluation of catchment models. *Hydrological Processes*, *17*(16), 3375–3378. https://doi.org/10.1002/hyp.5158

Wagener, T., & Gupta, H. V. (2005). Model identification for hydrological forecasting under uncertainty. *Stochastic Environmental Research and Risk Assessment*, *19*(6), 378–387. https://doi.org/10.1007/s00477-005-0006-5

Zhu, Q. (2013). *Study on the mechanism of surface-soil-ground water conversion in Huaibei Plain*. (Master's thesis). Hohai University. (in Chinese).