

Water Resources Research

RESEARCH ARTICLE

10.1029/2019WR025001

Key Points:

- Integration of water balance methods with a spatially explicit statistical model improves accuracy of streamflow prediction across the USA
- Study identifies large-scale natural and human controls on water delivery from the land to river networks and quantifies in-stream losses
- Approach improves understanding and parameterization of large-scale hydrological processes and scaling properties to support prediction

Supporting Information:

- Supporting Information S1

Correspondence to:

R. B. Alexander,
ralexander38@gmail.com

Citation:

Alexander, R. B., Schwarz, G. E., & Boyer, E. W. (2019). Advances in quantifying streamflow variability across continental scales: 1. identifying natural and anthropogenic controlling factors in the USA using a spatially explicit modeling method. *Water Resources Research*, 55, 10,893–10,917. <https://doi.org/10.1029/2019WR025001>

Received 21 FEB 2019

Accepted 12 NOV 2019

Accepted article online 18 NOV 2019

Published online 18 DEC 2019

Richard B. Alexander is retired (U.S. Geological Survey).
This article is a companion to 2019WR025037, 2019WR025037R and 2019WR025037RR.

©2019. The Authors.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Advances in Quantifying Streamflow Variability Across Continental Scales: 1. Identifying Natural and Anthropogenic Controlling Factors in the USA Using a Spatially Explicit Modeling Method

Richard B. Alexander¹ , Gregory E. Schwarz¹ , and Elizabeth W. Boyer² 
¹U.S. Geological Survey, Reston, VA, USA, ²Department of Ecosystem Science and Management, Pennsylvania State University, University Park, PA, USA

Abstract Despite considerable progress in hydrological modeling, challenges remain in the interpretation and accurate transfer of hydrological information across watersheds and scales. In the conterminous United States (CONUS), these limitations are related to spatial inconsistencies and constraints in hydrological model structures, including a lack of spatially explicit process components (streams, reservoirs, and watershed development) and restricted estimation of model parameters across watersheds. Collectively, such limitations can impede identification of the causes of streamflow variations across the diversity of watershed sizes and land uses in the CONUS and contribute to model imprecision and spatial inconsistencies in prediction uncertainties. We addressed these concerns with a new approach, the first hybrid (statistical-mechanistic) SPARROW (SPAtially Referenced Regression On Watershed attributes) model of long-term mean annual streamflow, applied across diverse environmental settings of the CONUS. The hybrid model coupled previous catchment-scale (1 km) water balance predictions of “natural” unit area runoff, which are inclusive of major water cycling processes, with additional explanatory variables (e.g., soils, vegetation, land use, topography, water losses in streams, and reservoirs) that account for the effects of natural and cultural water supply and demand processes that operate over large spatial scales and explain streamflow variability across CONUS river basins. Accounting for these statistically unique effects, including a nonlinear surface area-dependent scaling of water loss in river networks, significantly improved the accuracy of mean streamflow predictions in CONUS basins. Our hybrid modeling approach provides new methods for transferring hydrological information to ungauged locations in river networks, especially those in larger and more culturally diverse CONUS watersheds.

1. Introduction

Legacy and emerging environmental concerns require a quantitative ability to predict streamflow at regional and temporal scales in response to variations in climate, land use, and management. Progress in representing how streamflow responds to dominant hydrological processes across diverse watersheds and spatial scales has improved the methods for prediction in ungauged watersheds (Blöschl et al., 2013; Montanari et al., 2013; Sivapalan et al., 2003). These include improved understanding of the connections between small-scale physical models and large-scale hydrological responses (Hrachowitz et al., 2013; He et al., 2011), based on the use of multiple, complementary approaches (Gupta et al., 2013), including advances in parameter “upscaling” methods (Hrachowitz et al., 2013; Samaniego et al., 2010), regionalization techniques (e.g., Carey et al., 2010; Archfield & Vogel, 2010; Stuckey et al., 2014; Eng et al., 2007), and regional and continental applications of mechanistic and statistical hydrological models (e.g., Wolock & McCabe, 1999; Vogel et al., 1997, 1999; Vogel & Sankarasubramanian, 2000; Weiskel et al., 2014; Xia et al., 2012; Bock et al., 2016). These approaches have refined the resolution and skill of mechanistic hydrological models over catchment scales, with improved methods for identifying dominant large-scale controls and regions of hydrologic similarity that explain many spatial and temporal patterns in streamflow, related to climatic and landscape characteristics (e.g., Weiskel et al., 2014; Wolock & McCabe, 1999; Vogel et al., 1997, 1999; Church, 1995; Milly, 1994; Arnold et al., 1999), urbanization (e.g., DeWalle et al., 2000), and human alterations (e.g., Eng et al., 2007).

Despite considerable progress, challenges remain in the methods for identifying dominant hydrologic processes and accurately transferring hydrological information across watersheds (Montanari et al., 2013; Hrachowitz et al., 2013; He et al., 2011; McDonnell et al., 2007; Rode et al., 2010; Xia et al., 2012), especially over large regional and continental scales (Archfield et al., 2015) such as the conterminous United States (CONUS). Several limitations exist in the previous applications of hydrological models. First, the calibration of hydrologic models across large scales is typically performed independently across disjointed subsets of the spatial domain, causing discontinuities in estimated values of the parameters. Applications of physically based models can be especially challenging, leading to overparameterizations that complicate inferences about process effects and contribute to large prediction uncertainties when the calibrated models are transferred to ungauged locations (e.g., Xia et al., 2012; Beven, 2001; McDonnell et al., 2007; Jakeman & Hornberger, 1993). In contrast, many continental-scale applications have employed parsimonious mechanistic model structures with fewer controlling parameters (Archfield et al., 2015), but these simplified models have had mixed success at improving prediction accuracy (e.g., Bock et al., 2016; Arheimer et al., 2019). The methods for upscaling and spatial interpolation have included the following: the regionalization of catchment model parameters (e.g., Bock et al., 2016; Beck et al., 2016; Livneh & Lettenmaier, 2013) and measures of hydrological variability (e.g., Jehn et al., 2019; Addor et al., 2017) based on geographic proximity and similarities in hydrological and climatic conditions; the simultaneous calibration of models across representative catchments having similar watershed attributes (e.g., Arheimer et al., 2019); the use of spatial transfer functions based on the regression of catchment model parameters on watershed characteristics (e.g., Hundecha et al., 2016; Rakovec et al., 2016); and the aggregate use of model outcomes across large scales from independent calibrations in individual watersheds (e.g., Newman et al., 2015; Weiskel et al., 2014; Wolock & McCabe, 1999). Despite advances that have contributed to improved spatial sharing of hydrological information across continental scales (e.g., Bock et al., 2016; Hundecha et al., 2016; Rakovec et al., 2016), a persistent challenge over large scales is developing statistical methods for estimating parameters that are spatially and structurally consistent. Many of the approaches, especially in the CONUS, have entailed a more piecemeal sharing of hydrological information across the continental domain and the associated process components of the models (e.g., Bock et al., 2016; Livneh & Lettenmaier, 2013; Arheimer et al., 2019). The imposed spatial restrictions on the areal extent of the model calibrations and information sharing are partially influenced by the computational demands associated with the high temporal resolution of the models (e.g., daily, hourly) that add considerable complexity to the parameter estimation. These spatial constraints can adversely affect model precision (Schwarz et al., 2006, 2011), with reductions in the statistical uniqueness and sensitivity of model parameters.

Second, many of the continental-scale applications lack explicit parameterizations of river network features (e.g., evaporative loss, recharge, baseflow) that are estimated simultaneously with the climatic and landscape components of hydrological models. For example, mechanistic models (Wolock & McCabe, 1999; Xia et al., 2012; Weiskel et al., 2014; Bock et al., 2016; Beck et al., 2016; Newman et al., 2015) have typically generated streamflow from uniform landscape components (e.g., “hydrological response units” partitioned on the basis of uniform soils, land cover, and topography), with river networks used primarily to facilitate flow routing without accounting for water loss processes (e.g., evaporative losses, recharge, and water use) or estimating river loss processes using manual “trial-and-error” methods (Weiskel et al., 2014), which may be susceptible to large uncertainties. Models that include river and reservoir evaporative processes (e.g., Arheimer et al., 2019; Rakovec et al., 2016) have not consistently estimated the river network parameters simultaneously with catchment and regionalized parameters. Moreover, statistical models typically lack mass balance constraints and represent hydrological processes in a simplistic manner, using nonspatially distributed, averaged properties across watersheds (Vogel et al., 1997, 1999; Vogel & Sankarasubramanian, 2000). More broadly, approaches for regionalizing hydrological models to extend their capabilities to predict in ungauged rivers (He et al., 2011) have not considered the explicit role of river network and reservoir features on water budgets.

Finally, many of the continental-scale applications have emphasized natural hydrological processes based on monitoring in undeveloped headwater catchments (Archfield et al., 2015), without explicit components to separately account for cultural influences (e.g., Wolock & McCabe, 1999; Weiskel et al., 2014; Bock et al., 2016; Beck et al., 2016; Newman et al., 2015; Jehn et al., 2019; Livneh & Lettenmaier, 2013). Whereas the focus on undeveloped catchments is necessary to accurately describe natural water cycling processes, an explicit accounting of the effects of anthropogenic activities on streamflow is needed to advance process

understanding across large scales and to improve the management utility of hydrologic models. Although land use has been included in large-scale hydrological models with regionalized parameters in Europe (e.g., Hundecha et al., 2016; Rakovec et al., 2016) and in global hydrological models (e.g., Arheimer et al., 2019), these have been typically characterized using coarser spatial resolutions. Collectively, these limitations have contributed to uncertainties at continental spatial scales in both identifying the causes of streamflow responses and predicting streamflow in more developed watersheds and river basins (e.g., Bock et al., 2016; Rakovec et al., 2016). In all hydrological models, challenges remain in obtaining detailed water use and management data (Archfield et al., 2015) that would extend the predictive capabilities of the land use observations.

Although progress has been made on each of these issues, no single modeling approach has addressed all the limitations, with considerable gaps in applications and understanding remaining across continental scales. Methods are especially needed that improve the parameterizations of large-scale hydrological processes and spatial scaling properties, advances that can narrow the information divide between catchment-scale modeling that has focused on relatively undeveloped catchments and global models that have more limited spatial and process specificity (Archfield et al., 2015). Toward addressing these needs for predicting streamflow over continental scales, we present the first SPARROW (Spatially Referenced Regression On Watershed attributes) model application for quantifying long-term mean annual streamflow in rivers and streams across the CONUS. SPARROW is a spatially explicit, hybrid (statistical-mechanistic) mass-balance modeling approach with a parsimonious level of process complexity that is informed by nonlinear statistical estimation of model parameters using observed geospatial and stream measurements (Smith et al., 1997). SPARROW modeling techniques have been widely used over the past two decades to predict stream water quality across regional and CONUS scales and to quantify the major factors controlling spatial variation in hydrochemical and sediment loads in streams (e.g., Preston et al., 2011; Alexander et al., 2008, 2000; Ator et al., 2011, Brakebill et al., 2010; Anning & Flynn, 2014; Miller et al., 2017) and base flow in the Upper Colorado watershed (Miller et al., 2016). The hybrid features of SPARROW have been shown to improve the accuracy of the predictions and process interpretations in water quality models (e.g., Smith et al., 1997; Alexander et al., 2000, 2007, 2008; García et al., 2016; Preston et al., 2011; Schmadel et al., 2018, 2019).

Our approach represents an initial step toward improving understanding of large-scale hydrological processes and scaling properties by providing a spatially and structurally consistent, continental-scale modeling framework with an intermediate level of spatial complexity. In our approach, the influences of catchment-scale runoff processes were held constant (spatially and temporally) in the model estimation, thereby allowing an improved quantification of the effects of natural and cultural processes that explain spatial variability in streamflow across large watersheds and in streams and reservoirs of CONUS river networks. We coupled the SPARROW river network modeling structure (Figure 1a; ~62,000 reach watersheds) with previous catchment-scale (1 km) water balance predictions of “natural” unit area runoff (Wolock & McCabe, 2018) that are inclusive of major water cycling processes (e.g., precipitation, evapotranspiration, soil moisture, subsurface storage, and release; Figure 1b), conditioned on prevailing climate (precipitation and temperature) and soil moisture conditions. In the SPARROW model, the small-scale predictions of unit runoff were combined with additional explanatory variables (e.g., soils, vegetation, land use, topography, water losses in streams, and reservoirs) to statistically estimate the appreciable effects of natural and cultural water supply and demand processes that operate over large spatial scales and explain the observed streamflow at monitoring sites in CONUS watersheds (Figure 2). The sites monitor flow conditions in highly diverse environments with widely varying climate and watershed conditions, ranging from those in small undeveloped catchments to the more developed land uses in small to large watersheds. We demonstrate that accounting for the effects of these large-scale processes significantly improved the accuracy of the mean streamflow predictions in CONUS rivers. A major advance is the explicit representation of streams and reservoirs in the model and their simultaneous estimation with climatic and landscape effects; this revealed a nonlinear surface area-dependent scaling of water loss in river networks that accounts for the cumulative effects of water removal on streamflow. Our hybrid modeling approach has the objective of providing a spatially consistent estimation of hydrological model parameters and uncertainties, thereby advancing the methods for transferring hydrological information to ungauged locations in river networks, especially those in larger and more culturally diverse watersheds.

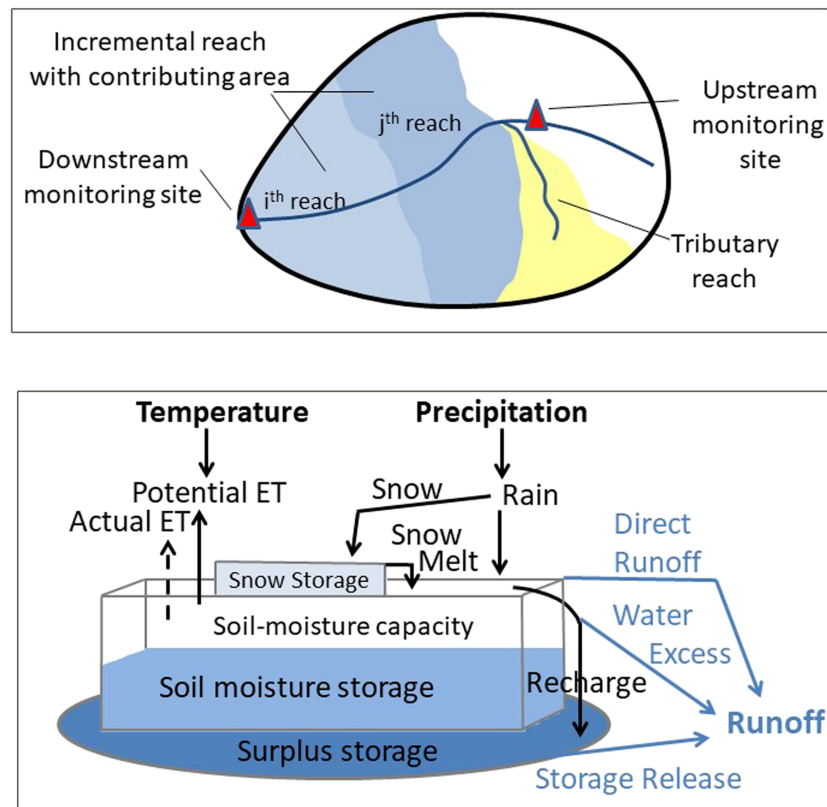


Figure 1. SPARROW streamflow conceptual model: (a) Illustration of the SPARROW reach network features for four incremental catchments; the SPARROW CONUS river network consists of 62,000 reach watersheds; (b) Thornthwaite monthly unit area runoff model and water cycling components, applied previously to 1-km cells in the conterminous United States (Wolock & McCabe, 2018); unit area runoff is the sum of three water surplus generation processes: direct runoff, soil water excess, and release from subsurface storage. For input to SPARROW, the 1-km monthly unit runoff values are summed for each year, with the multiyear annual values averaged across the 1-km cells in each incremental reach contributing area.

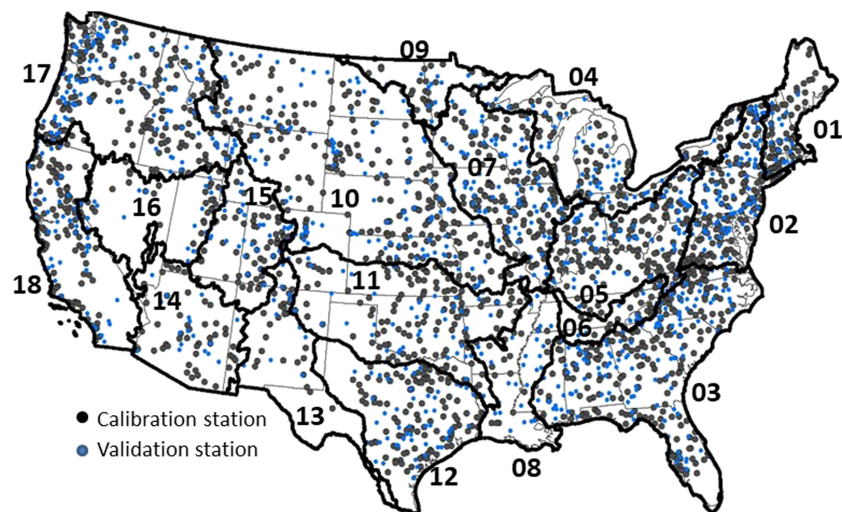


Figure 2. Monitoring sites for the SPARROW mean annual streamflow model. The map includes 1,778 calibration sites and 890 validation sites in the conterminous United States; boundaries identification numbers are shown for 18 Hydrologic Unit Code (HUC-2) regions.

Our study, presented here in the first of two companion studies, establishes a CONUS-scale SPARROW streamflow model with statistically unique explanatory variables, based on the use of spatially constant coefficients and model error variance. Results from this study as well as open-source SPARROW code are provided in the U.S. Geological Survey software release (Alexander & Schwarz, 2019; see also Alexander & Gorman Sanisaca, 2019), aiding in applications of our hybrid modeling approach to new studies or regions. Our initial study provides the foundational model that is used in the second study (Alexander et al., 2019) to evaluate and quantify hierarchical (regionally varying) effects on the model coefficients and prediction uncertainties, including streamflow measurement-related uncertainties. The analysis in the second study confirms the importance of the explanatory variables that we identified for the CONUS model and further demonstrates that spatial refinements in the model coefficients and uncertainties using hierarchical Bayesian methods improves the accuracy of the model predictions across regional and subregional scales.

2. Methods

2.1. Streamflow Conceptual Model

The foundational conceptual model for long-term mean annual streamflow is based on the standard spatially explicit SPARROW model for water quality mass loadings (Smith et al., 1997; Schwarz et al., 2006). The model structure separately represents three phases of streamflow: water generation, landscape and subsurface transport to water bodies, and stream and reservoir transport and loss processes (Figure 1a). These phases of streamflow generation and transport within the river network are specified according to a hybrid combination of process-related components and statistically estimated explanatory variables and model uncertainties. The spatial infrastructure for the streamflow model is defined by approximately 62,000 vector-based, one-dimensional stream channel flowlines, based on the U.S. Environmental Protection Agency CONUS River Reach File 1 (RF1) 1:500,000 scale (Nolan et al., 2002; see section 2.4). Note that the model equations presented in this section differ modestly in their syntax and structure from those presented in the SPARROW documentation (Schwarz et al., 2006).

In the model, long-term mean annual streamflow or water volume per unit time at the outlet of reach i is denoted V_i , and the mean water volume per unit time generated in the incremental area of reach i is denoted V_i^I . Incremental generated flow is expressed as a linear function of the “natural” unit area runoff source, R_i , the multiyear average of the annual sums of previously reported monthly unit area runoff for a 1-km grid for the CONUS (Wolock & McCabe, 2018). The unit area runoff, based on a prior application of monthly Thornthwaite (Thornthwaite & Mather, 1955) water balance conceptual models (Wolock & McCabe, 1999, 2018) to 1-km cells in the incremental catchment area of each reach (Figure 1b), describes the expected “natural” water runoff for the incremental drainage area (inclusive of inputs from subsurface water storage), conditioned on the prevailing climate (precipitation, temperature) and soil moisture conditions in the catchment, the principle variables in the Thornthwaite model. The Thornthwaite (Thornthwaite & Mather, 1955) model applies fundamental continuity principles to account for the primary factors that govern water cycling processes (Figure 1b). Unit runoff is predicted as a function of precipitation (P) minus energy demand from evapotranspiration (ET) and changes in subsurface water storage (S) that occur in response to changes in soil moisture conditions. Monthly unit runoff is calculated as the water surplus that results after the monthly climatic demand for water has been satisfied, based on the combination of water releases (contemporaneous and antecedent) from three sources, including direct runoff, water excess from soil storage, and the release of antecedent subsurface surplus water (Figure 1b). The potential ET (PET) is driven by the effects of solar radiation on evaporation and transpiration, which are estimated as a function of mean temperature and daylight hours for each month (Hamon, 1961). Actual ET (AET) depends on the available water from precipitation and soil moisture stores. Additional details of the previous implementation of the Thornthwaite model are given in Wolock and McCabe (1999, 2018).

Additional incremental flow is generated from Canadian regions outside the model domain. This flow is assumed to be proportional to foreign area within the catchment, denoted Y_i , with scaling coefficient, ψ , representing the flow yield from foreign drainage. Foreign drainage from Mexico is assumed to yield zero flow.

Unit runoff is scaled prior to entering the stream network at reach i by a reach-specific factor expressed as a function of M land-to-water delivery variables, Z_{mi} , $m = 1, \dots, M$, representing intensive measures of

incremental reach catchment properties. The specific functional relation between incremental flow generated from runoff in reach catchment i , modified by the land-to-water variables and external flow is given by

$$V_i^I = \alpha R_i \exp\left(\sum_m \theta_m Z_{mi}\right) + \psi Y_i \quad (1)$$

where α is a positive unit runoff source coefficient, θ_m are land-to-water variable-specific coefficients that mediate the relation between the land-to-water variables and the delivery of runoff to streams, and ψ is a fixed coefficient used to scale external flow. The θ_m coefficients can be interpreted as the rate of change of incremental generated flow from a unit increase in land-to-water variable Z_m . Consequently, θ_m has units that are the reciprocal of the units for Z_m . If Z_m is a log-transform of a catchment intensity property, then θ_m is the rate of change of incremental generated flow from a unit rate of change in the catchment intensity property, implying θ_m is unitless. To improve the interpretation of the source coefficient, α , the land-to-water variables are expressed as differences from their respective averages over the entire study area. Under this transformation, if the land-to-water variables are each symmetrically distributed, then the source coefficient represents the median share of incremental generated runoff that is delivered to the stream network. This transformation of the land-to-water variables is useful in the model specification phase of the analysis because the source coefficient tends to remain relatively constant as different land-to-water variables are included in the model. The land-to-water delivery variables account for additional natural and anthropogenic influences on water availability and demand that operate over large scales and are not explicitly included in the 1-km Thornthwaite model. These include natural climatic and landscape properties and cultural attributes, such as land use, vegetation, soils, topography, geology, and ancillary climatic factors (e.g., dryness).

In addition to the incremental contribution to flow (equation (1)), mean streamflow V_i also depends on the sum of water volumes from conflating reaches immediately upstream of reach i , \tilde{V}_j , for all $j \in J_i$, where J_i is the set of conflating reaches upstream of reach i (Figure 1a). For a purely dendritic network, the set J_i will everywhere consist of only two reaches, but any number of conflating streams is allowed. Not all flow at the upstream end of reach i is delivered to its outlet. If reach i represents a point of divergence in the reach network, then only a fraction, d_i , of the upstream flow is diverted to reach i , this fraction being defined as an attribute of the reach network. For a purely dendritic network, d_i would equal one for all reaches; for a network with divergences, the sum of d_i across all flowlines at a diversion equals one.

Additionally, the model allows for fractional delivery of upstream flow to the downstream outlet, this fraction being an inverse function of stream and reservoir attenuation variables, X_i^S and X_i^R , with associated mass transfer coefficients δ^S and δ^R . Water attenuation or loss can be caused by direct evaporation from the water surface, recharge to the subsurface, or transpiration by riparian vegetation. First-order surficial exchange processes such as these can be expressed as continuous functions of inverse areal hydraulic load with units of time length⁻¹ (Schwarz et al., 2006; Alexander et al., 2008).

The specific functional form assumed for the aquatic delivery fraction, denoted A_i , is

$$A_i(\delta^S, \delta^R, X_i^S, X_i^R) = \begin{cases} \exp\left(-\delta^S \frac{T_i^S}{D_i}\right), & \text{if reach } i \text{ is a stream,} \\ \frac{1}{1 + \delta^R H_i^{-1}}, & \text{if reach } i \text{ is a reservoir,} \end{cases} \quad (2)$$

where $X_i^S = T_i^S/D_i$, T_i^S being travel time in reach i and D_i being reach i 's depth; and $X_i^R = H_i^{-1}$, H_i^{-1} being the inverse areal hydraulic load, areal hydraulic load being the ratio of flow leaving the reservoir to surface area of the reservoir. The stream attenuation variable, X_i^S is an alternative representation of inverse areal hydraulic load, expressed to be independent of the surface area of the stream reach segment. Both X_i^S and X_i^R are expressed in units of time length⁻¹, such that the coefficients δ^S and δ^R are in units of length time⁻¹ (i.e., velocity). Additionally, incremental loading, V_i^I receives an areal adjustment, A_i' , prior to delivery to reach

i 's outlet, given that conceptually the reach incremental flow is assumed to be delivered to the reach mid-point. Accordingly, the incremental load delivery fraction for reaches equals the square root of the full reach delivery fraction ($A'_i = \sqrt{A_i}$). Conversely, incremental load entering the network at a reservoir or lake is assumed to be well mixed within the reach, so that for reaches classified as a reservoir or lake $A'_i = A_i$.

Modeled flow leaving reach i is expressed as the delivered sum of upstream flow routed to reach i , plus incremental load delivered to reach i 's outlet, implying

$$V_i(\beta) = A'_i(\delta) V_i^I(\alpha, \psi, \theta) + d_i A_i(\delta) \sum_{j \in J_i} \tilde{V}_j(\beta), \quad (3)$$

where $\delta = \{\delta^S, \delta^R\}$, $\theta = \{\theta_1, \dots, \theta_M\}$, and $\beta = \{\alpha, \psi, \delta, \theta\}$.

Equation (3) is processed in downstream order starting with the headwater reaches. In model estimation, the flow in the conflating upstream reaches takes one of two values:

$$\tilde{V}_j(\beta) = \begin{cases} V_{s(j)}^M, & \text{if flow in reach } j \text{ is monitored by station } s(j), \\ V_j, & \text{otherwise,} \end{cases} \quad (4)$$

where $V_{s(j)}^M$ denotes the monitored value of flow for station $s(j)$ at reach j , and V_j is flow determined from the processing of equation (3) in upstream flowlines. The substitution of monitored flow for modeled flow at monitored reaches implies that downstream values of modeled flow are conditioned on upstream monitored values, per conventional SPARROW methods (Schwarz et al., 2006). Modeled (simulated) streamflow predictions are denoted as “unconditioned” predictions.

If reach i is monitored by station $s(i)$, then model estimation relates the natural logarithm of monitored flow to the logarithm of conditioned modeled flow, inclusive of an additive model residual,

$$\ln V_{s(i)}^M = \ln(V_i(\beta)) + \epsilon_{s(i)}, \quad (5)$$

where $\epsilon_{s(i)}$ is the residual for the nested basin delineated by the monitoring station $s(i)$ at reach i and any upstream monitoring stations that are used to condition the modeled flow in estimation. In the baseline SPARROW model, $\epsilon_{s(i)}$ is assumed to be normally distributed and independent across nested basins with a mean of zero and common variance, σ_ϵ^2 .

2.2. Data and Model Estimation Procedures

The SPARROW streamflow model is applied to $8.1 \times 10^6 \text{ km}^2$ of drainage in the CONUS, using the RF1 river network with approximately 62,000 stream segments (mean reach catchment size = 60 km^2) and drainage areas from 1-km topography (Nolan et al., 2002). Estimates of the 11-year mean annual streamflow, the model response variable, were computed for gauging sites with complete records (no missing data) during the period 1997 to 2007 and located on RF1 stream reaches. The sites were obtained from a national inventory of daily streamflow gaging sites for the period 1900 to 2013, consisting of 23,739 daily flow gauges (Wolock & McCabe, 2018); the sites are operated by the U.S. Geological Survey (USGS) and publicly available from the USGS National Water Information System (NWIS) database (U.S. Geological Survey, 1994). The requirement of complete flow records had the objective of minimizing the effects of missing data and sampling-related influences on mean annual streamflow; this also provided a controlled set of observations for evaluating the hierarchical Bayesian methods in the companion paper (Alexander et al., 2019). The final 2,668 sites were selected from an initial set of 4,307 sites with complete records from 1997 to 2007 and RF1 reach identifiers. The final sites resulted from eliminating sites with more than 10% foreign drainage area (Canadian and Mexican areas that lack geospatial data), sites with known water diversions or agricultural return flows that are not accounted for in the model, sites on headwater reaches where the area is appreciably smaller than the reach drainage area, and sites in proximity to one another (a minimum river distance was imposed to ensure >5 stream reaches between sites; $\sim 300 \text{ km}^2$). From this set, two thirds of the sites (1,778) were randomly selected for model estimation, with the remaining sites (890) used for model validation (Figure 2). The median incremental drainage size is $1,250 \text{ km}^2$ for the 1,778 calibration sites (interquartile range = 475 to $3,225 \text{ km}^2$).

Monthly water balance estimates of unit runoff, PET, AET, and other water cycling components were available from a previous application (Wolock & McCabe, 2018) of Thornthwaite water balance models for each of 64 million 1-km grid cells for the CONUS. The models require monthly inputs of precipitation and temperature and estimates of the soil moisture storage capacity and assume that the unit runoff is 50% of the sum of excess contemporaneous precipitation and antecedent subsurface water storage. Long-term mean annual estimates of unit runoff for input to the SPARROW model were calculated as a mean of the annual sums of the monthly runoff volumes over 1997–2007. Additional explanatory climate variables evaluated in SPARROW included a measure of the runoff loss potential or aridity, based on the difference between PET and AET, and seasonal distributions of the annual precipitation volume.

Additional exploratory predictor variables for natural and cultural features of the catchments were obtained from a previous SPARROW modeling study for RF1 reaches in the CONUS (Anning & Flynn, 2014). These included variables for land cover/land use (National Land Cover Data for 21 classes for 2001, irrigation, tile drainage, population density, municipal and industrial water use, and vegetation index), soil properties (permeability, clay, sand, silt, and hydrological soils classification index), surface and subsurface hydrology/morphology/topography (baseflow index, saturation-excess and infiltration-excess overland flow, channel sinuosity, slope, and basin shape index), and snowfall frequency. Measures of the stream and reservoir hydraulic properties were obtained from RF1 network data (Nolan et al., 2002), including estimates of the channel water time of travel, calculated as a quotient of the mean annual stream water velocity and the stream channel length, and the areal hydraulic load for reservoirs (ratio of the outflow to the surface area). Stream channel depth was estimated as a function of the mean annual streamflow (Alexander et al., 2000). See Table S1 in the supporting information for a detailed listing of the exploratory explanatory variables.

Coefficients (α , δ^S , δ^R , θ_m) of the SPARROW streamflow models are estimated using nonlinear least squares (NLS) Marquardt-Levenburg methods, using the *nlfb* function in the *nlmrt* R package with constrained optimization (Nash, 2015). Model estimation uses the conditioned model predictions (Schwarz et al., 2006), which are calculated by substituting the observed mean annual streamflow for the model predicted streamflow on reaches with monitoring sites. Conditioned predictions provide the most accurate reach predictions to quantify streamflow and estimate model coefficients that describe the effects of natural and cultural processes on water availability and transport. Standard errors of the coefficients are estimated according to second-order Hessian methods. Model results were also found to be consistent with those obtained using the SAS SPARROW software (Schwarz et al., 2006). Decisions to accept or reject explanatory variables were informed by the physical interpretability of competing models (Schwarz et al., 2006; Alexander et al., 2008) and by evaluating multiple model diagnostics, including measures of the model accuracy—R-squared, root-mean-square error (RMSE; reported for the log-transformed scale), AIC (Akaike Information Criterion; Akaike, 1974), and prediction bias (Moriassi et al., 2007), and the statistical significance ($\alpha = 0.05$) and collinearity of the coefficients based on variance inflation factor (VIF) metrics (Schwarz et al., 2006). Additional diagnostics of model performance and interpretability include regional measures of model accuracy for the calibration sites and assessments of the model error assumptions (i.e., independent and homoscedastic residuals), based on Moran's I tests for spatial autocorrelation (Li et al., 2007; *spdep* R package, Bivand, 2015). The models are evaluated for convergence to a global minimum error by ensuring that the use of altered initial values yielded the same model coefficient metrics. A regional classification of the calibration sites according to the Hydrologic Unit Classification level-2 regions (Figure 2; Seaber et al., 1987) is used to assess spatial variability in the model prediction bias and precision; this classification has been featured in previous studies of annual streamflow (e.g., Vogel et al., 1998) for the 18 CONUS regions, six regions are associated with 20 to 50 sites, while seven regions have more than 100 sites.

Our strategy for variable selection in the SPARROW models involved comparing the performance and diagnostics of alternative models of varying complexity (type and number of variables; see Table 1), beginning with simple models that include either the Thornthwaite mean-annual predictions of “natural” unit runoff (Wolock & McCabe, 2018) or only the Thornthwaite model input variables (mean annual precipitation, temperature, and available soil water capacity). Subsequently, we added process complexity to these models to assess the importance of natural and anthropogenic properties of watersheds that are indicative of water supply and demand processes that operate over large spatial scales, which are not accounted for by the previous 1-km scale Thornthwaite unit runoff predictions. The added properties

Table 1
Evaluated Mean Annual SPARROW Streamflow Models for the CONUS

Model	Model explanatory variables				Model performance		
	Terrestrial water source	Land-to-water delivery	Aquatic loss	Number of variables	RMSE	Yield R^2	AIC
1	Precipitation	Temperature, AWC	No	3	1.139	0.582	5512
2	Runoff (fixed coef. = 1)	No	No	1	0.811	0.719	4303
3	Runoff (estimated coef.)	No	No	1	0.715	0.782	3856
4	Runoff (HUC-2 regional estimated coefs.)	No	No	18	0.625	0.835	3396
5	Runoff (estimated coef.)	All variables	No	14	0.541	0.876	2879
6	Runoff (fixed coef. = 1)	All variables	All variables	16	0.496	0.896	2570
7	Runoff (estimated coef.)	All variables	All variables	16	0.493	0.897	2545

Note. The terrestrial “runoff” water data source is the 1-km catchment-scale Thornthwaite water balance model predictions of unit runoff [Wolock & McCabe, 2018], aggregated for the incremental drainages of the 62,000 reaches (see methods). Coefficients are estimated using SPARROW NLLS (nonlinear least squares) methods for all explanatory variables, except for the unit runoff coefficient in models 2 and 6, which is fixed to a value of 1.0. “All variables” refers to the 12 land-to-water delivery explanatory variables or the 2 aquatic loss explanatory variables (streams, reservoirs) included in the final model 7 (see Table 2). Models 6 and 7 include water inputs from Canadian drainages. Performance metrics are based on use of unconditioned predictions of mean annual streamflow: RMSE (root-mean-square error), water-yield R-squared, and AIC (Akaike Information Criterion; models with smaller values have higher prediction accuracy). “AWC” is available water capacity. “HUC” is hydrologic unit code classification of CONUS drainage (see Figure 2).

Table 2
Results of the Nonlinear Least Squares (NLLS) Mean Annual Streamflow Model for the CONUS

Model component and explanatory variable	Variable units	Coef. units	Process category	NLLS estimates				
				Mean estimate	Standard error	<i>t</i> stat	<i>p</i> value	VIF
<i>Source</i>								
Unit runoff, α	m ³ /yr	dimenless.	Water balance	0.8272	0.0241	34.39	<0.0001	3.99
Canadian water yield, ψ	km ²	m ³ /km ² /yr	Water balance	7480	3851	1.94	0.0523	1.02
<i>Land-water delivery, θ_m</i>								
Forested land	percent	percent ^{−1}	Land cover/use	−0.0021	0.0008	−2.58	0.0100	1.64
Urban land	percent	percent ^{−1}	Land cover/use	0.0035	0.0011	3.10	0.0019	1.16
Vegetation index	dimenless.	dimenless.	Land cover/use	0.0136	0.0021	6.46	<0.0001	2.82
Farm tile drains	percent	percent ^{−1}	Land cover/use	0.0033	0.0010	3.41	0.0066	1.18
Stream density (log)	m km ²	dimenless.	Hydrology/Topog.	0.2839	0.0323	8.78	<0.0001	2.02
Channel sinuosity (log)	m/m	dimenless.	Hydrology/Topog.	−0.3806	0.1120	−3.40	0.0007	1.09
Land slope	degrees	degrees ^{−1}	Hydrology/Topog.	0.0183	0.0032	5.68	<0.0001	1.75
Soil clay content	percent	percent ^{−1}	Soils	0.0072	0.0014	5.00	<0.0001	1.15
Hydrologic soil group B	percent	percent ^{−1}	Soils	−0.0054	0.0007	−8.17	<0.0001	1.45
Hydrologic soil group C	percent	percent ^{−1}	Soils	−0.0051	0.0009	−5.92	<0.0001	1.50
PET and AET difference	% of PET	percent ^{−1}	Climate	−0.9565	0.1412	−6.77	<0.0001	2.47
Seasonal precipitation (May-June)	mm/yr	yr/mm	Climate	0.0204	0.0028	7.27	<0.0001	1.74
<i>Stream water losses</i>								
Streams, δ^S	day/m	m/day	Stream Corridor	0.0120	0.0011	10.64	<0.0001	2.08
<i>Reservoir water losses</i>								
Reservoirs, δ^R	yr/m	m/yr	Reservoirs	0.4919	0.1610	3.06	0.0023	1.11
<i>Model metrics</i>								
N observations	Calibration	Simulation	Validation					
	1,778	1,778	890					
RMSE (log m3/year)	0.4635	0.4925	0.5477					
R-squared yield (log units)	0.9090	0.8973	0.8689					
Prediction bias (%)	3.9	10.6	13.1					

Note. The model is defined in equation (3) and shown as Model 7 in Table 1. “dimenless.” = dimensionless; RMSE = root-mean-square error; Prediction bias is expressed as the quotient of the difference between the observed and predicted water volume and the observed volume. “Model metrics” are reported for “Calibration” (conditioned, monitoring-substituted predictions), “Simulation” (unconditioned predictions at the calibration sites), and “Validation” (unconditioned predictions at the validation sites). Unconditioned predictions are based on model execution in simulation mode without substitution of the monitored loads, using mean coefficients from the calibrated model, and are preferred for use in assessing the predictive skill of the model. Estimates of the Canadian water yield are sensitive to the initial value, although the choice has negligible effects on other model estimates; the final model is estimated using an initial value of $7,500 \text{ m}^3/\text{km}^2$, which is associated with the lowest model RMSE and highest statistical significance among the evaluated models with initial water yield values ranging from 3,000 to 9,000.

(Table 2; Table S1) include land-to-water delivery variables classified according to four general process categories (land cover and use, surface and subsurface hydrology and topography, soils, and climate), and river network properties that account for water loss in and along stream channels and reservoirs.

As a relative measure of the magnitude of the changes in natural runoff that are reflected in the streamflow predictions of the final model (Model 7; Table 1), which account for large-scale effects of natural and anthropogenic water supply and demand processes, we report the SPARROW Water Balance Accounting Ratio (*SWBAR*) for the *s*th calibration station and *k*th region of the 18 HUC-2 hydrological regions. This is defined as

$$SWBAR_{k,s} = \frac{V_{k,s}^C}{V_{k,s}^S}, \quad (6)$$

where $V_{k,s}^C$ and $V_{k,s}^S$ are the mean annual unconditioned streamflow volumes for the final estimated SPARROW model (C, complex; Model 7, Table 1) and the simple unit runoff model (S; Model 2 and Table 1). The simple model has a fixed (nonestimated) unit runoff coefficient, such that the model only reflects the 1-km scale natural water balance accounting processes contained in the previously estimated Thornthwaite predictions (Wolock & McCabe, 2018); these are “upscaled” by summing unit runoff throughout the river network without any statistical adjustment or accounting of large-scale processes. The explanatory variables and SPARROW statistical estimation of the complex model (Model 7; Table 2) provides an accounting of additional natural and cultural effects on the water balance over large scales. Thus, *SWBAR* provides a cumulative measure of the net change in water flows that occurs in the drainage above each monitoring site to account for large-scale water supply and demand processes, as quantified by the final complex model; below one *SWBAR* values indicate net water losses, whereas above one values indicate net water gains. Note that the unconditioned mean annual flow volumes are statistically adjusted to account for log retransformation bias (Schwarz et al., 2006). Although monitoring data are not required to compute *SWBAR*, the statistic was evaluated at calibration sites, given their location in larger watersheds that are more sensitive to large-scale processes that affect the aggregation of flow in the reach network, processes that are omitted from the natural runoff model. Therefore, *SWBAR* can be used to assess the potential weaknesses of simple natural runoff models when applied over larger geographic scales.

As a complementary metric to *SWBAR*, we report the COMparative Prediction Accuracy Statistic (COMPAS), a new station-specific metric of the relative differences in the magnitude of the prediction accuracy of nested models. It provides a measure of the accuracy of the SPARROW predictions of mean annual streamflow (based on Model 7; Table 1), attributable to large-scale water supply and demand processes that are omitted by the model predictions of natural streamflow (Model 2; Table 1), allowing further consideration of SPARROW prediction accuracy in undeveloped and developed basins of varying sizes across the CONUS. COMPAS is computed as the log ratio of the differences between the observed (O_s) and predicted (P_s) unconditioned streamflow values for the complex (C, Model 7) and simple (S, Model 2) models, defined as

$$COMPAS_s = \ln(O_s - P_s^C)^2 - \ln(O_s - P_s^S)^2 \quad (7)$$

for the *s*th station. A negative value of the COMPAS indicates a higher relative accuracy for the complex model.

3. Results

3.1. Evaluation of the Mean Annual Streamflow Models

In Table 1, we present a summary of the evaluated models with varying levels of complexity. The more complex models (Models 5–7) include the final selection of land-to-water delivery and aquatic loss variables. Among the evaluated models, Model 7 (the final selected model) has the lowest overall mean error, RMSE, highest R-squared for water yield, and smallest (most accurate) AIC value; details on the explanatory variables and performance metrics for this model are reported in Table 2. Model 7 includes unit runoff as the primary terrestrial water source, with small estimated water inputs from Canadian drainages ($7,480 \text{ m}^3 \cdot \text{km}^{-2} \cdot \text{year}^{-1}$; 7.5 mm/year) and 12 land-to-water delivery variables, associated with four

general process categories (Table 2) that control water availability and delivery to streams; water transport in streams and reservoirs is subjected to losses based on the estimated mass transfer rate coefficients. All explanatory variables in the final model are observed to have uniquely estimated coefficients (i.e., statistically separable from zero) that are uncorrelated with one another, based on their high levels of statistical significance and relatively low values (Schwarz et al., 2006) reported for the VIF multicollinearity metric (Table 2); the explanatory variables are statistically significant with $p < 0.01$, except for the Canadian water yield variable ($p = 0.052$) that is included to account for water sources outside of the CONUS.

Results for the evaluated models (Table 1) quantify the incremental improvements in model performance (RMSE, R-squared, AIC) that occur with the addition of process complexity. Beginning with the simplest models, the model that includes only the Thornthwaite water balance inputs (precipitation, temperature, and available water capacity; Model 1; Figure S1a) as explanatory variables is significantly exceeded in accuracy by accounting for natural water cycling processes that are present in the Thornthwaite unit runoff predictions (Models 2 and 3); for example, the RMSE is about 30% to 40% lower for the unit runoff models. Furthermore, the unit runoff model (Model 3; Figure S1b), with a statistically estimated coefficient, shows higher prediction accuracy (lower RMSE and higher R-squared) as compared to that for the unit runoff model (Model 2) with a fixed (nonestimated) coefficient value of 1.0; for example, the RMSE is about 12% lower for Model 3. Adding land-to-water delivery variables (Model 5) to the unit runoff model (Model 3) accounts for large-scale water supply and demand processes associated with the landscape that control water inputs to streams and reservoirs; these additional variables result in a 25% reduction in the model RMSE, with about a 12% increase in the water-yield R-squared. This model (5), with spatially constant coefficients for unit runoff and land-to-water delivery variables, also outperforms a unit runoff model with regionally variable coefficients (Model 4); for example, Model 5 has a 14% smaller RMSE than that for Model 4. Finally, the addition of the aquatic variables to account for in-stream and reservoir loss processes (Model 7) further reduces the model RMSE by nearly 10%, compared to that associated with Model 5, which lacks any explicit accounting for aquatic transport. The final model (Model 7), with 16 explanatory variables, also shows a slightly higher prediction accuracy compared to that for Model 6, which has the same explanatory variables but a fixed (nonestimated) unit runoff coefficient. Thus, statistical adjustment of the small-scale Thornthwaite unit runoff predictions as provided in Model 7 offers slightly improved predictive power over large spatial scales.

The selection of variables in the final model (Table 2) resulted from exploratory evaluations of streamflow models with different combinations of explanatory variables from a set of 38 measures of watershed properties (see Table S1 in the supporting information). The final selected variables (Table 2) were found to be the most statistically significant ($\alpha = 0.05$) and are uniquely identifiable among the alternative models, with most variables excluded from the final model observed to have very low levels of statistical significance ($p > 0.30$). In Table S1, we also report Spearman-Rho rank-order measures of monotonic correlation between all the explanatory variables for correlations above 0.20; this provides an approximate indicator of statistical similarities between the spatial patterns of the final selected variables and those for watershed properties found to be statistically less important predictors of mean annual streamflow. For example, negative statistical correlations were common in the spatial patterns among “closed” variables (e.g., land-use and soil content percentages), whereas positive and negative correlations were observed for the spatial distributions of hydrological and landscape metrics (e.g., saturation-excess or infiltration excess overland flow) that are derived from one or more controlling variables (e.g., depth to water). Our exploratory model evaluations included explanatory variables (Table S1) that represent a wide range of catchment properties (e.g., silt and sand content of soils, soil permeability) and estimates of specific processes (e.g., baseflow, tile drainage, saturation excess overland flow, municipal and irrigation water use) known to control streamflow. The monotonic correlations (Table S1) suggest that the explanatory variables identified in the final model show certain general similarities in their spatial patterns with those of various other streamflow-governing properties of catchments.

3.2. National and Regional Performance of the Final Model

Based on the final model error (Model 7 RMSE = 0.46 for conditioned predictions; Table 2), predictions of mean annual streamflow at the calibration sites are accurate to within $\pm 46\%$ (1 standard deviation of the mean) at the reach scale, with predictions showing an overall tendency for underestimation based on a

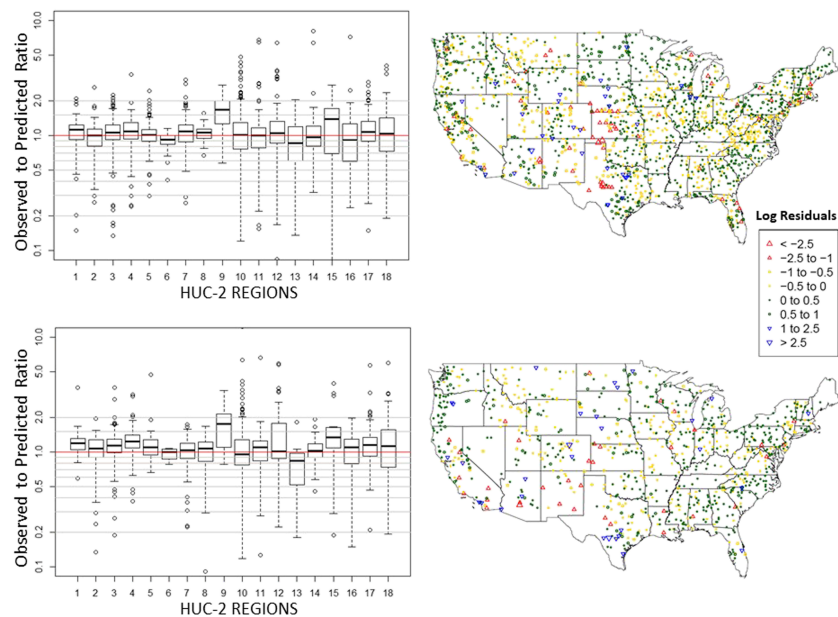


Figure 3. Calibration and validation results for the mean annual streamflow final model for the CONUS (Model 7; Table 1): (a) boxplots of observed to predicted streamflow for the calibrated model for stream reaches for unconditioned predictions by HUC-2 regions; (b) national map of model calibration log streamflow residuals for the unconditioned predictions (1,778 sites); (c) boxplots of observed to predicted water volume for the application of the model to stream reaches at validation sites by HUC-2 regions; (d) national map of model validation log residuals (890 sites). Residuals are computed as the observed log mean annual streamflow minus the predicted log mean annual streamflow.

positive mean prediction bias of 4%. Performance metrics associated with the unconditioned model predictions at calibration sites, indicate a slightly higher model error ($RMSE = 0.49$), larger prediction bias (10.6%), and similar water yield R -squared (0.90); these metrics provide an assessment of the general predictive skill of the model, independent of the effect of the monitoring adjusted (conditioned) predictions (Schwarz et al., 2006), which are most appropriate for use in estimating the model. Model performance metrics for the randomly sampled 890 validation sites show a 11% larger model error ($RMSE = 0.55$) and 24% larger average prediction bias (13.1%) as compared to the metrics reported for the simulation model predictions at calibration sites. For comparison with the model based on 1,778 calibration sites (Table 1), a model estimated using all the calibration and validation sites (2,668 sites) gives generally similar performance (unconditioned predictions: $RMSE = 0.51$, Bias = 9.3%); the explanatory variables also display similar coefficient significance levels, with mean coefficients typically within 10% to 25% of those in the final model. Additional details of the model residuals for the 1,778 calibration and 890 validation sites are presented in the supporting information (Figures S2 and S3).

Regional variations in prediction accuracy at the 1,778 calibration sites (Figure 3a and 3b) indicate a tendency toward underpredictions in most HUC-2 regions (12 regions under versus 6 over for the median ratio of the observed to predicted streamflow; 993 versus 785 sites, respectively), based on the observed to predicted streamflow ratios (Figure 4a) and model residuals for the unconditioned predictions (Figure 3b). In eastern regions (HUC-2 regions 1 to 8), the interquartile ranges of the observed to predicted ratios fall between 0.8 and 1.3 for all regions; region 6 (Tennessee) shows a preponderance of overpredictions, whereas underpredictions are common in Regions 1 (New England), 3 (Southeast-Gulf), 4 (Great Lakes), 5 (Ohio), 7 (Upper Mississippi), and 8 (Lower Mississippi); median ratios are typically between 0.9 and 1.1. In western regions (9 to 18), the spatial variability in the observed to predicted ratios are about 10% to 15% larger than observed for eastern regions, with interquartile ranges falling between 0.7 and 1.4, indicating the lower precision of the model in these regions; median ratios are typically between 0.9 and 1.1. Underpredictions of streamflow are more common in Regions 9 (Red-Rainy), 12 (Texas-Gulf), 15 (Lower Colorado), 17 (Pacific NW), and 18 (California), where the largest variability in prediction accuracy occurs. Clusters of underpredictions in streamflow (Figure 3b) are observed in watersheds in the midcontinent in the Dakotas, Nebraska, and Kansas in the north and in areas of Texas in the south. By contrast, clusters of overpredictions occur in

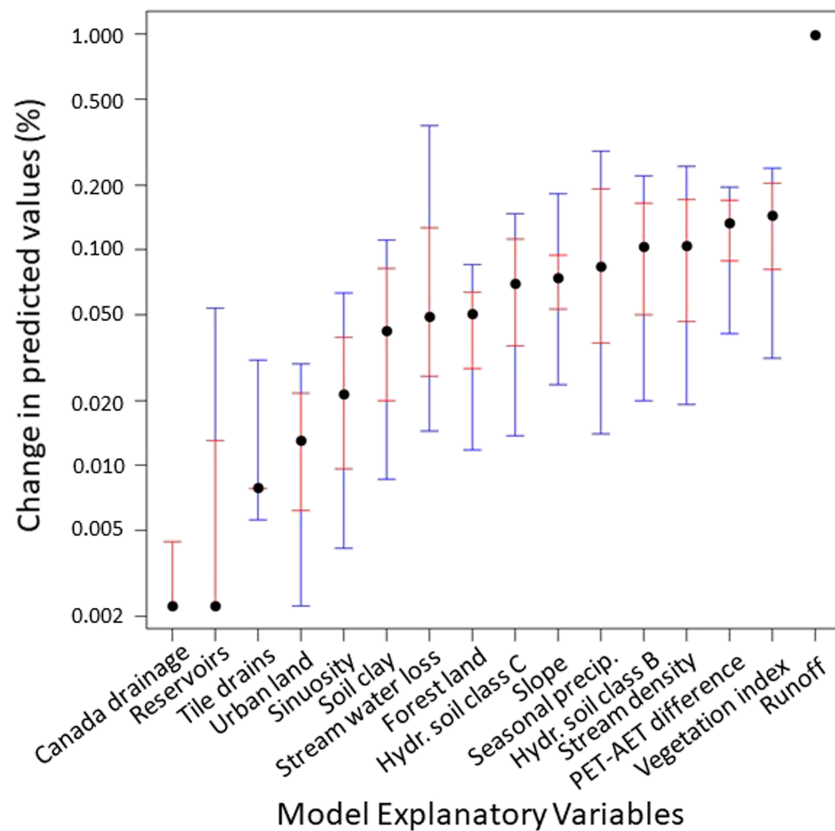


Figure 4. Sensitivity (percent change) in streamflow to a 1% change in the explanatory variable (inner interval = standard deviation; outer 95% interval) for the mean annual SPARROW streamflow model (Table 2). The sensitivities are summarized for streamflow predictions for reaches associated with the 1,778 calibration sites.

midcontinent watersheds in Nebraska, Oklahoma, and Texas as well as northern areas of California and in Florida. The areas of more extensive clustering of overprediction and underprediction occur in the midcontinent U.S. along the 100th longitudinal meridian, where an abrupt change occurs in annual precipitation volume that is associated with an abrupt shift from arid to humid climatic conditions (Seager et al., 2018).

For the validation sites ($n = 890$), regional variations in the model biases (Figures 3c and 3d) also indicate a tendency toward underpredictions in most regions (16 regions under versus 3 over; 549 versus 341 sites) but slightly more than observed for the calibration sites. In eastern regions, the interquartile range of the observed to predicted ratios typically extends from about 0.8 to 1.4, with regional medians observed in the range of 0.9 to about 1.2. In western regions, the observed to predicted ratios typically fall between 0.8 and 1.4 or higher for three regions (9, Red-Rainy; 12, Texas-Gulf; and 15, Lower Colorado). The clusters of underprediction and overprediction in streamflow (Figure 3d, triangles) show many similar spatial patterns that were observed for the calibration sites, with underpredictions in midcontinent areas of the Dakotas and Texas-Gulf and overpredictions in Kansas and Texas. There are also local underpredictions in similar locations in western regions as observed for the calibration sites.

A national test for spatial autocorrelation in the mean annual streamflow model residuals at the calibration sites (Moran's I with Euclidean distance weights; see Table S2) indicated a statistically significant ($\alpha = 0.10$) spatial correlation, with a tendency for the spatial clustering of overprediction and underprediction (positive test statistic). The test results were primarily influenced by regional positive spatial autocorrelation that was identified as statistically significant for five of the 18 HUC-2 regions (Table S2), with the Ohio (5), Missouri (10), and Texas-Gulf (12) showing the highest significance levels. These results suggest that the model residuals in these selected regions are not independently distributed as assumed; this may contribute to some

Table 3

Summary of comparative measures of prediction accuracy for the mean annual SPARROW streamflow model for the CONUS by region and site type classification for the 1,778 calibration sites

Region	Site type	Number of sites	Comparative model performance (COMPAS)		
			Median	25th	75th
East	HCDN	173	−0.265	−1.994	1.158
East	Non-HCDN	982	−0.456	−2.532	1.070
West	HCDN	87	0.058	−2.074	1.774
West	Non-HCDN	536	−2.613	−4.423	−0.627
East	Small basin	542	−0.146	−1.462	1.171
East	Large basin	613	−0.915	−3.449	1.001
West	Small basin	263	−1.343	−3.025	0.698
West	Large basin	360	−3.106	−4.770	−1.062
East	HCDN (Small basin)	130	−0.205	−1.776	0.960
East	Non-HCDN (Small basin)	412	−0.111	−1.388	1.211
West	HCDN (Small basin)	65	−0.034	−2.074	1.448
West	Non-HCDN (Small basin)	198	−1.712	−3.084	0.241

Note. Comparative performance is measured by the COMPAS metric (equation (7)) for the complex final streamflow model (Model 7; Table 1) and the simple natural runoff streamflow model (Model 2; Table 1); a negative value of COMPAS indicates a higher relative accuracy for the complex model. The “East” and “West” regions are defined according to the 100th CONUS longitudinal meridian. A total of 260 of the 1,778 calibration sites were previously identified as “HCDN” sites (Hydro-Climatic Data Network; Lins, 2012), with the complementary 1,518 sites classified as “Non-HCDN”. “Small” and “large” basins are defined according to total drainage areas below and above 1,154 km², the 75th percentile drainage area for the HCDN sites.

overestimation of the statistical significance of certain model coefficients and possible prediction biases in these regions.

Comparisons of the prediction accuracy of the final model (Model 7; Table 1) with that of the natural-runoff streamflow model (Model 2; Table 1), based on the upscaling of the 1-km predictions of natural runoff, provide a measure of the improved accuracy in SPARROW predictions of mean annual streamflow, attributable to natural and anthropogenic water balance processes that are omitted from the upscaled natural runoff streamflow predictions (Table 3). The model performance comparisons (Table 3) are based on values of the COMPAS metric (equation (7)), with negative values indicating a higher relative accuracy of the SPARROW streamflow model (Model 7; Table 1). To isolate the effects of human development and water use on the performance of the two models, COMPAS values are reported separately for 260 of the 1,778 calibration sites, previously identified as relatively undeveloped Hydro-Climatic Data Network (HCDN) sites (Lins, 2012), and for the complementary set of 1,518 more developed calibration sites. COMPAS values (Table 3) are additionally separated according to eastern and western stations, based on the 100th longitudinal meridian for the CONUS, a noted geographic division of humid and arid climatic and landscape conditions (Seager et al., 2018), and by drainage basin size as a measure of the geographic scale associated with the SPARROW final model accounting of large-scale natural and anthropogenic process effects on streamflow. The drainage size separation for “small basins” and “large basins” is defined by the 75th percentile drainage area of the HCDN sites (1,154 km²).

The results (Table 3) indicate that in the eastern CONUS the SPARROW streamflow model has a slightly higher median accuracy than the natural-runoff streamflow model at HCDN sites (−0.265) and a somewhat more pronounced median prediction accuracy for non-HCDN sites (−0.456). In the western CONUS, the SPARROW and natural-runoff models have nearly equal prediction accuracy at HCDN sites (median = 0.058), whereas the SPARROW model displays a large predictive advantage among non-HCDN sites (−2.613); the negative interquartile range for COMPAS values for non-HCDN sites in the West indicates that SPARROW has a clear advantage over the natural-runoff model in more arid climates where water losses via recharge and evapotranspiration and human alteration of streamflow are likely to be more pronounced. Results for the COMPAS performance metrics segregated by drainage basin size (“small” and “large”) indicate that prediction accuracy of the SPARROW streamflow model, relative to that of the natural-runoff model, is enhanced in both the east and west in large basins (median = −0.915 and −3.106) as compared to that in small basins (median = −0.146 and −1.343), with a more pronounced large-basin effect in western CONUS regions. Moreover, the comparatively higher accuracy of the SPARROW streamflow model relative to that of the natural-runoff model is also evident among the small basins for both undeveloped (HCDN) and developed (non-HCDN) sites, with evidence of a significantly higher prediction accuracy for the SPARROW streamflow model at non-HCDN sites in the western CONUS (median = −1.712).

3.3. Model Coefficients and Explanatory Variables

The model coefficients (Table 2) quantify statistically unique functional relations with mean annual streamflow. The unit runoff and the aquatic loss coefficients are constrained to be positive, with their observed high levels of statistical significance providing confirmation of the expected physical response of streamflow to spatial variations in landscape runoff and hydraulic characteristics of streams and reservoirs. The estimated stream and reservoir mass transfer coefficients are consistent with first-order net loss processes as specified by equation (2). Terrestrial water inputs to reach catchments, based on the aggregated previous predictions of natural Thornthwaite unit runoff (Wolock & McCabe, 2018), are mediated according to equation (1) by

the estimated unit runoff coefficient (0.83, Table 2) and the estimated effects of 12 land-to-water delivery variables. The land-to-water delivery coefficients can assume either positive or negative values, with the statistically estimated coefficients indicating the direction of the functional response of mean annual streamflow to spatial variations in each watershed attribute. As discussed below, the streamflow responses to these variables describe generally expected physical relations with streamflow and provide important insights into the types of watershed properties and processes (Table 2) that influence flow generation and transport over large spatial scales.

The explanatory variables in the final model (Table 2) vary somewhat in their levels of statistical significance (t statistics range from about 2 to 10), suggesting differences in the importance of the explanatory variables on the response of mean annual streamflow to spatial variations in watershed properties. As a measure of their relative importance, we report the coefficient sensitivity (Figure 4), expressed as the percent change in streamflow in response to a 1% unit-change in the explanatory variable coefficient (holding all other coefficients constant at their mean value). The sensitivities are positively correlated with coefficient t -statistics (Table 2) but provide a more informative relative measure of influence based on streamflow response to linear and nonlinear changes in the explanatory variables. Regional variations in the sensitivities are also reported for the HUC-2 regions in the supporting information (see Figures S4–S8).

Two general classes are observed that reflect more than 2 times the difference in median sensitivity values (Figure 4). The upper class includes 10 variables with median sensitivities greater than about 0.04% change in streamflow for a unit 1% change in the explanatory coefficient. Despite differences in median values within this class (ranging from 0.04% to more than 0.1%), considerable overlap is observed in the range of the sensitivities across the calibration sites, based on the standard deviations and 95% intervals. A lower-level class with five variables has median sensitivities of less than about 0.02% change in flow for a unit 1% change; these include the variables sinuosity, urban land, tile drains, reservoirs, and Canadian drainage (which represents a small external CONUS water input).

The highest streamflow sensitivities are observed for the vegetation index and the PET/AET difference variables (Figure 4), which are positively and negatively correlated with streamflow, respectively (Table 2). Consistent with their coefficient signs, the vegetation index serves as an indicator of available water supply in soils and groundwater and is controlled by climatic conditions and land use, whereas the PET/AET difference serves as an indicator of dryness (i.e., runoff loss potential). Thus, the variables indicate that lower quantities of streamflow are associated with more arid watersheds that also have less vegetative cover. Regional patterns in streamflow sensitivity and their magnitudes are similar for the two variables (Figures S4a and S6b, respectively), with the more eastern humid regions (HUC-2 = 1 to 7) displaying ~2 times greater sensitivities than those observed in most of the western arid regions, with notable exceptions for Regions 15 (Lower Colorado) and 18 (California) where widely varying sensitivities are observed. By comparison, regional patterns of flow sensitivities to the seasonal distribution of precipitation (Figure S6a) also reveal important climatic effects on streamflow, as quantified by the model. This variable is positively correlated with streamflow (Table 2), such that larger flows are associated with watersheds where a larger percentage of the annual precipitation occurs in the May–June months. The streamflow response to the higher proportion of precipitation in these months is observed to be greater in arid regions (Figure S6a), where appreciably higher streamflow sensitivities are observed (two to six times greater than in humid regions).

Relatively high streamflow sensitivities are also observed for two hydrological/morphological features of the landscape that are measured by land slope and stream density (Figure 4), which are both positively correlated with streamflow (Table 2). Over large spatial scales, these relations indicate that watersheds with higher stream density and land slope are associated with higher streamflow. The regional patterns in streamflow sensitivity to these variables show some general similarities (Figures S5a and S5b), with drainages in eastern regions displaying the lowest sensitivities; midcontinent and western regions show larger sensitivities for stream density, whereas far western regions have the largest sensitivities for land slope.

Among the soil land-to-water delivery variables, the clay content is positively correlated with streamflow, whereas hydrologic soils group B (silt loam types with moderate infiltration and drainage) and C (sandy clay loam types with low infiltration) were negatively correlated with streamflow (Table 1). Thus, watersheds with larger proportions of type B and C soils tended to have lower runoff potential and streamflow,

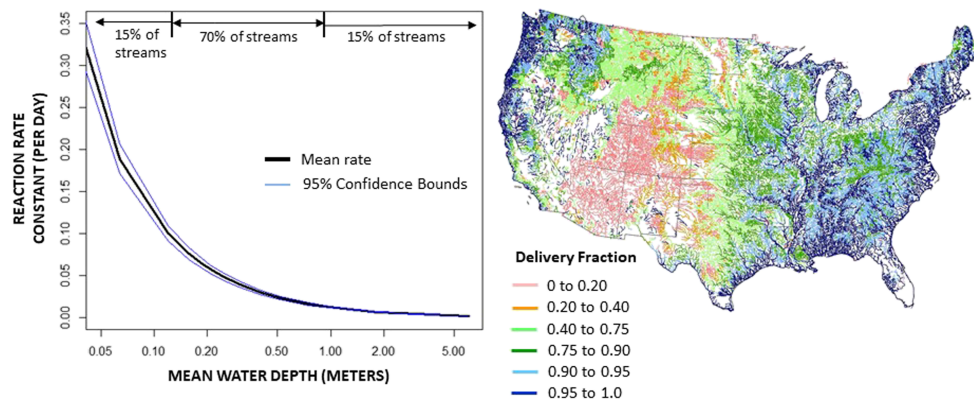


Figure 5. Estimates of net water loss in streams and reservoirs across the conterminous United States from the mean annual SPARROW streamflow model for the 1:500,000 scale: (a) In-stream reaction rate constants for net water loss in relation to mean water depth for streams, based on the mean annual streamflow model. Uncertainties account for model coefficient error only. The net water loss rates in streams and rivers generally decline with increases in streamflow and water depth. (b) Streamflow delivery fraction (dimensionless), the cumulative fraction of the streamflow volume delivered from each reach to the downstream terminus of U.S. river basins (coastal or international boundaries, interior closed basins) as a function of water losses in streams and reservoirs.

whereas watersheds with soils having larger proportions of clay tended to have less permeable conditions with higher runoff potential and streamflow. Each variable contributed statistically unique information (Table 2) about the magnitude of the streamflow response to spatial variations in these classifications of physical and hydraulic properties of soils. Results for the regional streamflow sensitivities to unit changes in these variables (Figures S7a–S7c) showed similarities in the generally higher sensitivities in more humid eastern watersheds but revealed large spatial differences in the magnitude of the streamflow response across regions of the midcontinent and western watersheds.

Measures of land use and farm practices in the model (Table 2), described by forest and urban area and tile drainage, indicated that statistically higher levels of watershed development (e.g., urbanization) were associated with larger streamflow, consistent with studies of regions of the CONUS using different approaches (Diem et al., 2018; Slater & Villarini, 2017). Forest land is potentially indicative of the effects of water losses from evapotranspiration on streamflow (Vadeboncoeur et al., 2018), with flow sensitivities to unit changes in forested lands displaying generally similar variations in magnitudes across CONUS regions (Figure S4b). Urban lands and tile drains displayed among the lowest streamflow sensitivities among the modeled variables (Figure 4), which may be partially explained by the generally patchy nonuniform distribution of these features on the landscape (Figures S4c and S4d) that likely contributed to observed areas of “hot spots” in flow sensitivity; streamflow sensitivities to urban land also showed a slightly elevated levels in western regions as compared to that in eastern regions.

The mean annual SPARROW model (Table 2) provides estimates of net water removal in streams and reservoirs mapped at the 1:500,000 scale across the CONUS (Figure 5). Water loss is quantified as a continuous function of the areal hydraulic load and an estimated mass transfer coefficient (equation (2)). The mean mass transfer coefficients for streams and reservoirs are 0.012 m/day (4.4 m/year) and 0.49 m/year, respectively (Table 2). We converted the mass transfer coefficient for streams to a reaction rate constant, expressed as the fraction removal per unit water travel time, by dividing by mean water depth. The average depth of water is proportional to the water volume that is in contact with the streambed. Mean depth is estimated as a function of mean streamflow from geomorphological relations (Alexander et al., 2000). The reaction rate constants demonstrate that the net mean annual water loss in streams is a water volume-dependent process that declines by 2 orders of magnitude from small to large stream sizes (see Figure 5a)—that is, from 20% to 0.2% of streamflow per day of water travel time for mean water depths from 0.06 to 6 m.

The general decline in the net water loss rates in streams with increases in streamflow and water depth (Figure 5a) contributes to a dendritic pattern of water delivery from individual reaches to the downstream terminus of rivers (Figure 5b). The dendritic pattern is more well defined in humid areas of the CONUS with sharper gradients in transport and loss, whereas more uniform large losses are estimated to occur across the midcontinent and in many western regions. The water losses are largest (some 60% or more) in the Upper

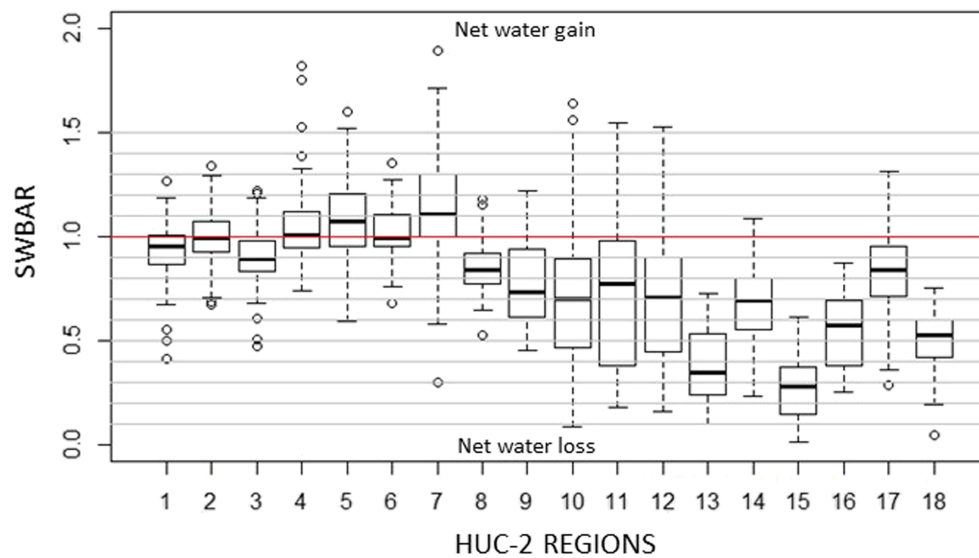


Figure 6. SPARROW Water Balance Accounting Ratios (SWBAR) for mean annual streamflow for the 1,778 calibration sites in HUC-2 regions of the CONUS. *SWBAR* provides a relative measure of the cumulative net change that occurred in natural streamflow (model 2) attributable to the final model 7 (Tables 1 and 2) and its accounting of the effects of large-scale water supply and demand processes on streamflow. Below one *SWBAR* values indicate net water losses, whereas above one values indicate net water gains. Estimates of the natural streamflow in the *SWBAR* calculation were based on small-scale (1-km) unit runoff predictions from model 2 (Table 1). Boxplots display the statistical distributions for the calibration sites in each HUC-2 region.

and Lower Colorado and Rio Grande Basins during transport to basin outlets (Figure 5b). In eastern streams, the largest losses occur over longer distances during transport from headwater streams to coastal areas, whereas relatively minor water losses occur during transport in larger rivers (less than 4%). Similar geographic patterns of water loss are obtained by quantifying the cumulative water losses in river corridors upstream of the modeled reaches in HUC-2 regions (Figure S9). This shows the largest cumulative losses of water in streams and reservoirs in midcontinent and selected western regions, where the interquartile range indicates that from 5% to as much as 40% of the water entering reaches is removed within river corridors; by contrast, less cumulative water loss occurs in eastern regions, where typically less than 5% of the water entering reaches is removed.

Streamflow sensitivities to unit changes in the areal hydraulic load in streams display the largest spatial variability across the calibration sites, compared with all other variables (Figure 4). The highest sensitivities occur in the more arid watersheds in midcontinent and selected western regions (Figure S8a). By comparison, streamflow sensitivities to reservoirs (Figure S8b) are relatively small nationally (median < 0.02%); however, much higher sensitivities are observed in midcontinent and western regions with many localized very high occurrences in these regions.

3.4. Water Balance Accounting Over Large Spatial Scales

Regional distributions of the *SWBAR* values are reported for the calibration stations in Figure 6. *SWBAR* provides a cumulative measure of the large-scale net effect of supply and demand processes on natural streamflow as represented by the previous upscaled predictions of catchment-level unit runoff. The effects of these large-scale processes are accounted for in SPARROW model 7 by (1) including additional explanatory variables that describe terrestrial and aquatic properties of watersheds associated with water cycling processes (Table 2), and (2) applying statistical nonlinear estimation methods over large spatial scales; these methods optimize estimates of the model coefficients such that errors in the predictions of mean annual streamflow in monitored drainages are minimized. The application of these methods facilitates inferences about the water balance effects of natural and cultural processes that operate across large spatial scales in CONUS river networks.

Across most CONUS regions (Figure 6), net water losses in natural runoff are typical within the drainages of most calibration sites as indicated by the predominant portions of the regional distributions (boxplots diagrams) with *SWBAR* values below 1. As expected, net water losses are concentrated in

the more arid midcontinent and western regions. In midcontinent regions (HUC-2 Regions 8–12), the losses vary widely within regions, with interquartile ranges indicating that from 10% to 60% of the natural streamflow is removed over large spatial scales. In western regions (HUC-2 Regions 13–18), the net losses range from 10% to 80% but typically span a narrower range within each region. These losses are potentially related to infiltration and long-term groundwater storage, evapotranspiration on the landscape and in river corridors, river recharge, and cultural activities such as agricultural irrigation that promotes transpiration (Winter, 2001). In selected midwestern regions (i.e., Ohio-5 and Upper Mississippi-7), the net supply of water to natural runoff is somewhat more typical within the drainages of most stations. In these regions, the large-scale additions of water, based on the upper 75th quantiles, range from 10% to 30% of the natural streamflow. These supply processes potentially reflect regional groundwater contributions (Winter, 2001) that are likely excluded from the small-scale runoff predictions but accounted for by SPARROW estimates of the incremental reach unit runoff and water delivery to streams. Land use and cultural activities (e.g., agriculture, urbanization) have also been identified as potential factors that increase water delivery to streams over large spatial scales, including midwestern river basins (Slater & Villarini, 2017) and southeastern river basins (Diem et al., 2018).

4. Discussion

4.1. Modeling Conceptual Approach

The SPARROW streamflow model presented here provides a spatially consistent, large-scale hydrological framework for predicting mean annual streamflow in watersheds and river networks of the CONUS. Our conceptual approach couples the spatially explicit SPARROW river-network modeling structure for the CONUS (Figure 1a; 62,000 reach watersheds) with previous catchment-scale (1-km) water balance predictions of unit area runoff (Wolock & McCabe, 2018) that account for major natural water cycling processes. The approach provides an expanded hydrological structure for the catchment-scale natural water balance predictions by accounting for additional natural and cultural water supply and demand processes that operate over large scales, processes that are critical to explain spatial variability in streamflow observations throughout CONUS watersheds and river basins. The model calibration uses streamflow observations from 1,778 sites (Figure 1a) that represent a diverse collection of CONUS watersheds of widely varying sizes, land cover types, and levels of human development.

The integration of the small 1-km scale unit runoff predictions into the SPARROW modeling framework is facilitated by the spatially distributed, hybrid mechanistic and statistical structure of SPARROW, which conceptually separates watersheds into landscape (water generation, land-to-water delivery factors) and aquatic (streams, reservoirs) process features. The nonlinear, mechanistic structure of the model includes mass balance constraints and nonconservative transport components. These components account for the supply and loss of water routed to streams via surficial and subsurface pathways and the removal of water within streams and reservoirs according to first-order reaction processes. The model optimization using NLLS methods is informed by its application across broad spatial scales, where large variations are observed in streamflow response to watershed properties, which enhances the precision of model coefficients and streamflow predictions (Schwarz et al., 2006; Smith et al., 1997). The data-informed, simple process structure of SPARROW also provides a flexible tool for evaluating hypotheses (Table 2) about the dominant factors that control spatial variability in streamflow. SPARROW methods, including SPARROW water quality models with integrated process-based components (Shih et al., 2010; García et al., 2016; Preston et al., 2011; Schmadel et al., 2018, 2019), have been shown previously to enhance model prediction accuracy and the interpretability of contaminant sources and stream and reservoir loss processes (Smith et al., 1997; Schwarz et al., 2006; Alexander et al., 2000, 2007, 2008).

Our application of SPARROW methods in this study led to the identification of a unique, parsimonious set of watershed properties that influence water supply and demand in CONUS watersheds (Table 2). Evidence of the higher accuracy of the SPARROW streamflow model, compared with that of a simple natural runoff model (Table 3), indicated that the SPARROW model identified unique effects of natural and anthropogenic processes on streamflow that occur across a wide range of geographic scales. The effects (Table 3) were identifiable at locations in small watersheds (less than 1,500 km²) having both undeveloped and developed conditions, but the effects were more prominent in larger watersheds (greater than 1,500 km²), and especially in

more arid regions west of the 100th longitudinal meridian where larger water losses occur during transport in river networks (Figure 6), related to such influences as groundwater recharge and evapotranspiration and human activities that alter streamflow (e.g., Eng et al., 2013). Our findings underscore the value of the SPARROW conceptual framework for streamflow prediction, illustrated by marked improvement in prediction accuracy (Tables 1 and 3) compared to simpler statistical and mechanistic model representations of water generation and stream routing in large, culturally impacted watersheds.

The SPARROW approach represents an initial effort to construct a hydrological modeling framework with spatially and structurally consistent components for estimating model parameters and uncertainties across large spatial scales. The model structure shares hydrological information within and across large river basins in a manner that broadens the spatial extent of the parameter estimation as compared to that of previous CONUS model applications (e.g., Weiskel et al., 2014; Bock et al., 2016; Livneh & Lettenmaier, 2013). The model structure is inclusive of the effects of natural and anthropogenic processes associated with large culturally diverse watersheds, including the cumulative effects of water loss in river networks, properties that have received less consideration in many prior CONUS hydrological applications (e.g., Archfield et al., 2015; Bock et al., 2016). For example, our analysis revealed nonlinear first-order scaling properties of water loss in streams and reservoirs that are emergent across large spatial scales (Table 2; Figure 5) and essential to quantify the cumulative effects of water removal along river corridors in CONUS river networks (Figure S9). These findings contribute new information about the effects of stream and reservoir loss processes on the annual water budget in CONUS rivers that is complementary to growing interests in river corridors as a foundational framework for understanding hydrological and biogeochemical functioning in river networks (Harvey & Gooseff, 2015; Ward & Packman, 2019). Moreover, the approach revealed the appreciable, spatially varying effects of large-scale water supply and demand processes on streamflow in river networks (Table 3; Figure 6), enabled by the spatially distributed, hybrid structure of the SPARROW model. The quantification of these spatial effects has relevance to further development of parameterizations and model structures that can more realistically represent large-scale processes and scaling properties in hydrological models. For example, future applications of SPARROW methods are needed to develop large-scale parameterizations of additional components of catchment-scale water balance models (e.g., PET, soil moisture, recharge) beyond the unit runoff component evaluated in the current study. Additionally, the SPARROW approach can be potentially extended to incorporate temporal variability across seasons, years, and multi-decades, based on recent progress in SPARROW seasonal dynamic and decadal static water-quality modeling (Ator et al., 2019).

4.2. Process-Related Effects on the Streamflow

Results from the sequential evaluation of a series of SPARROW models with increasing hydrological complexity (Table 1) indicate that spatial variations in mean annual streamflow is predominantly explained by natural climatic factors, as described by the major supply (precipitation) and demand (PET, AET) components of the water budget. These components, as accounted for by the previous uncalibrated catchment-scale natural Thornthwaite water balance predictions of unit runoff (Wolock & McCabe, 2018; model 2, Table 1), explain about 72% of the spatial variability in the observed log-transformed water yield at the 1,778 calibration sites, which represent watersheds with a wide range of land cover and levels of human development. Prior CONUS modeling studies of the effects of natural climatic factors on runoff (Vogel et al., 1999; Wolock & McCabe, 1999) have shown that spatial variations in mean annual streamflow in relatively undeveloped watersheds can be primarily explained by precipitation and temperature (e.g., log yield R-squared >85%), with additional climatic measures of water demand and supply and selected geomorphic properties (e.g., slope, elevation, drainage area) providing modest improvements in prediction accuracy.

We found that the inclusion of additional natural and culturally related explanatory variables and the large-scale statistical estimation of the model (Tables 1 and 2) led to major improvements in streamflow prediction accuracy, with the final model (Table 2) explaining nearly 90% of the spatial variability in the logged values of mean annual streamflow yields. This represents a 25% increase in the water yield R-squared and a 60% reduction in the model error (i.e., RMSE) compared with that of the natural unit runoff model (model 2; Table 1). In contrast to the simpler natural runoff model, the final model explicitly captures the large-scale effects of natural and culturally influenced water supply and demand processes in CONUS watersheds (Table 3; Figures 6 and S9).

Our results (Figure 6) revealed the appreciable changes that occur in natural runoff (Model 2), attributable to large-scale water supply and demand processes as identified by the final model (Model 7). The net cumulative effects of these processes are primarily expressed as large regionally variable net water losses at most calibration sites (Figure 6). By contrast, the large-scale net water additions to river basins are observed to be much smaller and are concentrated in a few midwestern regions. The processes that potentially explain spatial variations in these net changes in mean annual streamflow are varied. Water losses may be related to long-term groundwater storage, evapotranspiration on the landscape and in river corridors, river recharge, and cultural activities such as agricultural irrigation that promotes transpiration (Winter, 2001). Water supply processes may include regional groundwater contributions to streams (Winter, 2001) and the increased efficiency of water delivery to streams from modified landscapes, including urbanization and impervious cover (DeWalle et al., 2000; Diem et al., 2018) and agricultural land use that is associated with artificial drainage in many regions. For example, corn and soybean production has been shown to have increased water yields and baseflow, amplifying the effects of precipitation increases that have occurred in past decades in midwestern basins (Slater & Villarini, 2017).

The SPARROW model conceptually identifies three major watershed functions that are responsible for mean streamflow, including water generation, land-to-water delivery to streams, and loss processes in streams and reservoirs during transport in river networks. The water generation and land-water delivery components are closely coupled through multiplicative interactions (equation (1)) that determine the quantity of runoff that is transported to streams from the incremental area of catchments (Figure 1a). The water volume delivered to a stream reach from terrestrial sources in a catchment (small-scale Thornthwaite predictions of natural unit runoff) reflects interactions with the land-to-water delivery variables (and their estimated coefficients) that account for the net effect of watershed processes that can decrease (e.g., evapotranspiration, subsurface storage, consumptive use) or increase (e.g., shallow or deep groundwater, runoff from impervious land areas) water flow. Although the land-to-water delivery component is specified according to 12 statistically unique variables (Table 2), their estimated relations with streamflow do not necessarily reflect a physically based response to these specific variables. Streamflow relations with these variables may be indicative of the flow response to large-scale emergent hydrological processes that have been attributed to the complex interactions of small-scale physical processes (Hrachowitz et al., 2013; McDonnell et al., 2007). Thus, the model coefficients are conceptual or “effective,” rather than physically based, parameters (Hrachowitz et al., 2013) and are likely to describe large-scale dominant or emergent functional properties that are associated with the specific explanatory variables and their related watershed processes (Table 2).

In this context, the land-to-water delivery factors (product of the coefficient and reach-level values of the variable) provide spatially distributed functional properties that describe how spatial variability in mean streamflow is influenced by major types of watershed processes (Table 2) related to climate (i.e., aridity and vegetative cover) and landscape features (i.e., soil structure, channel development, topographic relief, and land use). We observed a general consistency in the streamflow response to the explanatory variables in each process class (Table 2), although each variable contributes uniquely to explaining the reach-level spatial patterns in mean streamflow. Many of these variables and general process types (Table 2) have been previously identified in CONUS models as important predictors of spatial variability in mean streamflow (Vogel et al., 1997, 1999; Weiskel et al., 2014; Bock et al., 2016; Bishop & Church, 1992; Church, 1995). Our findings indicated that spatial variability in mean streamflow is most sensitive to climatic factors that describe the extent of aridity (PET/AET difference) and vegetative cover (and associated soil moisture and groundwater supply) of watersheds, with higher streamflow associated with less arid and more vegetated drainages and the highest flow sensitivity occurring in eastern CONUS regions. The level of stream channel development (stream density) and topographic relief (land slope) in watersheds were also found to significantly influence spatial variability in streamflow (positive relations), with generally higher flow sensitivities observed in western regions. Additionally, the structural characteristics of soils, related to permeability and infiltration (soil clay content, B/C hydrologic soil classifications), were observed to significantly affect streamflow variability, with the clay content and soil classifications displaying unique functional relations with streamflow related to runoff potential in watersheds. We found that streamflow generally displayed higher sensitivities to these soil properties in eastern regions. Moreover, streamflow was found to be influenced by the type of land use and level of cultural development as measured by forest and urban land area and farm practices (tile drainage), with higher levels of urban development and tile drainage, and lower

levels of forested land area, associated with higher streamflow. Streamflow sensitivity to spatial variations in forested area was found to be relatively uniform across the CONUS, whereas “hot spots” of flow sensitivity were observed for urban area and tile drainage, consistent with their more nonuniform distribution in the CONUS that contribute to more localized effects on streamflow. The urban land predictor variable (Table 2) may be inclusive of the effects of municipal and industrial water use and consumption, as suggested by their spatial correlation (Table S1); runoff from impermeable surfaces in urban areas may also have a dominant effect on streamflow (Diem et al., 2018). The inclusion of land cover/use variables in SPARROW shares some functionalities of the Curve Number (CN) method that has been widely used in hydrologic models to account for land-use effects on water runoff (Limbrunner et al., 2005).

We also found that the net water loss in stream channels (Table 2; Figure 5a) can be described as an approximate first-order surface area-dependent process, based on evidence of a decline in the fraction of the streamflow that is removed per day of water travel time (i.e., estimated as a mass transfer coefficient) with increasing stream size. This relation revealed large regional differences in the rates of water loss in streams across the CONUS (Figure 5b) that are caused by spatial variations in both the reaction rate coefficients, related to mean stream channel depth, and the water time of travel (quotient of length and water velocity); both are properties that scale proportionally with flow magnitude (Figure 5a). Accordingly, the mean annual net cumulative losses in stream channels displayed a dendritic pattern (Figure 5b), with the largest losses occurring in small streams and more arid midcontinent and western regions. These patterns (Figure 5) are similar to those observed for nitrogen, phosphorus, and dissolved organic carbon in previous CONUS and regional SPARROW models that are related to the effects of hydrological and biogeochemical processes (Alexander et al., 2000, 2008; Boyer et al., 2006; Preston et al., 2011; Shih et al., 2010). The specific causes of the net water losses are not identified by the streamflow model but may include a combination of processes along river corridors, such as direct evaporation from water surfaces, recharge to the subsurface (especially in western streams), and transpiration from riparian vegetation. In western streams, the losses may also be influenced by water withdrawals and consumptive use by agricultural and urban sectors; evaluations of these effects in future studies could refine the water loss estimates but data on withdrawals are limited.

The functional relations that we observed between the spatial patterns in mean streamflow and those associated with the SPARROW explanatory variables (natural runoff, land-to-water delivery, aquatic water loss) and their connections with watershed processes (Table 2) provide information that can assist efforts to advance the hydrological understanding needed to improve streamflow predictions in ungauged watersheds. For example, the hydrological community (Hrachowitz et al., 2013) has advocated advances in the methods of predicting streamflow response to catchment properties using data-driven approaches, including classification methods and modeling frameworks with catchment-integrated process representations and effective parameters. These methods have been cited as having utility for identifying spatial patterns in the hydrological functioning of catchments (Sivapalan et al., 2003; Hrachowitz et al., 2013) and dominant process effects over large spatial scales that can inform improved process understanding (Sivapalan et al., 2003). The regionalization of the SPARROW model coefficients, as described in our companion study (Alexander et al., 2019), provided spatial refinements to the functional relations that explain mean streamflow in CONUS watersheds.

4.3. Model Prediction Uncertainties

Model performance metrics for the final model predictions of streamflow across the CONUS indicated generally acceptable levels of model precision and bias nationally (Table 2), based on median prediction accuracies that are less than $\pm 10\%$ in most of the 18 hydrological regions (Figure 3a). The regional accuracy of the model was improved by the addition of land-to-water delivery factors and aquatic water loss variables, with the final model (Table 2) showing higher regional accuracy as compared to that of simpler models (Table 1). The improved accuracy of the final model is likely related to the many different combinations of unit runoff and land-to-water delivery factors in CONUS watersheds that provide considerable regional and sub-regional specificity to the model; evidence of sub-regional accuracy included the lack of statistically significant spatial autocorrelation in the model residuals in most CONUS water-resource regions (Table S2).

However, we found evidence of regional variations in the bias and precision of the model predictions (Figure 3a) that highlight the need for improvements in the specification of the SPARROW model. For example, prediction biases were indicated by the general tendency for the underprediction of mean annual

streamflow at more than half of the stations in about two-thirds of the regions, with differences in the magnitude of the biases among the individual regions (Figure 3a). Many of the largest underpredictions occur in drier regions, especially in the Red-Rainy (HUC-2 = 9) and the Lower Colorado (15) basins. In addition, there is evidence of subregional prediction biases, related to spatial clustering of the underprediction and overprediction in five regions (regional spatial autocorrelation tests; Table S2), many of which are concentrated in midcontinent regions along the 100th longitudinal meridian, where abrupt changes occur in precipitation (Seager et al., 2018). These prediction biases suggest that certain unspecified regionally specific hydrological processes need to be additionally accounted for in the SPARROW model, which may include small-scale processes that are potentially deficient in the previous catchment-scale (1-km) water balance estimates of unit runoff (Wolock & McCabe, 2018). The regionalization of the model parameters and uncertainties, as described in our companion study (Alexander et al., 2019), provides an initial step toward addressing the observed prediction biases in the model.

Additionally, we found evidence of two types of nonconstant (heteroscedastic) patterns in the streamflow model residuals. First, much larger model errors were observed in midcontinent and western regions (HUC-2 Regions 9 to 18; Figures 3a and 3b), where prediction accuracies at monitoring sites typically varied from -40% (underprediction) to $+30\%$ (overprediction), as compared to eastern regions where prediction accuracies were generally smaller, ranging from -30% to $+20\%$ (based on the interquartile ranges). Larger departures from these norms were more evident in regions in the western half of the CONUS. The streamflow model errors are assumed to be spatially constant across all watersheds, and thus, the SPARROW model uncertainties are generally overestimated and underestimated in eastern and western regions, respectively. Our observation of major East-West differences in model prediction errors is generally consistent with that from previous statistical models of annual streamflow in HUC-2 hydrological CONUS regions (Vogel et al., 1998, 1999), in which regional model errors were reported to be positively correlated with a dryness/aridity index; similar regional differences in model performance have been previously reported for the CONUS in applications of mechanistic hydrological models (Bock et al., 2016; Newman et al., 2015; Arheimer et al., 2019; Beck et al., 2016). However, it is unknown whether measurement errors in the observations of mean annual streamflow at the monitoring stations may partially explain regional variations in model uncertainties. Although typically small, stream discharge measurement errors have been identified as being important sources of uncertainty in hydrological models (e.g., Kuczera et al., 2010; Westerberg et al., 2011), with recent evidence (J. Kiang, USGS, written comm., April 24, 2015) indicating the somewhat larger measurement errors in field rating-curve flow estimates at certain midcontinent and western U.S. gauges. Second, we observed that the model error variance was negatively correlated with basin size and streamflow magnitude (Figure S1). This is a common pattern observed in many SPARROW water quality models (Preston et al., 2011), which may be partially caused by spatial averaging of model prediction errors over large areas related to errors in the explanatory data. Independent of the causes, it remains to be determined whether regional or scale-dependent heteroscedasticities are statistically identifiable in different areas of the CONUS, a question that is addressed in our companion study (Alexander et al., 2019).

The presence of regional biases and imprecision in the SPARROW streamflow model suggests that further improvements may be possible in the regional specification of the model, either through the inclusion of additional explanatory variables or regionalization of the model parameters. The parameters of the SPARROW streamflow model are assumed to be constant across all watersheds and may be overly constrained spatially. Parameter regionalization is a common practical approach (e.g., Bock et al., 2016; He et al., 2011; Shen, 2018; Archfield et al., 2015; Beck et al., 2016) to improve model accuracy and the transferability of model parameters spatially across watersheds in cases where the underlying causes of model errors are unknown or related to latent variables. In our companion study (Alexander et al., 2019), we found that model accuracy and interpretability were improved by using hierarchical Bayesian methods to regionalize the model coefficients and uncertainties. We also obtained more precise estimates of the model uncertainties by employing hierarchical Bayesian “state-space” methods to separate streamflow measurement errors from process-related model errors. The CONUS streamflow model developed in the present study provides a spatially consistent infrastructure capable of supporting these regional refinements in model specificity to account for spatial variations in latent hydrological processes and model uncertainties.

5. Conclusions

Our integration of predictions of unit runoff from small-scale water balance methods with a spatially explicit large-scale hybrid (statistical-mechanistic) model improved the accuracy of streamflow predictions in rivers and streams within watersheds across the CONUS. The model accounted for statistically unique effects of natural and anthropogenic processes that operate over large spatial scales and influence water availability and transport to streams and downstream water bodies. This study advances understanding of spatial variability in streamflow response to these large-scale processes and can inform efforts to improve the regional and subregional applicability of hydrological models to support the transfer of hydrological information from gauged to ungauged watersheds.

Acknowledgments

The supporting information accompanying this manuscript provides supplementary model results in tabular and graphical form. A digital archive of the models presented in this study is publicly available as a U.S. Geological software release (Alexander & Schwarz, 2019; data sources are referenced and documented in the digital archive). Thanks to Lillian Gorman Sanisaca (USGS Maryland-Delaware-DC Water Science Center) for computer-related support on the R-based software used to estimate the SPARROW models, Dave Wolock of USGS for discussions on water balance methods and for providing access to national water balance results, and to Dave Anning of USGS for assistance with the final selections of the streamflow monitoring stations, based on his extensive knowledge of site locations. Thanks to the journal reviewers for their helpful comments and suggestions. This research was funded by the National Water Quality Assessment Project of the USGS National Water Quality Program; support was also provided by the Integrated Modeling and Prediction Division (IMPD) of the USGS Water Mission Area. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government. We appreciate additional research support at Penn State University from the U.S. Department of Agriculture National Institute of Food and Agriculture.

References

- Addor, N., Newman, A. J., Mizukami, N., & Clark, M. P. (2017). The CAMELS data set: Catchment attributes and meteorology for large-sample studies. *Hydrology and Earth System Sciences*, 21(10), 5293–5313. <https://doi.org/10.5194/hess-21-5293-2017>
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716–723. <https://doi.org/10.1109/TAC.1974.1100705>
- Alexander, R. B., Boyer, E. W., Smith, R. A., Schwarz, G. E., & Moore, R. B. (2007). The role of headwater streams in downstream water quality. *JAWRA Journal of the American Water Resources Association*, 43(1), 41–59. <https://doi.org/10.1111/j.1752-1688.2007.00005.x>
- Alexander, R. B., & Gorman Sanisaca, L. (2019). RSPARROW: An R system for SPARROW modeling (Software Release), U.S. Geological Survey. <https://doi.org/10.5066/P9UAZ6FO>
- Alexander, R. B., & Schwarz, G. E. (2019). A spatially explicit model of long-term mean annual streamflow for the conterminous United States (Software release). U.S. Geological Survey. <https://doi.org/10.5066/P9NT387B>
- Alexander, R. B., Schwarz, G. E., & Boyer, E. W. (2019). Advances in quantifying streamflow variability across continental scales: 2. Improved model regionalization and prediction uncertainties using hierarchical Bayesian methods. *Water Resources Research*. <https://doi.org/10.1029/2019WR025037>
- Alexander, R. B., Smith, R. A., & Schwarz, G. E. (2000). Effect of stream channel size on the delivery of nitrogen to the Gulf of Mexico. *Nature*, 403(6771), 758–761. <https://doi.org/10.1038/35001562>
- Alexander, R. B., Smith, R. A., Schwarz, G. E., Boyer, E. W., Nolan, J. V., & Brakebill, J. W. (2008). Differences in phosphorus and nitrogen delivery to the Gulf of Mexico from the Mississippi River basin. *Environmental Science & Technology*, 42(3), 822–830. <https://doi.org/10.1021/es0716103>
- Anning, D. W., & Flynn, M. E. (2014). *Dissolved-solids sources, loads, yields, and concentrations in streams of the conterminous United States* (U.S. Geological Survey Scientific Investigations Report 2014-5012) (p. 101). Retrieved from <https://doi.org/10.3133/sir20145012>
- Archfield, S. A., Clark, M., Arheimer, B., Hay, L. E., McMillan, H., Kiang, J. E., et al. (2015). Accelerating advances in continental domain hydrologic modeling. *Water Resources Research*, 51, 10,078–10,091. <https://doi.org/10.1002/2015WR017498>
- Archfield, S. A., & Vogel, R. M. (2010). Map correlation method: Selection of a reference streamgage to estimate daily streamflow at ungauged catchments. *Water Resources Research*, 46, 2009WR008481. <https://doi.org/10.1029/2009WR008481>
- Arheimer, B., Pimentel, R., Isberg, K., Crochemore, L., Andersson, J. C. M., Hasan, A., & Pineda, L. (2019). Global catchment modelling using World-Wide HYPER (WWH), open data and stepwise parameter estimation. *Hydrology and Earth System Sciences Discussions*, 1–34. <https://doi.org/10.5194/hess-2019-111>
- Arnold, J. G., Srinivasan, R., Muttiah, R. S., & Allen, P. M. (1999). Continental scale simulation of the hydrologic balance. *Journal of the American Water Resources Association*, 35(5), 1037–1051. <https://doi.org/10.1111/j.1752-1688.1999.tb04192.x>
- Ator, S. W., Brakebill, J. W., & Blomquist, J. D. (2011). *Sources, fate, and transport of nitrogen and phosphorus in the Chesapeake Bay watershed: An empirical model* (U.S. Geological Survey Scientific Investigations Report 2011-5167). Retrieved from <http://pubs.usgs.gov/sir/2011/5167/>
- Ator, S. W., García, A. M., Schwarz, G. E., Blomquist, J. D., & Sekellick, A. J. (2019). Toward explaining nitrogen and phosphorus trends in Chesapeake Bay Tributaries, 1992–2012. *JAWRA Journal of the American Water Resources Association*, 55(5), 1149–1168. <https://doi.org/10.1111/1752-1688.12756>
- Beck, H. E., van Dijk, A. I. J. M., de Roo, A., Miralles, D. G., McVicar, T. R., Schellekens, J., & Bruijnzeel, L. A. (2016). Global-scale regionalization of hydrologic model parameters. *Water Resources Research*, 52, 3599–3622. <https://doi.org/10.1002/2015WR018247>
- Berghuijs, W. R., Sivapalan, M., Woods, R. A., & Savenije, H. H. G. (2014). Patterns of similarity of seasonal water balances: A window into streamflow variability over a range of time scales. *Water Resources Research*, 50, 5638–5661. <https://doi.org/10.1002/2014WR015692>
- Beven, K. (2001). How far can we go in distributed hydrological modelling? *Hydrology and Earth System Sciences*, 5(1), 1–12. <https://doi.org/10.5194/hess-5-1-2001>
- Bishop, G. D., & Church, M. R. (1992). Automated approaches for regional runoff mapping in the northeastern United States. *Journal of Hydrology*, 138(3–4), 361–383. [https://doi.org/10.1016/0022-1694\(92\)90126-G](https://doi.org/10.1016/0022-1694(92)90126-G)
- Bivand, R. (2015). Package “spdep”: Spatial dependence, weighting schemes, statistics and models. Retrieved from <https://cran.r-project.org/web/packages/spdep/spdep.pdf>
- Blöschl, G., Sivapalan, M., Wagener, T., Viglione, A., & Savenije, H. (Eds.) (2013). *Runoff prediction in ungauged basins: Synthesis across processes, places and scales*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139235761>
- Bock, A. R., Hay, L. E., McCabe, G. J., Markstrom, S. L., & Atkinson, R. D. (2016). Parameter regionalization of a monthly water balance model for the conterminous United States. *Hydrology and Earth System Sciences*, 20(7), 2861–2876. <https://doi.org/10.5194/hess-20-2861-2016>
- Boyer, E. W., Alexander, R. B., Parton, W. J., Li, C., Butterbach-Bahl, K., Donner, S. D., et al. (2006). Modeling denitrification in terrestrial and aquatic ecosystems at regional scales. *Ecological Applications*, 16(6), 2123–2142. [https://doi.org/10.1890/1051-0761\(2006\)016\[2123:MDITAA\]2.0.CO;2](https://doi.org/10.1890/1051-0761(2006)016[2123:MDITAA]2.0.CO;2)

- Brakebill, J. W., Ator, S. W., & Schwarz, G. E. (2010). Sources of suspended-sediment flux in streams of the Chesapeake Bay watershed: A regional application of the SPARROW model. *JAWRA Journal of the American Water Resources Association*, 46(4), 757–776. <https://doi.org/10.1111/j.1752-1688.2010.00450.x>
- Carey, S. K., Tetzlaff, D., Seibert, J., Soulsby, C., Buttle, J., Laudon, H., et al. (2010). Inter-comparison of hydro-climatic regimes across northern catchments: Synchronicity, resistance and resilience. *Hydrological Processes*, 24(24), 3591–3602. <https://doi.org/10.1002/hyp.7880>
- Church, M. (1995). Maps of regional evapotranspiration and runoff/precipitation ratios for the northeast United States. *Journal of Hydrology*, 168(1-4), 283–298. [https://doi.org/10.1016/0022-1694\(94\)02640-W](https://doi.org/10.1016/0022-1694(94)02640-W)
- DeWalle, D. R., Swistock, B. R., Johnson, T. E., & McGuire, K. J. (2000). Potential effects of climate change and urbanization on mean annual streamflow in the United States. *Water Resources Research*, 36(9), 2655–2664. <https://doi.org/10.1029/2000WR900134>
- Diem, J. E., Hill, T. C., & Milligan, R. A. (2018). Diverse multi-decadal changes in streamflow within a rapidly urbanizing region. *Journal of Hydrology*, 556, 61–71. <https://doi.org/10.1016/j.jhydrol.2017.10.026>
- Eng, K., Carlisle, D. M., Wolock, D. M., & Falcone, J. A. (2013). Predicting the likelihood of altered streamflows at ungauged rivers across the conterminous United States. *River Research and Applications*, 29(6), 781–791. <https://doi.org/10.1002/rra.2565>
- Eng, K., Milly, P. C. D., & Tasker, G. D. (2007). Flood regionalization: A hybrid geographic and predictor-variable region-of-influence regression method. *Journal of Hydrologic Engineering*, 12(6), 585–591. [https://doi.org/10.1061/\(ASCE\)1084-0699\(2007\)12:6\(585\)](https://doi.org/10.1061/(ASCE)1084-0699(2007)12:6(585))
- García, A. M., Alexander, R. B., Arnold, J. G., Norfleet, L., White, M. J., Robertson, D. M., & Schwarz, G. (2016). Regional effects of agricultural conservation practices on nutrient transport in the upper Mississippi River Basin. *Environmental Science & Technology*, 50(13), 6991–7000. <https://doi.org/10.1021/acs.est.5b03543>
- Gupta, H. V., Perrin, C., Kumar, R., Blöschl, G., Clark, M., Montanari, A., & Andréassian, V. (2013). Large-sample hydrology: a need to balance depth with breadth. *Hydrology and Earth System Sciences Discussions*, 10(7), 9147–9189. <https://doi.org/10.5194/hessd-10-9147-2013>
- Harvey, J., & Gooseff, M. (2015). River corridor science: Hydrologic exchange and ecological consequences from bedforms to basins. *Water Resources Research*, 51(9), 6893–6922. <https://doi.org/10.1002/2015WR017617>
- He, Y., Bárdossy, A., & Zehe, E. (2011). A review of regionalisation for continuous streamflow simulation. *Hydrology and Earth System Sciences*, 15, 3539–3553. <https://doi.org/10.5194/hess-15-3539-2011>
- Hrachowitz, M., Savenije, H. H. G., Blöschl, G., McDonnell, J. J., Sivapalan, M., Pomeroy, J. W., et al. (2013). A decade of Predictions in Ungauged Basins (PUB)—A review. *Hydrological Sciences Journal*, 58(6), 1198–1255. <https://doi.org/10.1080/02626667.2013.803183>
- Hundecha, Y., Arheimer, B., Donnelly, C., & Pechlivanidis, I. (2016). A regional parameter estimation scheme for a pan-European multi-basin model. *Journal of Hydrology: Regional Studies*, 6, 90–111. <https://doi.org/10.1016/j.ejrh.2016.04.002>
- Jakeman, A. J., & Hornberger, G. M. (1993). How much complexity is warranted in a rainfall-runoff model? *Water Resources Research*, 29(8), 2637–2649. <https://doi.org/10.1029/93WR00877>
- Jehn, F. U., Bestian, K., Breuer, L., Kraft, P., & Houska, T. (2019). Clustering CAMELS using hydrological signatures with high spatial predictability. *Hydrology and Earth System Sciences Discussions*, 1–21. <https://doi.org/10.5194/hess-2019-129>
- Kuczera, G., Kavetski, D., Renard, B., & Thyer, M. (2010). A limited-memory acceleration strategy for MCMC sampling in hierarchical Bayesian calibration of hydrological models. *Water Resources Research*, 46. <https://doi.org/10.1029/2009WR008985>
- Li, H., Calder, C. A., & Cressie, N. (2007). Beyond Moran's I: Testing for spatial dependence based on the spatial autoregressive model. *Geographical Analysis*, 39(4), 357–375. <https://doi.org/10.1111/j.1538-4632.2007.00708.x>
- Limbrunner, J. F., Vogel, R. M., & Chapra, S. C. (2005). A parsimonious watershed model. In *Computer Models of Watershed Hydrology* (2nd ed.). CRC Press.
- Lins, H. F. (2012). USGS Hydro-climatic data network 2009 (HCDN-2009). U.S. Geological Survey Fact Sheet 2012-3047. Retrieved from <https://pubs.usgs.gov/fs/2012/3047/pdf/fs2012-3047.pdf>; https://water.usgs.gov/osw/hcdn-2009/HCDN-2009_Station_Info.xlsx
- Liu, Y., & Gupta, H. V. (2007). Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework. *Water Resources Research*, 43. <https://doi.org/10.1029/2006WR005756>
- Livneh, B., & Lettenmaier, D. P. (2013). Regional parameter estimation for the unified land model. *Water Resources Research*, 49, 100–114. <https://doi.org/10.1029/2012WR012220>
- McDonnell, J. J., Sivapalan, M., Vaché, K., Dunn, S., Grant, G., Haggerty, R., et al. (2007). Moving beyond heterogeneity and process complexity: A new vision for watershed hydrology. *Water Resources Research*, 43. <https://doi.org/10.1029/2006WR005467>
- Miller, M. P., Buto, S. G., Lambert, P. M., & Rumsey, C. A. (2017). *Enhanced and updated spatially referenced statistical assessment of dissolved-solids load sources and transport in streams of the Upper Colorado River Basin* (U.S. Geological Survey Scientific Investigations Report 2017-5009) (p. 23).
- Miller, M. P., Buto, S. G., Susong, D. D., & Rumsey, C. A. (2016). The importance of base flow in sustaining surface water flow in the Upper Colorado River Basin. *Water Resources Research*, 52(5), 3547–3562. <https://doi.org/10.1002/2015WR017963>
- Milly, P. C. D. (1994). Climate, soil water storage, and the average annual water balance. *Water Resources Research*, 30, 2143–2156. <https://doi.org/10.1029/94WR00586>
- Montanari, A., Young, G., Savenije, H. H. G., Hughes, D., Wagener, T., Ren, L. L., et al. (2013). “Panta Rhei—Everything flows”: Change in hydrology and society—The IAHS scientific decade 2013–2022. *Hydrological Sciences Journal*, 58(6), 1256–1275. <https://doi.org/10.1080/02626667.2013.809088>
- Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., & Veith, T. L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE*, 50(3), 885–900. <https://doi.org/10.13031/2013.23153>
- Nash, J. C. (2015, February 20). Package “nlmrt”: Functions for nonlinear least squares solutions. Retrieved from <https://cran.r-project.org/web/packages/nlmrt/nlmrt.pdf>
- Newman, A. J., Clark, M. P., Sampson, K., Wood, A., Hay, L. E., Bock, A., et al. (2015). Development of a large-sample watershed-scale hydrometeorological data set for the contiguous USA: Data set characteristics and assessment of regional variability in hydrologic model performance. *Hydrology and Earth System Sciences*, 19(1), 209–223. <https://doi.org/10.5194/hess-19-209-2015>
- Nolan, J. V., Brakebill, J. W., Alexander, R. B., & Schwarz, G. E. (2002). *Enhanced river reach file 2 (U.S. Geological Survey Open-File Report 02-40)*. Reston, VA: U.S. Geological Survey.
- Preston, S. D., Alexander, R. B., & Wolock, D. M. (2011). Sparrow modeling to understand water-quality conditions in major regions of the United States: A featured collection introduction. *JAWRA Journal of the American Water Resources Association*, 47(5), 887–890. <https://doi.org/10.1111/j.1752-1688.2011.00585.x>

- Rakovec, O., Kumar, R., Mai, J., Cuntz, M., Thober, S., Zink, M., et al. (2016). Multiscale and multivariate evaluation of water fluxes and states over European River Basins. *Journal of Hydrometeorology*, 17(1), 287–307. <https://doi.org/10.1175/JHM-D-15-0054.1>
- Rode, M., Arhonditsis, G., Balin, D., Kebede, T., Krysanova, V., van Griensven, A., & van der Zee, S. E. A. T. M. (2010). New challenges in integrated water quality modelling. *Hydrological Processes*, 24(24), 3447–3461. <https://doi.org/10.1002/hyp.7766>
- Samaniego, L., Kumar, R., & Attinger, S. (2010). Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale. *Water Resources Research*, 46. <https://doi.org/10.1029/2008WR007327>
- Schmadel, N. M., Harvey, J. W., Alexander, R. B., Schwarz, G. E., Moore, R. B., Eng, K., et al. (2018). Thresholds of lake and reservoir connectivity in river networks control nitrogen removal. *Nature Communications*, 9(1), 2779. <https://doi.org/10.1038/s41467-018-05156-x>
- Schmadel, N. M., Harvey, J. W., Schwarz, G. E., Alexander, R. B., Gomez-Velez, J. D., Scott, D., & Ator, S. W. (2019). Small ponds in headwater catchments are a dominant influence on regional nutrient and sediment budgets. *Geophysical Research Letters*, 46(16), 9669–9677. <https://doi.org/10.1029/2019GL083937>
- Schwarz, G. E., Alexander, R. B., Smith, R. A., & Preston, S. D. (2011). The regionalization of national-scale SPARROW models for stream nutrients I. *JAWRA Journal of the American Water Resources Association*, 47(5), 1151–1172. <https://doi.org/10.1111/j.1752-1688.2011.00581.x>
- Schwarz, G. E., Hoos, A. B., Alexander, R. B., & Smith, R. A. (2006). *The SPARROW surface water-quality model: Theory, application, and user documentation (U.S. Geological Survey Techniques and Methods Report, Book 6, Chapter B3.)*. Reston, VA: U.S. Geological Survey. <https://pubs.er.usgs.gov/publication/tm6B3>
- Seaber, P. R., Kapinos, F. P., & Knapp, G. L. (1987). *Hydrologic unit maps (U.S. Geological Survey Water-Supply Paper 2294)* (p. 20). Reston, VA: U.S. Geological Survey.
- Seager, R., Lis, N., Feldman, J., Ting, M., Williams, A. P., Nakamura, J., et al. (2018). Whither the 100th meridian? The once and future physical and human geography of America's arid–humid divide. Part I: The story so far. *Earth Interactions*, 22(5), 1–22. <https://doi.org/10.1175/EI-D-17-0011.1>
- Shen, C. (2018). A transdisciplinary review of deep learning research and its relevance for water resources scientists. *Water Resources Research*, 54(11), 8558–8593. <https://doi.org/10.1029/2018WR022643>
- Shih, J.-S., Alexander, R. B., Smith, R. A., Boyer, E. W., Schwarz, G. E., & Chung, S. (2010). *An initial SPARROW model of land use and in-stream controls on total organic carbon in streams of the conterminous United States (U.S. Geological Survey Open-File Report 2010-1276)* (p. 22). Reston, VA: U.S. Geological Survey. Retrieved from <http://pubs.usgs.gov/of/2010/1276>
- Sivapalan, M., Takeuchi, K., Franks, S. W., Gupta, V. K., Karambiri, H., Lakshmi, V., et al. (2003). IAHS decade on predictions in ungauged basins (PUB), 2003–2012: Shaping an exciting future for the hydrological sciences. *Hydrological Sciences Journal*, 48(6), 857–880. <https://doi.org/10.1623/hysj.48.6.857.51421>
- Slater, L., & Villarini, G. (2017). Evaluating the drivers of seasonal streamflow in the U.S. midwest. *Water*, 9(9), 695. <https://doi.org/10.3390/w9090695>
- Smith, R. A., Schwarz, G. E., & Alexander, R. B. (1997). Regional interpretation of water-quality monitoring data. *Water Resources Research*, 33(12), 2781–2798. <https://doi.org/10.1029/97WR02171>
- Stuckey, M., Koerkle, E., & Ulrich, J. (2014). Estimation of baseline daily mean streamflows for ungaged locations on Pennsylvania streams, water years 1960–2008 (U.S. Geological Survey Scientific Investigations Report 2012-5142) (p. 61).
- Thorntwaite, C. W., & Mather, J. R. (1955). *The water balance*. Publications in Climatology, 8(1), Drexel Institute of Technology, Laboratory of Climatology, Centerton, NJ.
- Geological Survey (1994). *USGS water data for the Nation*. U.S. Geological Survey National Water Information System database. Reston, VA: U.S. Department of Interior, U.S. Geological Survey. <https://doi.org/10.5066/F7P55KJN>
- Vadeboncoeur, M. A., Green, M. B., Asbjornsen, H., Campbell, J. L., Adams, M. B., Boyer, E. W., et al. (2018). Systematic variation in evapotranspiration trends and drivers across the Northeastern United States. *Hydrological Processes*, 32(23), 3547–3560. <https://doi.org/10.1002/hyp.13278>
- Vogel, R. M., Bell, C. J., & Fennessey, N. M. (1997). Climate, streamflow and water supply in the northeastern United States. *Journal of Hydrology*, 198(1–4), 42–68. [https://doi.org/10.1016/S0022-1694\(96\)03327-6](https://doi.org/10.1016/S0022-1694(96)03327-6)
- Vogel, R. M., & Sankarasubramanian, A. (2000). Spatial scaling properties of annual streamflow in the United States. *Hydrological Sciences Journal*, 45(3), 465–476. <https://doi.org/10.1080/02626660009492342>
- Vogel, R. M., Tsai, Y., & Limbrunner, J. F. (1998). The regional persistence and variability of annual streamflow in the United States. *Water Resources Research*, 34(12), 3445–3459. <https://doi.org/10.1029/98WR02523>
- Vogel, R. M., Wilson, I., & Daly, C. (1999). Regional regression models of annual streamflow for the United States. *Journal of Irrigation and Drainage Engineering*, 125(3), 148–157. [https://doi.org/10.1061/\(ASCE\)0733-9437\(1999\)125:3\(148\)](https://doi.org/10.1061/(ASCE)0733-9437(1999)125:3(148))
- Ward, A. S., & Packman, A. I. (2019). Advancing our predictive understanding of river corridor exchange. *Wiley Interdisciplinary Reviews: Water*, 6(1), e1327. <https://doi.org/10.1002/wat2.1327>
- Weiskel, P. K., Wolock, D. M., Zarriello, P. J., Vogel, R. M., Levin, S. B., & Lent, R. M. (2014). Hydroclimatic regimes: a distributed water-balance framework for hydrologic assessment, classification, and management. *Hydrology and Earth System Sciences*, 18(10), 3855–3872. <https://doi.org/10.5194/hess-18-3855-2014>
- Westerberg, I. K., Guerrero, J.-L., Younger, P. M., Beven, K. J., Seibert, J., Halldin, S., et al. (2011). Calibration of hydrological models using flow-duration curves. *Hydrology and Earth System Sciences*, 15(7), 2205–2227. <https://doi.org/10.5194/hess-15-2205-2011>
- Winter, T. C. (2001). The concept of hydrologic landscapes. *Journal of the American Water Resources Association*, 37(2), 335–349. <https://doi.org/10.1111/j.1752-1688.2001.tb00973.x>
- Wolock, D. M., & McCabe, G. J. (1999). Explaining spatial variability in mean annual runoff in the conterminous United States. *Climate Research*, 11(2), 149–159. <https://doi.org/10.3354/cr011149>
- Wolock, D. M., & McCabe, G. J. (2018). Water balance model inputs and outputs for the conterminous United States, 1900–2015 (Data release). U.S. Geological Survey. <https://doi.org/10.5066/F71V5CWN>
- Xia, Y., Mitchell, K., Ek, M., Cosgrove, B., Sheffield, J., Luo, L., et al. (2012). Continental-scale water and energy flux analysis and validation for North American Land Data Assimilation System project phase 2 (NLDAS-2): 2. Validation of model-simulated streamflow. *Journal of Geophysical Research*, 117(D3). <https://doi.org/10.1029/2011JD016051>