USGS
science for a changing world

Prepared in cooperation with U.S. Army Corps of Engineers, Chicago District

Effect of Uncertainty of Discharge Data on Uncertainty of Discharge Simulation for the Lake Michigan Diversion, Northeastern Illinois and Northwestern Indiana

Scientific Investigations Report 2022–5102

U.S. Department of the Interior
U.S. Geological Survey

Cover: Photograph showing downstream view at the low-water control and overbank at the Long Run near Lemont, Illinois, streamgage (U.S. Geological Survey [USGS] station 05537500) on April 22, 2008. Gage height was 1.07 feet, and discharge was 12.0 cubic feet per second. Photograph was taken by David Schrader, USGS.

Effect of Uncertainty of Discharge Data on Uncertainty of Discharge Simulation for the Lake Michigan Diversion, Northeastern Illinois and Northwestern Indiana

By David T. Soong and Thomas M. Over

Prepared in cooperation with U.S. Army Corps of Engineers, Chicago District

Scientific Investigations Report 2022–5102

U.S. Department of the Interior
U.S. Geological Survey

U.S. Geological Survey, Reston, Virginia: 2022

For more information on the USGS—the Federal source for science about the Earth, its natural and living resources, natural hazards, and the environment—visit https://www.usgs.gov or call 1–888–ASK–USGS.

For an overview of USGS information products, including maps, imagery, and publications, visit https://store.usgs.gov/.

Suggested citation:
Soong, D.T., and Over, T.M., 2022, Effect of uncertainty of discharge data on uncertainty of discharge simulation for the Lake Michigan Diversion, northeastern Illinois and northwestern Indiana: U.S. Geological Survey Scientific Investigations Report 2022–5102, 54 p., https://doi.org/10.3133/sir20225102.

Associated data:
Over, T.M., Soong, D.T., and Sortor, R.N., 2022, Models, inputs, and outputs for estimating the uncertainty of discharge simulations for the Lake Michigan Diversion using the Hydrological Simulation Program – FORTRAN model: U.S. Geological Survey data release, https://doi.org/10.5066/P9UC21B0.

U.S. Geological Survey, 2014, National Land Cover Database (NLCD) 2011 Land Cover Conterminous United States: U.S. Geological Survey data release, https://doi.org/10.5066/P97S2IID.

U.S. Geological Survey, 2020, USGS water data for the Nation: U.S. Geological Survey National Water Information System database, https://doi.org/10.5066/F7P55KJN.

ISSN 2328-0328 (online)

Acknowledgments

John Doherty, developer of the Model Independent Parameter Estimation (PEST) software package, graciously responded to many questions regarding the design of the methodologies used in the study. Ben Renard and Jérôme Le Coz (Institut national de recherche pour l'agriculture, l'alimentation et l'environnement, France) provided the Bayesian rating curve estimation (BaRatin) software and assisted its application. The Lake Michigan Diversion Accounting project managers Dr. Tzuoh-Ying Su (retired) and Mr. Jeff Fuller of the U.S. Army Corps of Engineers-Chicago District provided operational project support. Mr. Jeff Fuller also provided hourly meteorological and precipitation data used in this study.

Marvin Harris, Crystal Prater, and Ryan Beaulin of the U.S. Geological Survey (USGS) provided discharge measurement information and assisted in its interpretation. Kevin Kho and Peter Regan provided programming assistance. Amy Russell provided guidance on and facilitated workflows and participated in report organization and review.

The help of all these individuals is sincerely appreciated.

Contents

Figures

Tables

Conversion Factors

U.S. customary units to International System of Units

Multiply	By	To obtain
	Length	
inch (in.)	2.54	centimeter (cm)
inch (in.)	25.4	millimeter (mm)
foot (ft)	0.3048	meter (m)
mile (mi)	1.609	kilometer (km)
	Area	
square mile (mi^2)	259.0	hectare (ha)
square mile (mi^2)	2.590	square kilometer (km^2)
	Volume	
cubic foot (ft^3)	0.02832	cubic meter (m^3)
	Flow rate	
cubic foot per second (ft^3/s)	0.02832	cubic meter per second (m^3/s)
inch per hour (in/h)	0.0254	meter per hour (m/h)

Temperature in degrees Fahrenheit (°F) may be converted to degrees Celsius (°C) as follows:
°C = (°F – 32) / 1.8.

Datum

Horizontal coordinate information is referenced to the North American Datum of 1983 (NAD 83).

Supplemental Information

A water year is the 12-month period from October 1 through September 30 and designated by the calendar year in which it ends (for example, water year 2019 is the period beginning October 1, 2018, and ending September 30, 2019).

Abbreviations

BaRatin	Bayesian rating curve estimation
CCPN	Cook County Precipitation Network
CI	confidence interval
CV	coefficient of variation
ET	evapotranspiration
FDC	flow-duration curve
HSPF	Hydrological Simulation Program–FORTRAN
LMDA	Lake Michigan Diversion Accounting
NAD 83	North American Datum of 1983
NLCD	National Land Cover Database
NWS–COOP	National Weather Service Cooperative Observer Program
PEST	Model-Independent Parameter Estimation and Uncertainty Analysis package
POR	period of record
POS	period of study
SPD	stage-period-discharge
TRC	Technical Review Committee
TSPROC	Time Series Processor
USACE	U.S. Army Corps of Engineers
USGS	U.S. Geological Survey
V_N	normalized variability index
WCF	weir-channel-flood plain
WY	water year

Effect of Uncertainty of Discharge Data on Uncertainty of Discharge Simulation for the Lake Michigan Diversion, Northeastern Illinois and Northwestern Indiana

By David T. Soong and Thomas M. Over

Abstract

Simulation models of watershed hydrology (also referred to as "rainfall-runoff models") are calibrated to the best available streamflow data, which are typically published discharge time series at the outlet of the watershed. Even after calibration, the model generally cannot replicate the published discharges because of simplifications of the physical system embedded in the model structure and uncertainties of the input data and of the estimated model parameters, which, although optimized for the given calibration data, remain uncertain. The input data errors are caused by uncertainties in the forcing data, such as precipitation and other climatological data, and in the published discharges used for calibration. In the numerical algorithms used for calibration, the published discharges are often assumed to be without error, but they are themselves uncertain, typically having been computed using ratings, which are models fitted to uncertain discharge measurements.

In this study, uncertainty of published daily discharge data and how the discharge uncertainty is transmitted to the parameter values of the Hydrological Simulation Program–FORTRAN (HSPF) rainfall-runoff model and to the simulated discharge at both calibration and prediction locations were investigated for the Lake Michigan diversion in northeastern Illinois and northwestern Indiana. The HSPF model used in this study is used by the U.S. Army Corps of Engineers as part of quantifying the diversion of water from Lake Michigan by the State of Illinois. In this study, the model is calibrated jointly at two watersheds in the study area; the resulting model is considered the base model in this study. Seven other gaged watersheds in the study area are used for testing predictive simulations. A Bayesian rating curve estimation (BaRatin) approach, the BaRatin stage-period-discharge (SPD) method, was used to estimate the uncertainty of the published discharge from the calibration watersheds. To characterize the effect of the discharge uncertainty on parameter values, the HSPF model parameters were recalibrated to 17 nonrandomly selected pairs of discharge series from the BaRatin SPD analysis. To provide an indicator of the effect of parameter uncertainty to compare to the effect of discharge uncertainty, 1,000 parameter sets also were randomly generated from the estimated parameter covariance matrix of the base model. The recalibrated and random parameter sets were then used in HSPF simulations of discharge at the two calibration watersheds and at the seven prediction watersheds. Selected discharge summary statistics—the period-of-study (POS, water years 1997 to 2015) mean discharge, selected flow-duration curve (FDC) quantiles, and water year mean discharges—are used to characterize the variability between simulated and published discharge.

A normalized variability index (V_N) is used as a measure of the uncertainty of flow statistics arising from the uncertainty of the sources considered in this study. When this index is at least 1, the variability of the simulations is large enough to explain the median error between simulated and published values, although offsetting errors from other sources are also likely. When the index is appreciably less than 1, the variability of the simulations is clearly insufficient to explain the median error between simulated and published values. At the two calibration watersheds and for results of the two simulation sets considered together, the V_N values ranged from 0.2 to 0.8 for POS mean discharge, from 0.3 to 0.6 in the median for a set of FDC quantiles, and from 0.1 to 0.2 in the median for water year mean discharges. These values indicate that substantial uncertainty remains unexplained. Even though two watersheds were used in calibration, that calibration was highly constrained because it was applied to the watersheds simultaneously and was subject to parameter regularization that constrained the adjustment of the parameters from their initial values. These constraints were applied to avoid overfitting to the calibration watersheds and thus to increase the likelihood that the resulting parameters would give accurate results at watersheds not used in the calibration, but they created a parameter transfer error in the calibration watershed results shown by the balancing of errors between the two watersheds. Additional remaining error sources include model structural error and meteorological forcing error to the degree that the calibration was unable to adjust the parameters to account for these errors. At the prediction watersheds, the corresponding V_N values were almost always substantially lower

than those values at the calibration watersheds. This result is expected because the prediction watersheds have additional uncertainty, including parameter transfer error.

The work described in this report provides preliminary estimates of a limited range of sources of error in predicted discharge uncertainty. Future work would be beneficial to obtain a better statistical characterization of the effect of the uncertainty of calibration discharge series and to address additional sources of uncertainty, such as from precipitation input data used in calibration and prediction and from structural (model) errors.

Introduction

With increasing needs for streamflow information but scarcity of available observations, statistical and physical models are used for predicting discharge in ungaged or sparsely gaged basins. Uncertainty in discharge predictions can be appreciable, and the International Association of Hydrological Sciences conducted a decade-long (2003–12) initiative aiming to reduce uncertainties from the hydrological data and hydrological modelling on prediction in ungaged basins (Sivapalan, 2003; Sivapalan and others, 2003; Bloschl and others, 2013; Hrachowitz and others, 2013). Uncertainty quantification is an analysis to estimate the likelihood of certain outcomes if some aspects of the system are not exactly known (National Research Council, 2012). In the rainfall-runoff modeling context, using assumptions and simplifications to describe hydrologic processes that are poorly known or are unknown is necessary. These simplifications have led to a diversity of rainfall-runoff models, and as a result such models are described according to the governing equations used (as empirical, conceptual, and physical models) or according to the representation of spatial processes (as lumped, semi-distributed, and distributed models), making users aware of model strengths and weaknesses (Sitterson and others, 2017). Once a rainfall-runoff model is selected to simulate flows at the selected watershed, estimating parameter values so the output matches the selected discharge quantities, that is, model calibration, constitutes a major effort in the success of model applications. Because a rainfall-runoff model cannot perfectly simulate all the environmental processes (Doherty, 2010; Sitterson and others, 2017) and the data that drive the model and are used for calibrating parameters or evaluating results contain noise or errors (Hrachowitz and others, 2013), a calibrated parameter set inevitably contains uncertainty. Model calibration generally reduces but cannot eliminate misfits between simulated and observed discharges. Misfits for a calibrated model generally increase when the characteristics of climatic inputs or targeted discharge have appreciable changes. Slight changes in parameter values often result in large changes in simulated discharges; thus, when it is used in the typical way, with a single parameter set and single forcing dataset, a rainfall-runoff model produces a deterministic result, but this result represents a single realization of many possible outcomes of model simulations when uncertainty of the modeling system is considered (Farmer and Vogel, 2016). As a result, modeling results are best treated probabilistically. Estimation of predictive uncertainty is considered good practice in any environmental modeling activity (Refsgaard and others, 2007), especially when the model simulation serves as a tool in providing information to support decision making in watershed management.

Background

A canal construction project in northeastern Illinois from 1892 to 1922 (Hill, 2007) diverted the Chicago and Calumet Rivers from flowing into Lake Michigan to the Illinois River by way of the Chicago Sanitary and Ship Canal, the Calumet-Sag Channel, and the Des Plaines River (fig. 1). Because the Illinois River discharges to the Mississippi River rather than the Great Lakes – St. Lawrence Seaway (not shown on figure), flows in this canal constitute an interbasin water transfer or diversion with effects on the water resources of multiple States and Canada. Following multiple U.S. Supreme Court decrees related to the diversion, most recently modified in 1980 (Wisconsin v. Illinois, 449 U.S. 48, 1980), the Lake Michigan Diversion Accounting (LMDA) system was developed by the U.S. Army Corps of Engineers (USACE) Chicago District for estimating the diversion of water from Lake Michigan by the State of Illinois. The 1986 Water Resources Development Act gave the USACE the responsibility to monitor and compute the Illinois diversion of water from Lake Michigan (42 U.S.C. 1962d-20; U.S. Army Corps of Engineers, Chicago District, undated a, b).

The diversion has various components; for example, the wastewater treatment plant discharges to the Chicago Sanitary and Ship Canal, originating from water supplied by pumpage for water supply from Lake Michigan (Espey and others, 2019). Surface-water runoff from the Lake Michigan watershed, which prior to the canal construction (approximately 673 square miles [mi^2]) flowed to the Chicago and Calumet Rivers and then to Lake Michigan, is also considered a diversion. Catchments in the diverted watershed are mostly ungaged. To estimate discharges from the ungaged catchments, the USACE applies the Hydrological Simulation Program–FORTRAN (HSPF; Bicknell and others, 2005). The HSPF model is a process-based, continuous, spatially lumped-parameter model developed for simulating the movement of water on pervious and impervious surfaces, in soil storages, and within streams and well-mixed reservoirs (Bicknell and others, 2005). The USACE established a regional approach for determining the parameters of the HSPF model (that is, calibrating the model) used for the LMDA project (for example, Espey and others, 1993). A regional approach for parameter calibration means that a common set of parameter values is used throughout a region and is developed by evaluation of simulations on multiple basins, which facilitates

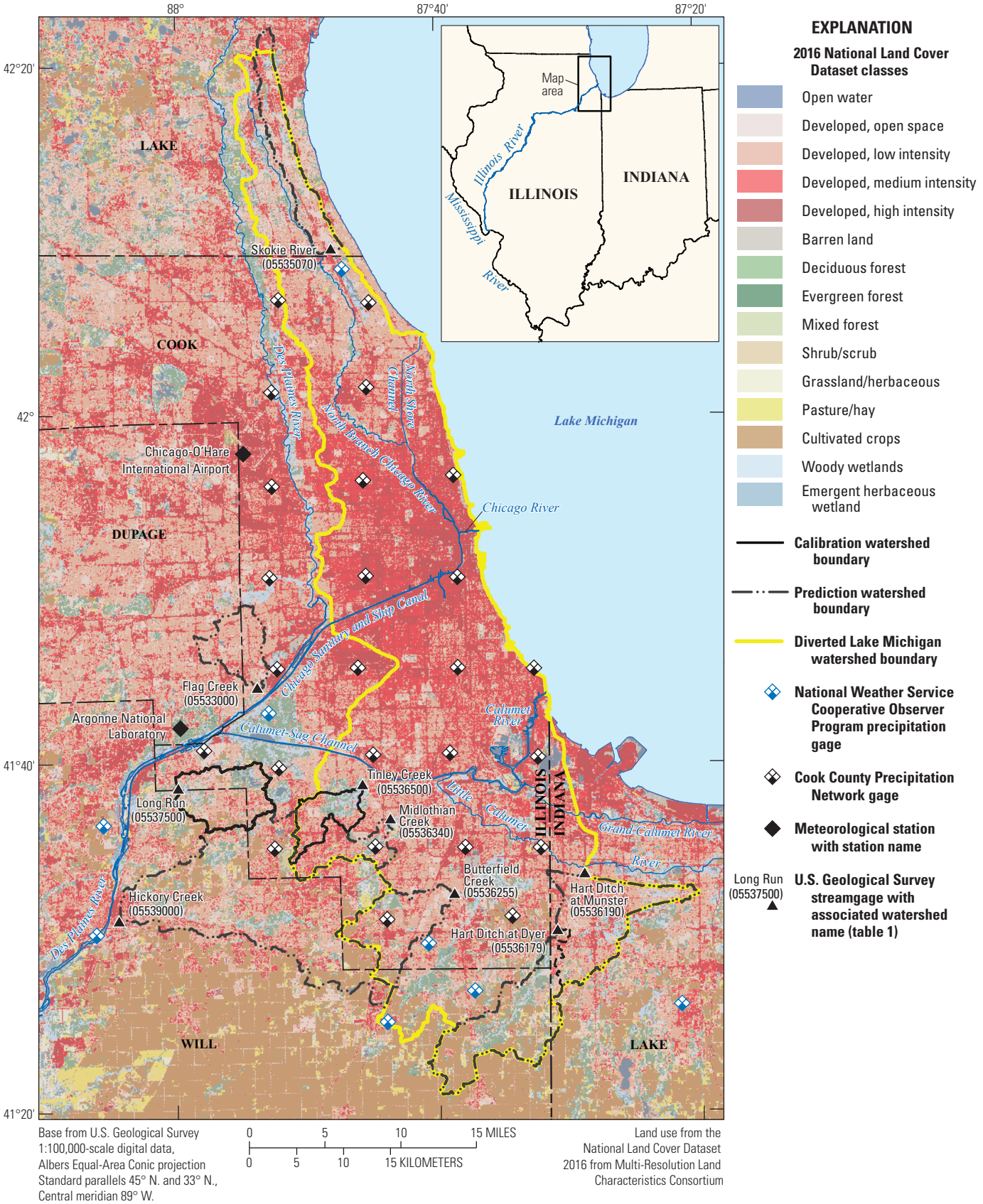

Figure 1. Location of diverted Lake Michigan watershed and the Cook County Precipitation Network gages, U.S. Geological Survey streamgages and associated watersheds, and National Weather Service Cooperative Observer Program precipitation gages used in this study.

the use of the model in ungaged basins in the region. In the construction of the LMDA HSPF models, the catchment land uses are aggregated into three type of land covers: grassland, forest, and directly connected impervious land. Each land-cover type has a separate set of parameters, and the same parameter values are applied to that land-cover type for all HSPF models in the region. Parameter values of each land-cover type do not change spatially (because HSPF is lumped) nor seasonally (that is, HPSF parameters that allow monthly variations are not varied seasonally), and land-use fractions in the model do not vary in time. Snow accumulation and snowmelt processes on all three land-cover types are simulated using HSPF's energy balance approach. The regional HSPF parameter approach and the original calibration of HSPF met the Supreme Court requirement of using the "best current engineering practice and scientific knowledge" according to the Fifth Technical Review Committee (TRC; Espey and others, 2004, p. ii). However, because of a lack of documentation of the regional parameters and their transfer (that is, how the regional parameter set was determined from multiple calibration watersheds and the accuracy of their transfer), the Fifth TRC recommended additional checks of simulated discharges to confirm the accuracy of the HSPF model (Espey and others, 2004). Parameters governing runoff generation from pervious land segments (that is, those in the PWATER section of PERLND module modeling grassland and forest land) were targeted for revisions (Soong and Over, 2015, p. 7–8). Revisions of the parameters pertinent to runoff generation from impervious lands and routing were not considered. In LMDA operations, the flow routing is handled by a separate hydraulic model. Revisions to the snow process parameters were considered (Over and others, 2010), but the Sixth TRC recommended that changes to snow parameters should be made based on a large system-wide study (Espey and others, 2009). The impervious land parameters were most recently revised in the late 1990s (Soong and Over, 2015). At present (2020), nine regional HSPF parameter sets have been investigated at different times in the LMDA system (Soong and Over, 2015; Espey and others, 2019).

Among the nine HSPF parameter sets, the parameter sets from the original calibration were determined from calibration of published daily discharges from four gaged watersheds adjacent to the diverted watershed in the 1960s and 1970s (for example, Espey and others, 2004, p. 53). Subsequent parameter sets were developed using different types of data or watersheds in the calibration (U.S. Army Corps of Engineers, Chicago District, 2009; Soong and Over, 2015). The predictive accuracy of six of the nine parameter sets was evaluated by Soong and Over (2015) using daily discharge records between water years (WYs) 1995 and 2011 available at nine smaller watersheds in or near the diverted watersheds. In addition to their evaluations of historical parameter sets, Soong and Over (2015) developed a new parameter set by simultaneously calibrating two watersheds using the Model-Independent Parameter Estimation and Uncertainty Analysis package (PEST; Doherty, 2015; Watermark Numerical Computing, 2020a, b) as the platform for optimization (hereafter referred to as the "two-watershed parameter set"). The two-watershed parameter set achieved better overall performance than the six parameter sets evaluated for the study period (WYs 1995 to 2011) and is used as the base-model parameter set in this study.

Watersheds and Model-Input Data

The study area in northeastern Illinois and northwestern Indiana, including land use, the nine gaged watersheds used in this study, the diverted watershed where the runoff is predicted using the HSPF model, and the raingages used in the HSPF modeling, is shown in figure 1. Summary information concerning the nine study watersheds is provided in table 1. The study watersheds are small, ranging in area from 11.2 to 107 mi^2 in drainage area, and have long discharge records, with seven of nine beginning in 1951 or earlier and the shortest record beginning in late 1989 (Hart Ditch at Dyer, Indiana [U.S. Geological Survey (USGS) station 05536179; hereafter referred to as "Hart Ditch at Dyer"]; U.S. Geological Survey, 2020). Discharge has been computed using stage-discharge ratings at all study streamgages except for Hart Ditch at Munster, Ind. (U.S. Geological Survey station 05536190; hereafter referred to as "Hart Ditch at Munster"), which has been operated differently because of backwater effects and, as a result, has reduced accuracy (Don Arvin, U.S. Geological Survey, written commun., August 15, 2013).

The diverted watershed area is covered by a precipitation gage network, the 25-gage Cook County Precipitation Network (CCPN; Bauer and Westcott, 2017), which began operation in 1990 (Peppier, 1991) and provides hourly precipitation data for LMDA HSPF simulations, which are simulated at an hourly time step but calibrated to published daily discharges. The calibration watersheds, Tinley Creek near Palos Park, Ill., monitored by USGS station number 05536500 [hereafter referred to as "Tinley Creek"] and Long Run near Lemont, Ill., monitored by USGS station number 05537500 [hereafter referred to as "Long Run"], are within the coverage of the CCPN, whereas for the other seven watersheds, the CCPN precipitation data are supplemented with data from eight National Weather Service Cooperative Observer Program (NWS–COOP) precipitation gages (National Weather Service, 2020). The NWS–COOP data were retrieved from the archive operated by the National Centers for Environmental Information (National Centers for Environmental Information, 2018). Except for part of the record at one station, which has hourly records, the eight NWS–COOP precipitation gages have daily records, and the daily precipitation records were subsequently disaggregated to hourly time steps for use in HSPF simulation. The Thiessen polygon method (Chow, 1964) was used for distributing precipitation data from a gage to its corresponding subwatershed. Other meteorological model inputs consist of measured air temperature, wind speed, and cloud cover from O'Hare Airport; solar radiation

Table 1. Background information for the U.S. Geological Survey streamgages for nine study watersheds.

[mi², square mile]

Station number (fig. 1)	Station name	Short name	Drainage area (mi²)	Impervious land[1] (percent)	Grassland (percent)	Forest (percent)	Cumulative area above dams/drainage area above streamgage	Period of published daily discharge[2]
05533000	Flag Creek near Willow Springs, Illinois	Flag Creek	16.5	37.6	60.4	2.0	0.0	7/26/1951–present
05535070	Skokie River near Highland Park, Illinois	Skokie River	21.1	31.4	55.5	13.1	0.0	8/21/1967–present
05536255	Butterfield Creek at Flossmoor, Illinois	Butterfield Creek	23.5	29.9	62.7	7.4	0.122	5/17/1948–present
05536340	Midlothian Creek at Oak Forest, Illinois	Midlothian Creek	12.6	36.6	53.3	10.1	1.207	10/1/1950–present
[3]05536500	Tinley Creek near Palos Park, Illinois	Tinley Creek	11.2	28.8	46.3	24.9	0.025	7/11/1951–present
[3]05537500	Long Run near Lemont, Illinois	Long Run	20.9	23.3	59.8	16.8	0.127	7/1/1951–present
05539000	Hickory Creek at Joliet, Illinois	Hickory Creek	107.5	23.2	64.1	12.8	0.101	10/1/1944–present
05536179	Hart Ditch at Dyer, Indiana	Hart Ditch at Dyer	37.6	7.7	69.9	22.5	0.0	9/19/1989–present
05536190	Hart Ditch at Munster, Indiana	Hart Ditch at Munster	70.7	19.6	63.4	17.0	0.0	10/1/1942–present

[1]Impervious land is the hydraulically connected impervious land; the fractions were determined based on the 2011 version of the 2006 National Land Cover Database (Homer and others, 2015).

[2]Present refers to year 2020.

[3]Calibration watersheds.

from Argonne National Laboratory; and potential evaporation computed from these data (Soong and Over, 2015). The hourly CCPN and meteorological time series were provided by the USACE (Jeff Fuller, USACE, written commun., November 2017; the hourly CCPN data were collected by the Illinois State Water Survey [ISWS] Prairie Research Institute and provided to USACE; the data are not available from the ISWS. Contact USACE for further information). The period for parameter calibration and discharge simulation is WYs 1996 to 2015. In HSPF modeling, the values of state variables used to start a simulation are not exactly known, and a simulation period of approximately a year is required for the results to converge; therefore, the period from WY 1997 to 2015 is used for analyzing the HSPF results. In this study, the discharge simulation period (WYs 1996 through 2015) is called the period of record (POR), and the period from which results are used in the analysis (WYs 1997 through 2015) is called the period of study (POS).

Areal coverages for three types of land covers in the nine watersheds were determined using the 2011 version of the 2006 National Land Cover Database (NLCD; U.S. Geological Survey, 2014; Homer and others, 2015). The NLCD land-cover classes were converted to the three HSPF land-cover types with a weight matrix (Sharpe and Soong, 2015). For catchments in the diverted watershed, Sharpe and Soong (2015) determined the fractions of the three LMDA land-cover types in each catchment changed little among the 2001, 2006, and 2011 versions of NLCD. However, such evaluation was not done for the nine study watersheds.

Flood-control reservoirs have been built in the study area since the 1960s (Resource Coordination Policy Committee, 1998). Reservoir(s) are present in the Midlothian Creek, Hickory Creek, Butterfield Creek, Long Run, and Tinley Creek watersheds. A ratio of cumulative area above dams to drainage area above streamgage was computed (table 1; Soong and Over, 2015), with a value of 0.025 for the Tinley Creek watershed and 0.127 for the Long Run watershed. The ratio can be greater than 1 if multiple reservoirs are above the gage because the area above each dam is counted independently, such as in the Midlothian Creek watershed (Soong and Over, 2015). Depending on their sizes, locations, and operations, flood-control reservoirs can attenuate the flood peaks and alter

the temporal distributions of streamflows, but these hydraulic effects are not modeled in Soong and Over (2015) nor in this study. In the HSPF land-cover types used in this report, reservoir areas are considered impervious surfaces, as are all areas classified as "open water" in the NLCD. Published discharges at Flag Creek near Willow Springs, Illinois (USGS station number 05533000; hereafter referred to as "Flag Creek"), Hickory Creek at Joliet, Ill. (USGS station number 05539000; hereafter referred to as "Hickory Creek"), Hart Ditch at Dyer, and Hart Ditch at Munster include wastewater-treatment plant effluent. Effluent discharge data were retrieved from U.S. Environmental Protection Agency websites and distributed from their reporting interval (most monthly) uniformly to hourly as reported in Soong and Over (2015). Reservoirs and effluent outflows can be assigned to the subwatershed where they are located because all nine watersheds have been divided into subwatersheds using the Thiessen method as described above. Unlike the LMDA HSPF models that do not have a flow routing component, simple flow routing functions (rudimentary stage-discharge relations based on topographic information) are included in the nine watershed models to approximate the hydraulic effects of streams within each watershed. In LMDA, flow routing is handled by a hydraulic model separately (for example, Espey and others, 2019). Tinley Creek and Long Run (fig. 1) were selected as calibration watersheds to develop the model parameter sets by Soong and Over (2015) because of the consistency of precipitation gage coverage by the CCPN, having different fractions of the grassland and forest lands and therefore providing an opportunity for determining parameters that distinguished between the two land covers, and the lack of effluent discharges.

Uncertainty of HSPF Model Simulations

Errors in model structure, in model parameter values, and in the data used for climatic forcing and calibration contribute to uncertainty in rainfall-runoff simulations. Even with an optimized parameter set, rainfall-runoff models are not capable of reproducing all aspects and portions of the hydrograph equally well, which is a consequence of structural error (Gupta and others, 2006). Structural error of a model originates from the underlying governing equations and numerical algorithms used to interpret the governing equations (Sitterson and others, 2017). In addition to structural error, errors in the model parameters and in the meteorological forcing and discharge data used for calibration, evaluation, and translating real-world watershed properties into the model's conceptualization also generate uncertainty in model-simulated discharges. For the LMDA HSPF model, the conversion of a mosaic of land uses into three land-cover types, using simplified routing that ignores the detailed hydraulic properties, the derivation of a regional parameter set, and the transferability of the regional parameters to the ungaged portion of the diverted watershed, contributes additional uncertainty to simulated discharges.

Calibration, also known as history matching, is the process of conditioning the model parameters to historical system response data (Tolson and Shoemaker, 2007). For rainfall-runoff models, the system response data generally are from the observed discharge time series at streamgages, from which components and statistics of hydrographs are selected to form an objective function to be minimized as parameters are adjusted (Lumb and others, 1994). Parameter calibration is an inverse problem in which parameter values are sought from a set of observations. The results depend on the quality of the data and representativeness of the model as well as parameters for the problem to be solved (Gábor and Banga, 2015). Because some HSPF model parameters cannot be calibrated from observed discharge data independently of each other (Lumb and others, 1994), parameter estimation for HSPF can be considered as an ill-posed inverse problem, which implies nonuniqueness of parameter sets (termed "equifinality" by Beven and Freer, 2001). Parameter nonuniqueness results in predictive nonuniqueness (Doherty and Johnston, 2003).

Through calibration, the inaccuracy of model simulations, represented as the simulation-to-observation misfits, can be reduced but not eliminated. However, the uncertainty remains because misfits may or may not stay within the same range as in the calibration when the calibrated parameter set are applied in predictions (Doherty, 2010). Structural error could be the dominant contributor to the misfits and forcing parameters to achieve minimum misfits may cause parameters to be overfitted (that is, given values that will make prediction errors larger than other values for which calibration errors are larger to compensate for errors in model structure; Doherty and Welter, 2010). Parameters may be assigned incorrect values if the processes they represent are not well-represented in the observations used in the objective function, and parameters of different governing equations in the same physical processes may become correlated in adjustments during the calibration (Doherty, 2010). Also, noise or errors in precipitation input and discharge data for calibration introduce additional errors to parameter calibrations.

Researchers have attempted to identify structural and parameter errors when the model is used to make predictions (for example, Doherty and Christensen, 2011; Diaz-Ramirez and others, 2013). Predictive error is estimated as the difference between validation data and model predictions based on the calibrated parameters from the calibration data (Gábor and Banga, 2015). The expected prediction errors can be rewritten as a combination of three terms: squared bias (where bias is the difference of the calibrated model prediction from the true validation data), variance of the prediction owing to the uncertainty of the parameters, and squared measurement error (Gábor and Banga, 2015). Doherty and Welter (2010) showed the deleterious effect of structural error on the model-calibration process and investigated a variety of ways to mitigate the effect. Regularization (Moore and Doherty, 2005; Hunt and others, 2007), where expert knowledge about parameter values and hydrologic system responses is introduced as prior information and used to construct an additional objective

function alongside the one based on observed data, can lead to a conditionally unique (that is, a different unique solution is expected if the expert information used is changed) solution of the ill-posed inverse problem. Regularization also induces tradeoffs between bias and variance of the two types of objective functions (Gábor and Banga, 2015). Gábor and Banga (2015) also demonstrated how regularized estimations ensure the best tradeoffs between the bias and variance, reducing overfitting and allowing the incorporation of prior knowledge in a systematic way. Doherty and Christensen (2011) used simple and complex models side by side to explore difficulties in model calibration and to quantify calibration-induced predictive bias. In summary, the model structural error is an integral term that represents the combination of structural deficiency, parameter compensation, and final parameters selected to represent the measurements. Separating the model uncertainty from parameter uncertainty is difficult not only because the probability distributions are not known (Doherty, 2010, 2015) but also because a unique and separate evaluation process is difficult to obtain for the structural errors (Michael Fienen, USGS, written commun., March 2021).

Transferring model parameterizations from gaged to ungaged watersheds generally has been implemented through regionalization efforts (Hrachowitz and others, 2013, who cited work by Dunn and Lilly, 2001, and Wagener and others, 2004). Hrachowitz and others (2013) indicated the success of parameter transferability improves with increasing physical similarity and spatial proximity between gaged and ungaged locations. However, the understanding of the processes underlying the basin responses and of the degree of feedback among these processes is incomplete at the scale of individual watersheds. Such deficiency in understanding of watershed-scale responses and of uncertainty of model predictions impedes the transfer of the understanding of hydrological response patterns obtained from gaged to ungaged basins (Hrachowitz and others, 2013).

Purpose and Scope

This study aims to quantify the uncertainties in the HSPF-simulated discharges at the ungaged catchments in the diverted Lake Michigan watershed used in LMDA computations. This report presents the first set of results of the study. The primary objectives of this report are to describe how uncertainty in the published discharge records used for HSPF parameter calibration is transmitted to calibrated parameter values and how the resulting parameter uncertainty affects simulated discharge volumes. The results are based on HSPF models of nine gaged watersheds in or near the diverted Lake Michigan watershed (see fig. 1 and discussions in the "Watersheds and Model-Input Data" section). Among the nine watersheds, the Tinley Creek and Long Run watersheds are used in model calibration and the other seven are used for discharge predictions. Two secondary objectives of the report are to describe (1) the adaption and application of a method for estimating the uncertainty of published daily discharge records at the calibration watersheds, and (2) the uncertainty estimate of simulated discharge arising from parameter uncertainty at the calibration and prediction watersheds.

All the HSPF PEST calibrations and HSPF simulations were forced by hourly climatic data, whereas the simulated discharge data were aggregated from hourly to daily time steps to compare with the published streamgage records (Soong and Over, 2015, p. 13). The period of parameter calibration and simulation is from WY 1996 to 2015. Following the guidance of the TRCs for the LMDA, the parameters being considered in calibration are the parameters used in HSPF to model runoff generation from pervious land segments. Parameters used to model runoff from impervious land segments, routing in streams, and snow processes were not considered. The analysis is based on the pervious land segment parameter values obtained from the two-watershed calibration procedure of Soong and Over (2015).

Methods

The effects of the uncertainty of published discharge records and of the uncertainty of model parameters on HSPF predictions for LMDA are characterized in this study with two sets of HSPF simulations, one for each type of uncertainty, and these two sets of simulations are based on investigations of the underlying uncertainty. The models, input, and output data are available from Over and others (2022). Published discharge uncertainty is estimated at the two calibration watersheds, Tinley Creek and Long Run, with a simulation approach to computed discharge uncertainty estimation based on Bayesian inference (Everitt and Skrondal, 2010, p. 37) that was developed by Mansanarez and others (2019). The uncertainty of the HSPF model parameters is also characterized with a simulation approach, which is based on an estimate of the covariance matrix of the parameters of a base model. This base model and its parameter set are derived using the two-watershed calibration scheme of Soong and Over (2015) using the calibration watersheds and the published discharge records for the period from WY 1996 to 2015.

The simulations used to characterize the effect of uncertainty of the published discharge records are based on parameters obtained by recalibrating the base model with discharge series selected from realizations of the uncertain published discharges from the calibration watersheds to derive respective optimal parameter sets. The simulations used to characterize the parameter uncertainty are based on parameters obtained from Monte Carlo simulations of parameter sets based on the estimated covariance matrix of the base-model parameters. The base, recalibrated, and random parameter sets are then used for HSPF simulations from the nine study watersheds (that is, the two calibration and the seven prediction watersheds). In this study, the uncertainties are analyzed using selected discharge summary statistics: the POS (WYs 1997

to 2015) mean discharge, selected flow-duration curve (FDC) nonexceedance probability quantiles, and WY mean discharges.

This section has four subsections. In the first, the estimation of the uncertainty of the published discharge records at the two calibration watersheds is described. In the second, the two-watershed calibration scheme adopted with modification from Soong and Over (2015) and used for recalibrating parameters with uncertain published discharge time series is reviewed. In the third, the estimation of the uncertainty of the base-model parameters and the simulation of corresponding random parameter sets are presented. In the fourth and final subsection, the recalibration of the HSPF model to discharge series selected from those characterizing the uncertainty of the published discharge is discussed.

Estimating Uncertainty of Published Discharge

As discussed previously, the discharge series at the streamgages used in this study, particularly those at the calibration streamgages (Tinley Creek and Long Run), are computed with stage-discharge rating curves. That is, for each stage measurement h_t (which are recorded with a resolution of plus or minus 0.01 foot (ft) and at time steps t separated usually by 15 minutes for these streamgages), the discharge is computed as $\widehat{Q}_{USGS,t} = f_{RC}(h_t)$, where $f_{RC}(h_t)$ is the rating curve, which is usually a monotonically increasing function, and the $\widehat{\ }$ notation indicates that computed discharge $\widehat{Q}_{USGS,t}$ is an estimate. Shapes of these rating curves vary with time as the stream-channel conditions change for a variety of reasons, including scour and fill of the channel during high-flow events, seasonal and long-term (years) changes in channel vegetation, and human channel modifications. In USGS streamgaging practice, rating curves are developed, adjusted, and revised following standard USGS procedures based on field observations and analysis that may include considerable judgement by the hydrographer (Kennedy, 1983, 1984).

The USGS rating-curve development process is described briefly here to provide a qualitative understanding of the uncertainty involved and to provide perspective on the applicability of the selected uncertainty estimation method. A rating curve is a curve fit to pairs of stage and discharge measurements. The curve is usually fit by the hydrographer who made the measurements and is fit by judgement, based on their knowledge of stream conditions (for example, stages corresponding to changes in hydraulic control, such as the top of the channel banks) and of the discharge measurement uncertainties, which are recorded using qualitative codes assigned at the time of the measurement (Turnipseed and Sauer, 2010). Hydrographers continuously evaluate the rating with ongoing discharge measurements and note changes observed during site visits. Based on the discharge measurements, ratings are adjusted locally using temporary shifts when measurements plot farther from the rating curve than the measurement uncertainty. The shape of the shift as a function of stage and time is determined based on site conditions and the stage record. The shifts may be adjusted repeatedly when new discharge measurements are made. When a consistent pattern of temporary shifts in one direction is observed, a new rating curve is developed. Because high-flow measurements are rare and because shifts often affect primarily the lower flows, high-flow measurements from previous years are used in the new rating curve, whereas low-flow measurements are primarily taken from the recent period when the consistent shifts were observed.

Because of the high degree of judgement involved in the determination and adjustment of stage-discharge rating curves following the USGS development process, no quantitative method for assigning uncertainty to the rating curves or to the resulting computed discharge has been developed. Lacking such a method, the approach taken here is to apply a state-of-the-art automated method for rating-curve fitting and discharge computation with built-in uncertainty estimation and to adapt the uncertainty results to the published USGS daily discharge by a ratio technique.

The selected automated method is the stage-period-discharge (SPD) extension of the Bayesian rating curve estimation (BaRatin) method (hereafter referred to as BaRatin SPD) proposed by Mansanarez and others (2019), which extends the original BaRatin method (Le Coz and others, 2014) to rating curve and discharge computation at streamgages with multiple known rating-curve periods. Bayesian inference is used in the BaRatin method, computed by using Markov Chain Monte Carlo sampling (Everitt and Skrondal, 2010, p. 269). This method is used to combine prior information regarding the hydraulic geometry of the streamgage with the stage-discharge measurement pairs and their uncertainties to simulate a set of equally likely posterior (Everitt and Skrondal, 2010, p. 334) rating curves that characterize the rating curve and its uncertainty. The simulated posterior rating curves can then be used to predict equally likely discharge time series given a recorded stage-time series. In the SPD extension of BaRatin, the user specifies the rating-curve periods and which parameters are thought to vary by period (for example, the elevation of the lowest control) and which parameters are static across all time periods. The user also defines and parameterizes models for the development of the period-dependent parameters, beginning with the present period and moving back in time. Based on these specifications and stage-discharge pairs for each rating-curve period, BaRatin SPD computes posterior realizations of period-dependent rating curves. The BaRatin SPD method does not explicitly simulate temporary shifts within a rating-curve period. Instead, the variations in the rating curve indicated by the stage-discharge measurements that were used to construct the shifts in the USGS method will appear as scatter around the fitted rating curve and, thus, will be reflected in the uncertainty of the estimated posterior rating curves produced by the BaRatin SPD method. The magnitude of this component of

the BaRatin SPD rating-curve uncertainty is unknown because a method to estimate the uncertainty of shifts in the USGS method has not been developed.

Incorporating Uncertainty in Rating-Curve Estimation

The channel at the Tinley Creek streamgage (fig. 1) is winding with a gravel and loam bed and vegetated banks and a forest- and grass-covered flood plain on the right side looking downstream (fig. 2). The low-water control is a gravel bar about 50 ft downstream from a bridge near where the continuous stage-measurement equipment is housed. Shifting of the stage-discharge rating is appreciable because of the accumulation of leaves and debris on the low-water control, scour and fill during floods, and effects of ice. Furthermore, debris often collects downstream from the streamgage, which can cause backwater at medium and high stages. The streamgage was established in 1951 and has a drainage area of 11.2 mi^2 (table 1).

The Long Run streamgage (fig. 1) was also established in 1951 at a bridge crossing the stream and has a drainage area of 20.9 mi^2 (table 1). The channel at the streamgage has a sand and fine loam bed with low and heavily wooded banks and flood plains (fig. 3). The low-water control has moved twice during the HSPF analysis period. Prior to April 1995, the control was a gravel bar about 200 ft downstream from the streamgage. Subsequently, the control became a rock riffle about 1000 ft downstream, and since May 9, 2001, the low-water control has been a rock dam built about 20 ft downstream from the bridge. Flows at the streamgage can have large shifts because of vegetation changes, debris jams, and beaver dams.

The rating-curve periods that cover the HSPF analysis period and the number and range of discharge measurements within each at the Tinley Creek and Long Run streamgages are given in table 2. Historical high-flow measurements used for rating-curve fitting by the USGS are not included in the table.

Discharge measurements at the streamgages shown in table 2 during the HSPF analysis period were made using mechanical current meters until 2005, when the technology was switched to acoustic methods, primarily acoustic Doppler velocimeters, with acoustic Doppler current profiler measurements when flows were deeper. At shallow depths (4.5 ft at Tinley Creek and 3.1 ft at Long Run), velocities were obtained while wading, otherwise velocities were measured from the bridge at the streamgage. In the USGS, when mechanical current meters and acoustic Doppler velocimeters are used, discharge is generally computed using the midsection method (Turnipseed and Sauer, 2010). To estimate discharge measurement uncertainty for use with BaRatin SPD, the subjective field-assessed discharge measurement qualities of Excellent, Good, Fair, and Poor were mapped to standard deviations of 3, 5, 8, and 15 percent, respectively (Turnipseed and Sauer, 2010). When the measurement quality was not specified, a standard deviation of 15 percent was also used.

Figure 2. Downstream view at the low-water control and overbank at the Tinley Creek near Palos Park, Illinois, streamgage (U.S. Geological Survey [USGS] station 05536500) on December 2, 2003. Gage height was 2.45 feet, and discharge was 3.47 cubic feet per second. Photograph was taken by Perry Draper, USGS.

Figure 3. Downstream view at the low-water control and overbank at the Long Run near Lemont, Illinois, streamgage (U.S. Geological Survey [USGS] station 05537500) on April 22, 2008. Gage height was 1.07 feet, and discharge was 12.0 cubic feet per second. Photograph was taken by David Schrader, USGS.

Table 2. Properties of U.S. Geological Survey (USGS) rating-curve periods that cover water years 1995 to 2016 at Tinley Creek near Palos Park, Illinois (USGS station 05536500), and Long Run near Lemont, Illinois (USGS station 05537500) streamgages.

[SPD, stage-period-discharge; Qms, discharge measurements; ft^3/s, cubic feet per second]

Rating-curve number	BaRatin SPD period	Range of dates	Number of Qms	Range of Qms (ft^3/s)
		Tinley Creek		
23	5	10/05/1991–11/11/1995	29	0.29–121
24	4	11/11/1995–11/24/1999	36	0.50–2,100
25	3	11/24/1999–10/15/2007	54	0–189
26	2	10/15/2007–10/26/2012	31	0.057–89.9
27	1	10/26/2012–02/14/2017	28	0.13–870
		Long Run		
27	6	10/01/1990–07/17/1996	39	0.613–315
28	5	07/17/1996–03/09/1998	12	1.40–2,740
29	4	03/09/1998–01/23/1999	7	1.90–35.5
30	3	01/23/1999–10/05/2001	22	1.79–246
31	2	10/05/2001–10/25/2010	53	1.95–648
32	1	10/25/2010–01/24/2017	39	0.54–903

Two hydraulic configurations for the application of BaRatin SPD to the streamgages in the calibration watersheds were tested, and the weir-channel-flood plain (WCF) configuration, which combines a rectangular weir (for low flows), a channel (for in-channel flows), and a channel with flood plain (for overbank flows) to simulate the hydraulics at both streamgages was selected. Following the notation of Mansanarez and others (2019), the general flow formula for the weir is

$$Q_h = C_r B_w \sqrt{2g}\left(h-b\right)^c = a\left(h-b\right)^c \quad (1)$$

where

Q_h is the discharge being computed;
C_r is the discharge coefficient;
B_w is the effective weir width perpendicular to the flow;
g is the acceleration of gravity;
h is the stage;
b is the offset;
c is an exponent that generally is taken to be 3/2 for a rectangular weir; and
a is the stage factor for a weir, equal to $C_r B_w \sqrt{2g}$.

Also following the notation of Mansanarez and others (2019), the general flow formula for the wide rectangular channel is

$$Q_h = K_s B_c \sqrt{S_0}\left(h-b\right)^c = a\left(h-b\right)^c \quad (2)$$

where

Q_h is the discharge being computed;
K_s is the Strickler flow resistance coefficient;
B_c is the effective channel width;
S_0 is the bed slope;
h is the stage;
b is the offset;
c is an exponent that generally is taken to be 5/3 for a wide rectangular channel; and
a is the stage factor for a wide rectangular channel, equal to $K_s B_c \sqrt{S_0}$.

Assuming only the weir and channel offsets $b_1^{(k)}$ and $b_2^{(k)}$ vary by rating-curve period k, the discharge estimated with WCF configuration for the kth rating period is expressed as

$$\widehat{Q}_{h,k} = \begin{cases} a_1\left(h-b_1^{(k)}\right)^{c_1}, \text{ if } \kappa_1^{(k)} \le h < \kappa_2^{(k)} \\ a_2\left(h-b_2^{(k)}\right)^{c_2}, \text{ if } \kappa_2^{(k)} \le h < \kappa_3 \\ a_2\left(h-b_2^{(k)}\right)^{c_2} + a_3\left(h-b_3\right)^{c_3}, \text{ if } h \ge \kappa_3 , \end{cases} \quad (3)$$

where

$\widehat{Q}_{h,k}$ is the estimated discharge at stage h for the kth rating period;
a_1 is the stage factor for the first (low-flow) control modeled as a weir, that is, $C_r B_w \sqrt{2g}$, where C_r is the discharge coefficient, B_w is the effective weir width perpendicular to the flow, and g is gravitational acceleration;
h is stage;
$b_1^{(k)}$ is the stage offset for the first control and the kth rating-curve period;
c_1 is the stage exponent for the first control;
$\kappa_1^{(k)}$ is the activation stage for the first control and the kth rating period;
$\kappa_2^{(k)}$ is the activation stage for the second control and the kth rating period;
a_2 is the stage factor for the second (channel) control modeled as a wide rectangular channel, that is, $K_s B_c \sqrt{S_0}$, where K_s is the Strickler flow resistance coefficient, B_c is the effective channel width, and S_0 is the bed slope;
$b_2^{(k)}$ is the stage offset for the second control and the kth rating-curve period;
c_2 is the stage exponent for the second control;
a_3 is the stage factor for the flood plain part of the third control, that is, $K_s B_c \sqrt{S_0}$ for the flood plain, where K_s is the Strickler flow resistance coefficient, B_c is the effective channel width, and S_0 is the bed slope; and
κ_3 is the activation stage for the third control.

The values of the activation stages are deduced from continuity (Mansanarez and others, 2019). When a control is added, as defined here beginning with the low-flow weir and when moving from the main channel control to the main channel with flood-plain control, then continuity implies that $k_i=b_i$ (therefore, $k_1=b_1$ and $k_3=b_3$); however, when a control is replaced, as defined here when moving from weir control to channel control, numerical solution for the crossing of the two control equations is used, for which there may be 0, 1, or 2 solutions (Mansanarez and others, 2019).

The hydraulic-geometry dimensions and prior values of the parameters to be used for Bayesian inference by BaRatin SPD were estimated from the measurement section of the most recent rating period and using prior probability distributions following guidance and examples in Le Coz and others (2014) and Mansanarez and others (2019). Ideally, BaRatin SPD is parameterized using measured cross-section geometry; however, such information is not typically used by the USGS in the development of the stage-discharge ratings (other than zero-flow and top-of-bank stages). Therefore, prior parameter values used in this study were based on estimates rather than physical measurements. Widths of the low-flow weir and channel were estimated from field notes and aerial maps. The Strickler coefficients for the channel and flood plain were

estimated from field notes and photographs. Bed slope was estimated from topographic maps. In this study, for the initial prior parameter estimates at the Tinley Creek streamgage site (figs. 1 and 2), the continuity equation for the weir-channel activation k_2 stage had no solution, and adjustments had to be made by trial-and-error. The final specified and derived prior parameter values and probability distributions for the application of BaRatin SPD at the Tinley Creek and Long Run streamgages for this study are given in tables 3 and 4, respectively. The incremental change parameters $\delta l^{(k)}$ and $\delta g^{(k)}$ simulate the local (weir-only) and global (weir and channel) shifts in stage between rating-curve periods, respectively. The effects of the shifts were incorporated into the prior distributions of $b_1^{(k)}$and $b_2^{(k)}$, $k > 1$, by Monte Carlo simulation, using R language computer code (R Core Team, 2019) given in Mansanarez and others (2019).

The stage-discharge measurement pairs for each rating period, along with the uncertainty of the discharge measurements inferred from the quality of the measurement noted by the hydrographers, are then used to infer the static and variable parameter values. The BaRatin SPD estimated discharge $\hat{Q}$ is expressed as

$$\hat{Q} = f(h, k \,|\, \boldsymbol{\beta}), \tag{4}$$

where

- $f(\,)$ is the function giving $\hat{Q}$;
- h is stage;
- k is the rating-curve period;
- $\boldsymbol{\beta}$ is the vector of static and period-specific rating-curve parameters; and
- $k|\boldsymbol{\beta}$ represents the condition of k given $\boldsymbol{\beta}$.

The parameters are inferred using uncertain stage, discharge, and period definition data. The error model for measured discharges is

$$\widetilde{Q}_i = Q_i + \varepsilon_{Qi}, \tag{5}$$

where

- $\widetilde{Q}_i$ is the ith discharge measurement,
- Q_i is the true discharge corresponding to the ith measurement, and
- ε_{Qi} is its measurement error, the collection of which (for all i) is assumed to be mutually independent and distributed as $\varepsilon_{Q_i} \sim N\left(0, \sigma_{Q_i}\right)$; that is, a normal (Gaussian) distribution with mean 0 and standard deviation σ_{Q_i}. The standard deviations σ_{Q_i} are specified depending on discharge magnitude and measurement quality through the mapping of the field-assessed measurement qualities to standard deviations in percent as described above.

In terms of the BaRatin SPD rated discharge (eq. 3), the ith true discharge Q_i is given as

$$Q_i = \widehat{Q}_i + \varepsilon_{f,i} = f\left(h_i, k_i | \boldsymbol{\beta}\right) + \varepsilon_{f,i}, \tag{6}$$

where

- $\hat{Q}_i$ is the BaRatin SPD-estimated discharge corresponding to the ith discharge measurement;
- $\varepsilon_{f,i}$ is a structural error term, whose values are assumed to be mutually independent and distributed as $\varepsilon_{f,i} \sim N\left(0, \gamma_1 + \gamma_2 \widehat{Q}_i\right)$. Therefore, the structural error term has mean zero, but its standard deviation is dependent on the discharge. The added parameters γ_1 and γ_2 are assumed to be unknown and are estimated as part of the inference process. The prior vdistributions used here for the structural error parameters γ_1 and γ_2 are given in tables 3 and 4;
- $f(\,)$ is the function giving $\widehat{Q}_i$;
- h_i is the stage corresponding to the ith discharge measurement;
- k_i is the rating-curve period corresponding to the ith discharge measurement;
- $\boldsymbol{\beta}$ is the vector of rating-curve parameters, and
- $k|\boldsymbol{\beta}$ represents the condition of k given $\boldsymbol{\beta}$.

Combining equations 5 and 6 to eliminate the true discharge yields

$$\widetilde{Q}_i = f\left(h_i, k_i | \boldsymbol{\beta}\right) + \varepsilon_{Qi} + \varepsilon_{f,i}, \tag{7}$$

where

- $\widetilde{Q}_i$ is the ith discharge measurement,
- $f(\,)$ is the function giving Q_i,
- h_i is the stage corresponding to the ith discharge measurement,
- k_i is the rating-curve period corresponding to the ith discharge measurement,
- $\boldsymbol{\beta}$ is the vector of rating-curve parameters,
- $k_i|\boldsymbol{\beta}$ represents the condition of k_i given $\boldsymbol{\beta}$,
- ε_{Qi} is the measurement error of the ith discharge measurement, and
- $\varepsilon_{f,i}$ is the structural error term.

Equation 7 gives the distribution of a measured discharge $\widetilde{Q}_i$ in terms of the measured stage h_i, the rating period k_i, the vector of rating-curve parameters $\boldsymbol{\beta}$, the measurement error ε_{Qi}, and the structural error $\varepsilon_{f,i}$. Because the measurement and structural errors are assumed to be Gaussian, the distribution of $\widetilde{Q}_i$ is likewise Gaussian. Based on equation 7, the joint

Table 3. Prior information for physical parameters and their distributions used for BaRatin stage-period-discharge (SPD) analysis of Tinley Creek near Palos Park, Illinois, streamgage (U.S. Geological Survey station 05536500).

[ft, foot; $N(\mu; \sigma)$, Gaussian distribution with mean μ and standard deviation σ; $LN(\mu_{log}; \sigma_{log})$, lognormal distribution with median $e^{\mu_{log}}$ and coefficient of variation $\sqrt{e^{\sigma_{log}} - 1}$ and whose logarithm has mean μ_{log} and standard deviation σ_{log}; ft/s^2, foot per squared second; --, not applicable; m$^{1/3}$/s, cube root of a meter per second; <, less than; ft^3/s, cubic foot per second]

Specified parameter	Definition	Prior distribution	Inferred parameter	Definition	Prior distribution
			Control 1—Low-flow weir		
b_1	Offset (ft)	N(2.316; 0.656)	$b_1^{(1)}$	Offset, period 1 (ft)	N(2.316; 0.656)
B_w	Weir width (ft)	$LN(ln(6.562); 0.246)$	$b_1^{(2)}, \ldots, b_1^{(5)}$	Offsets, periods 2–5 (ft)	([1])
C_r	Discharge coefficient (unitless)	$LN(ln(0.4); 0.246)$	a_1	Stage factor for a weir, equal to $C_r B_w \sqrt{2g}$ (eq. 1)	$LN(ln(21.06); 0.349)$
g	Gravitational acceleration (ft/s^2)	$LN(ln(32.19); 0.0005)$	--	--	--
c_1	Exponent (unitless)	N(1.581; 0.025)	--	--	--
			Control 2—Channel flow		
b_2	Offset (ft)	N(2.339; 0.656)	$b_2^{(1)}$	Offset, period 1 (ft)	N(2.339; 0.656)
B_{c_2}	Channel width (ft)	$LN(ln(32.81); 0.2)$	$b_2^{(2)}, \ldots, b_2^{(5)}$	Offsets, periods 2–5 (ft)	([1])
K_{S_2}	Strickler coefficient (m$^{1/3}$/s)	$LN(ln(25); 0.1)$	a_2	Stage factor for the channel, equal to $K_{s_2} B_{c_2} \sqrt{S_0}$ (eq. 2)	$LN(ln(47.29), 0.324)$
S_{0_2}	Bed slope (unitless)	$LN(ln(0.0015); 0.5)$	--	--	--
c_2	Exponent (unitless)	N(1.631; 0.025)	--	--	--
			Control 3—Flood plain plus channel flows		
b_3	Offset (ft)	N(8.061; 0.820)	a_3	Stage factor for the flood plain, equal to $K_{s_2} B_{c_3} \sqrt{S_0}$ (eq. 2)	$LN(ln(271.9); 0.449)$
B_{c_3}	Flood-plain width (ft)	$LN(ln(262.5); 0.3)$	--	--	--
K_{S_3}	Strickler coefficient (m$^{1/3}$/s)	$LN(ln(18); 0.25)$	--	--	--
S_{0_3}	Bed slope (unitless)	$LN(ln(0.0015); 0.5)$	--	--	--
c_3	Exponent (unitless)	N(1.67; 0.05)	--	--	--
			Incremental change in stage (2<k<5)		
$\delta l^{(k)}$	Local change (ft)	N(0; 1.64)	--	--	--
$\delta g^{(k)}$	Overall change (ft)	N(0; 1.64)	--	--	--
			Structural error		
γ_1	Intercept (ft^3/s)	$LN(ln(35.31); 1)$	--	--	--
γ_2	Slope (unitless)	$LN(ln(0.5); 1)$	--	--	--

[1]See "Uncertainty of Stage-Discharge Rating Curves" section.

Table 4. Prior information for physical parameters and their distributions used for BaRatin stage-period-discharge (SPD) analysis of Long Run near Lemont, Illinois, streamgage (U.S. Geological Survey station 05537500).

[ft, foot; $N(\mu; \sigma)$, Gaussian distribution with mean μ and standard deviation σ; $LN(\mu_{log}; \sigma_{log})$, lognormal distribution with median $e^{\mu_{log}}$ and coefficient of variation $\sqrt{e^{\sigma_{log}} - 1}$ and whose logarithm has mean μ_{log} and standard deviation σ_{log}; ft/s^2, foot per squared second; --, not applicable; m$^{1/3}$/s, cube root of a meter per second; <, less than; ft^3/s, cubic foot per second]

Specified parameter	Definition	Prior	Inferred parameter	Definition	Prior
			Control 1—Low-flow channel		
b_1	Offset (ft)	$N(0.307; 0.656)$	$b_1^{(1)}$	Offset, period 1 (ft)	$N(0.304; 0.656)$
B_w	Weir width (ft)	$LN(ln(3.937); 0.246)$	$b_1^{(2)}, \ldots, b_1^{(6)}$	Offsets, periods 2–6 (ft)	([1])
C_r	Discharge coefficient (unitless)	$LN(ln(0.4); 0.246)$	a_1	Stage factor for a weir, equal to $C_r B_w \sqrt{2g}$ (eq. 1)	$LN(ln(12.62); 0.349)$
g	Gravitational acceleration (ft/s^2)	$LN(ln(32.2); 0.0005)$	--	--	--
c_1	Exponent (unitless)	$N(1.5; 0.025)$	--	--	--
			Control 2—Channel flow		
b_2	Offset (ft)	$N(0.504; 0.656)$	$b_2^{(1)}$	Offset, period 1 (ft)	$N(0.499; 0.656)$
B_{c_2}	Channel width (ft)	$LN(ln(26.25); 0.246)$	$b_2^{(2)}, \ldots, b_2^{(6)}$	Offsets, periods 2–6 (ft)	([1])
K_{S_2}	Strickler coefficient (m$^{1/3}$/s)	$LN(ln(24); 0.198)$	a_2	Stage factor for the channel, equal to $K_{s_2} B_{c_2} \sqrt{S_0}$ (eq. 2)	$LN(ln(29.61), 0.339)$
S_{0_2}	Bed slope (unitless)	$LN(ln(0.001); 0.246)$	--	--	--
c_2	Exponent (unitless)	$N(1.667; 0.025)$	--	--	--
			Control 3—Flood plain plus channel flows		
b_3	Offset (ft)	$N(3.281; 0.2)$	a_3	Stage factor for the flood plain, equal to $K_{s_3} B_{c_3} \sqrt{S_0}$ (eq. 2)	$LN(\ln(43.21); 0.339)$
B_{c_3}	Flood-plain width (ft)	$LN(ln(114.8); 0.246)$	--	--	--
K_{S_3}	Strickler coefficient (m$^{1/3}$/s)	$LN(ln(8); 0.198)$	--	--	--
S_{0_3}	Bed slope (unitless)	$LN(ln(0.001); 0.246)$	--	--	--
c_3	Exponent (unitless)	$N(1.667; 0.025)$	--	--	--
			Incremental change in stage ($2<k<6$)		
$\delta l^{(k)}$	Local change (ft)	$N(0; 1.64)$	--	--	--
$\delta g^{(k)}$	Overall change (ft)	$N(0; 1.64)$	--	--	--
			Structural error		
γ_1	Intercept (ft^3/s)	$LN(ln(35.31); 1)$	--	--	--
γ_2	Slope (unitless)	$LN(ln(0.5); 1)$	--	--	--

[1]See "Uncertainty of Stage-Discharge Rating Curves" section.

likelihood across all measurements can be derived and the prior distributions specified, which are in turn combined to compute the posterior distribution. For details, see Mansanarez and others (2019).

Using the characterization of the prior parameters as discussed previously and the posterior distribution developed based on equation 7, inference of the BaRatin SPD model parameters was determined using a Markov Chain Monte

Carlo sampler as described in and using the R language computer code provided in Mansanarez and others (2019). Although the numerical parameters of the Markov Chain Monte Carlo sampler can be adjusted by the user, the default values were used for this study. These default values include performing 100,000 realizations, of which the first half were discarded as burn-in, which is the initial set of iterations needed before the samples reach the posterior distribution (Everitt and Skrondal, 2010, p. 66). Then, to avoid autocorrelation, every 100th of the latter 50,000 realizations were retained, resulting in 500 quasi-stationary, quasi-independent realizations of rating-curve parameters for further use in this study. The sequences of rating-curve parameter values were examined visually and analyzed numerically to verify stationarity and independence.

Prediction of Published Discharge Uncertainty

To obtain estimates of discharge uncertainty, the rating-curve realizations simply need to be forced with appropriate stage-data time series. Before their use with the rating-curve realizations, the stage data at the streamgages in the calibration watersheds, which were available mostly at a 15-minute time step on the quarter hour, were preprocessed as follows: (1) the data were interpolated to be uniformly on the quarter hour; (2) the data were analyzed for gaps, and calendar days with data gaps longer than 12 hours or fewer than 48 stage values were set to missing; and (3) 500 realizations (1 per rating-curve realization) of independent mean-zero random Gaussian noise with a standard deviation of 0.01 ft (following USGS stage accuracy requirements [Sauer and Turnipseed, 2010, p. 3]) were added to the interpolated stage data (hereafter referred to as "noisy stage data").

The 500 realizations from the discharge time series, simulated at 15-minute intervals that were obtained from applying the rating curves to the realizations of noisy stage data, were post-processed following three steps: (1) the discharge data were aggregated to daily mean values using the trapezoidal rule (Greenspan and Benney, 1973, p. 287–291); (2) missing days and days that were estimated in the published record were estimated; and (3) the resulting daily discharge realizations were rescaled to the published values so that the mean of the realizations agreed with the published value each day. Further details on the second and third post-processing steps are provided in the following paragraphs.

The missing days and days when discharges were estimated in the published record were determined using USGS methodology of hydrographic comparison (Rantz, 1982, chap. 15). This methodology is based on use of adjusted versions of observations at comparison streamgages with similar hydrographs. Such similarity is usually a result of similar basin size and proximity. Thus, these basins would be expected to have similar weather and physiography. The hydrographic comparison estimates were applied at a daily time scale and were made by adjusting the discharge at the most correlated available comparison station to equal the discharge on the last day before the missing period or period to be estimated and on the first day after the period. Because the discharge adjustment involves matching the values before and after the estimated periods, it is customized for each realization and thus randomized. The days that were estimated in the published discharge record were estimated with this hydrographic comparison methodology in the BaRatin SPD realizations, even if some stage data were available on those days, for two reasons: (1) because the published record was not computed using the stage-discharge rating on those days, it would not be meaningful to compare to an estimate obtained in that manner; and (2) often these days were estimated because, although stage data were available, they were known to be invalid for discharge computation, such as days when the stream was ice-covered. During WYs 1995–2016, which consists of 8,036 days, 200 days were determined as missing in the BaRatin SPD discharge realizations for the records at the Tinley Creek streamgage, including 4 days that were not considered as estimated in the published record. Those 4 days, plus the 881 days that were indicated as estimated in the published record, were estimated in the BaRatin SPD realizations. The realizations at the Long Run streamgage during the same period had no missing days, but the published USGS record had 875 estimated days, and those 875 days were also estimated in the BaRatin SPD realizations.

The final post-processing step, rescaling, is crucial and was applied to adapt the uncertainty estimate provided in the BaRatin SPD realizations to the published discharge record, which is used as the basis for the rainfall-runoff model calibrations. BaRatin SPD realizations do not generally center on the published values, and this result is expected because these realizations were obtained using an independent, alternative rating-curve estimation procedure. To address this discrepancy, the BaRatin SPD realizations were rescaled to enforce day-by-day centering on the published daily record using the following equation:

$$Q_{j,T}^{pred.rescaled} = Q_{j,T}^{pred} \times \left[Q_T^{record} / \overline{Q_T^{pred}} \right], \qquad (8)$$

where

$Q_{j,T}^{pred.rescaled}$ is the jth rescaled BaRatin SPD realization on day T;

$Q_{j,T}^{pred}$ is the jth realization on day T before rescaling;

Q_T^{record} is the published daily value on day T; and

$\overline{Q_T^{pred}}$ is the average over the 500 realizations on day T.

It can be deduced from equation 8 that when $Q_T^{record} / \overline{Q_T^{pred}}$ is less than one, the variance of $Q_{j,T}^{pred.rescaled}$ will be reduced by rescaling; the extreme case of this variance reduction is when $Q_T^{record} = 0$; in that case, $Q_{j,T}^{pred.rescaled}$ will also be zero for all realizations. However, in this study, zero-valued published daily discharges are rare: the Tinley Creek record has 18 zero values during the BaRatin SPD simulation period, and the Long Run record has none. Overall, coefficients of variation are similar

before and after rescaling. The ratios of the published mean discharge to the mean of the mean simulated discharges before rescaling are 0.9996 and 1.0949, and the mean values of the daily rescaling ratios $Q_T^{record} / \overline{Q_T^{pred}}$ are 1.1808 and 1.1038, for the Tinley Creek and Long Run streamgages, respectively. Details on the properties of the rescaled BaRatin SPD discharge series realizations are presented in the "Uncertainty of Published Discharge" section.

Two-Watershed Calibration for Regional Hydrological Simulation Program—FORTRAN Parameters

The two-watershed HSPF calibration scheme used in this study is based on the scheme developed by Soong and Over (2015, p. 35–39) and is applied with some modifications. The scheme uses the Time Series PROCessor (TSPROC, Westenbroek and others, 2012), PEST, and two HSPF watershed models. TSPROC computes and transmits statistics and time series information needed for the computation of the objective function from the discharge time series that were simulated by HSPF with the current parameter set to PEST, which then seeks a set of optimal parameter values that obtain the minimum misfit for the measurement objective function, and passes the attained parameter set to the HSPF models for recomputing the discharge time series. The search stops when the preset convergence criterion is met. The PEST control files were set up as recommended in Doherty (2013), which includes use of regularized inversion to address parameter nonuniqueness and singular-value decomposition for numerical simulation stability (with the eigenvalue ratio threshold for truncation set at 5×10^{-7}; Doherty, 2013). The two calibration watersheds are Tinley Creek and Long Run, as explained above. The primary difference in the scheme used in this study compared to that of Soong and Over (2015, p. 37) is that the stopping condition was modified to limit the number of model iterations, as described below.

HSPF parameters related to runoff generation from pervious lands (abbreviated explanations of pervious land parameters analyzed are included in table 1.1) were optimized; the parameters of runoff generation from impervious lands, snow process-related parameters, and routing parameters remained fixed (see "Background" section). Calibration was limited to pervious land parameters in accordance with the work done for the nine historical parameter sets. All pervious land parameters except the INFEXP, INFILD, and BASETP parameters were set as adjustable. INFEXP and INFILD are generally not calibrated (U.S. Environmental Protection Agency, 2000), and BASETP is fixed at a value of 0.0 in the LMDA regional parameter sets (RUST Environment and Infrastructure, 1993; Espey and others, 2019). Thus, 17 parameters are in each of the 2 land-cover types (grassland and forest) that are calibrated with the discharge timeseries for a total of 34 adjustable parameters. For a full description of the HSPF parameters, see U.S. Environmental Protection Agency (2000).

Several transformations are applied to the parameters before their use in the calibration process, as was done in Soong and Over (2015). First, AGWRC and IRC (see table 1.1 for the definitions of parameters) were transformed as $Xtrans = X/(1–X)$, where X is either IRC or AGWRC, and are denoted as AGWRCTR and IRCTR thereafter. This transformation allows their values to cover a wider range during estimation. Second, to permit certain parameters to have different values between the forest and grassland segments but within a reasonable range of covariation, LZSN, INFILT, AGWRCTR, AGWETP, CEPSC, UZSN, INTFW, and LZETP for forest lands were all redefined as ratios to the grassland values. The ratio versions of these parameters are denoted with a suffix "R." Finally, all parameter values were $\log_{10}$-transformed. The logarithms of the parameters make inversion process more numerically stable (Watermark Numerical Computing, 2020a, p. 41) because additive changes in the transformed parameter values make proportional changes in the untransformed parameter values.

The measurement objective function comprises eight observation data groups organized from daily discharge records of the two calibration watersheds for WYs 1996–2015. The eight observation groups are as follows:

1. daily discharge,
2. daily quick flow derived from daily discharge by base-flow separation,
3. daily base flow derived from daily discharge by base-flow separation,
4. storm hydrographs for events with daily maxima above a selected threshold (200 cubic feet per second [ft^3/s] was used in the study),
5. monthly runoff volume,
6. annual runoff volume,
7. total runoff volume of the calibration period, and
8. FDC defined by 11 nonexceedance probabilities.

Magnitudes of the elements making up these groups (for example, the daily discharge values in the daily discharge group) may differ by orders of magnitudes; therefore, element weights were assigned to equalize the contribution of errors in each value to the group's contribution to the measurement objective function. For each element of a group, its weight was set equal to the inverse of its value, except for the FDC group for which a constant weight was used for each of the nonexceedance probabilities. Also, to avoid infinite weights associated with zero-valued daily discharges, 0.01 ft^3/s was added to the daily discharges before the inversion. Then, in addition to the weighting of the elements of the observation groups described above, the groups themselves were weighted to equalize their effect on the objective function (using the PWTADJ1 utility; Doherty, 2013, p. 3).

When using regularized inversion, as was done here, two objective functions are used: a measurement objective function and a regularization objective function. The PEST optimization is constrained in seeking a user-defined target measurement objective function value subject to the constraint that the regularization objective function (departures of parameters from their default values or preferred condition) is minimized (Doherty, 2010). The lower the measurement objective function value, the better the simulation to observation fit. The lower the regularization objective function, the less the parameter values depart from the expert knowledge or prior information expressed through the initial parameter values. The procedure for optimization when using PEST with regularization recommended by Doherty (2013) is to set the two variables that control the upper limit (PHIMLIM) and the acceptable level (PHIMACCEPT) of the measurement objective function in the "Regularization" block of the PEST control file to extremely low values (1.0^{-10} and 1.05^{-10}, respectively) for the first iteration. Users then update the PHIMLIM and PHIMACCEPT using the minimum measurement objective function value and the optimal parameter values attained in the first iteration, followed by iteratively adjusting the two variables and updating parameter values until PHIMLIM is more stable. For this study, to minimize the effect of expert knowledge in the calibrated parameter values, the process was stopped after the second iteration and the resulting parameter set was used if the resulting PHIMLIM was within about 2 percent of the value of PHIMLIM obtained after the first iteration. If the PHIMLIM obtained in the second iteration was much higher or lower than that achieved in the first iteration, then PHIMLIM was set to a value approximately 5 percent higher than the minimum PHIMLIM obtained in the previous iteration, but the parameter values were not updated. Iteration was then repeated with PHIMLIM reset up or down by 5 percent or less until the objective function did not vary by more than 2 percent of the original value, and the resulting parameter set was used.

For PEST to start a search, a set of initial parameter values and parameter bounds are assigned. For the base model, the initial parameter values and the lower and upper bounds determined from Soong and Over (2015) were adopted with the bounds slightly modified (table 1.1). For the following recalibrations that use uncertain published discharges, the optimal parameter values obtained from the base model were set to be the initial values in the new calibration.

Estimating Uncertainty of Base-Model Parameters

If linearity can be assumed for model outputs with respect to the model parameterization, the posterior probability function of parameters can be inferred from a Bayesian update of its prior probability density function using information learned from model calibration (Doherty, 2010, p. 79). However, the probability distribution of the prior parameters is unknown. If the distribution is assumed to be multivariate Gaussian (Doherty, 2015, p. 31–32) and if the data noise is also characterized by multi-Gaussian distribution, then the joint (multivariate Gaussian) probability distribution for the posterior (optimized) parameters can be calculated (Doherty, 2010, p. 79–80). For such a distribution, the width of the posterior parameter distribution in any direction of parameter space, and the degree to which random vector element indicates statistical interrelatedness, is characterized by a covariance matrix. The variances of individual parameters, together with the covariances between parameters, are combined in the covariance during the optimization with the covariance from measurement noise through the Jacobian matrix (Doherty, 2015, p. 55–56). In linear analysis, as presented here, predictive uncertainty and error variance depends only on the stochastic characterization of parameter variability as expressed by the covariance matrix.

The posterior covariance matrix is not generated by PEST as part of the optimization process in the model described here because of the use of regularization in the optimization process. To obtain a linear approximation to the posterior parameter covariance matrix, the PEST utility function PREDUNC7 (Watermark Numerical Computing, 2020b, p. 138) was used. PREDUNC7 is parameterized with Gaussian prior parameter uncertainty estimates, which can be a full covariance matrix or, as used in this study, standard deviation estimates for each adjustable parameter; parameter data from the PEST control file; and the Jacobian matrix computed by PEST during optimization iterations (Watermark Numerical Computing, 2020b). Using a vector of standard deviations instead of a full covariance matrix for the prior distribution of parameters is equivalent to an assumption of independence among the parameters; this assumption was made because of a lack of prior knowledge of the covariances among the parameters. The prior standard deviations were estimated using the bounds of the parameters (see table 1.1) by taking $(\log_{10}(\text{upper bound}) - \log_{10}(\text{lower bound}))/4$; note that in the application of PEST in this study, all adjustable parameters are $\log_{10}$-transformed. When estimating the prior standard deviations in this way, it is assumed that the $\log_{10}$-transformed parameters are normally distributed and the upper and lower bounds demarcate their 95-percent confidence intervals (CIs; Doherty, 2013, p. 11). PREDUNC7 then computes and writes a full covariance matrix pertinent to the posterior parameter probability distribution (Watermark Numerical Computing, 2020b) by extracting information regarding correlation among parameters from the Jacobian matrix. The posterior parameter covariance matrix computed by PREDUNC7, along with mean values from the $\log_{10}$-transformed base-model parameter values, was then sampled under the assumption of multivariate

Gaussianity using the PEST RANDPAR utility (Watermark Numerical Computing, 2020b, p. 164–166) to generate 1,000 random correlated parameter sets, where the sampled values were constrained to lie within the specified upper and lower bounds (table 1.1).

Model Recalibration with Uncertain Published Discharge Time Series

The discharge series realizations derived from the BaRatin SPD analysis at the two streamgages in the two calibration watersheds are independent from each other; therefore, the simplest approach to selecting datasets for recalibration would be as random selection of a large number (that is, from 100 to all 500 of them) of samples from the two sets of discharge series realizations. This approach would also be the most useful, because, being based on random sampling, the results would approximate the full probability distribution of uncertainty arising from uncertainty in published discharge and could be added to the uncertainty arising from the model parameters. However, random sampling of such a large number of these series is beyond the scope of this study because the subsequent calibration is computationally demanding and is not automatable because the stopping criterion requires expert judgement.

Instead of random sampling from the two sets of 500 BaRatin SPD discharge realizations for the calibration watersheds, a deterministic approach was developed to select pairs of realizations that cover most of the possible combinations of the range of simulated volumes at the two watersheds. According to this approach, the sets of realizations at the two watersheds were ranked according to the total runoff volume during the calibration period (WYs 1996–2015). The discharge realizations for each watershed with volumes closest to the 1-percent, 10-percent, 50-percent, 90-percent, and 99-percent nonexceedance probability quantiles of volume for that watershed (thus, five realizations for each watershed) were selected. Among the 25 possible pairwise combinations of these two sets of five realizations, 17 pairs of discharge series realizations were selected (fig. 4) to limit the computational expense and because we considered that the additional combinations would not offer much additional information. The naming convention QTp_TLp_L is used hereafter for designating the pairs, where Q is discharge, T indicates Tinley Creek, p_T is the Tinley Creek nonexceedance probability percentile, L indicates Long Run, and p_L is the Long Run nonexceedance probability percentile. For example, QT99L1 means the calibration discharge data comprising the realizations with the 99th nonexceedance probability percentile in volume from Tinley Creek and the 1st nonexceedance probability percentile in volume from Long Run. Once the calibration data pairs were selected, the procedures for calibrating the models are described in the "Two-Watershed Calibration for Regional Hydrological Simulation Program–FORTRAN Parameters" section.

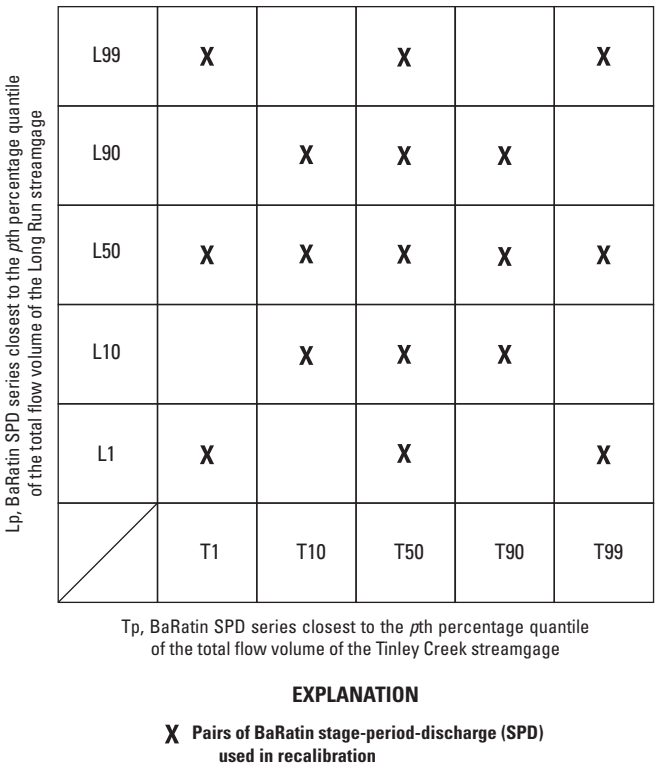

Figure 4. Pairing of BaRatin stage-period-discharge series realizations for the Tinley Creek near Palos Park, Illinois (U.S. Geological Survey station 05536500) and Long Run near Lemont, Illinois (U.S. Geological Survey station 05537500) streamgages for investigating Hydrological Simulation Program–FORTRAN parameter uncertainty arising from the uncertainty in published discharge records.

Uncertainty of Published Discharge

The multiperiod stage-discharge rating-curve-based BaRatin SPD method was applied to estimate the uncertainty of the observed discharge at the streamgages in the calibration watersheds at Tinley Creek and Long Run. The results of its application are described and discussed in this section, beginning with the rating curves and proceeding to the daily discharge time series.

Uncertainty of Stage-Discharge Rating Curves

Rating-curve realizations resulting from the application of the BaRatin SPD method as described in the "Estimating Uncertainty of Published Discharge Methods" section are illustrated, for selected periods, in figure 5. As expected, the 95-percent quantile CIs contain nearly all the stage-discharge measurement pairs, whereas the 1σ (one standard deviation) CIs contain most measurement pairs. Other than at lower flows where CIs widen, as does the range of measurements, the width of the CIs (in log-log space, so in proportional terms) does not vary appreciably with discharge. The 95-percent

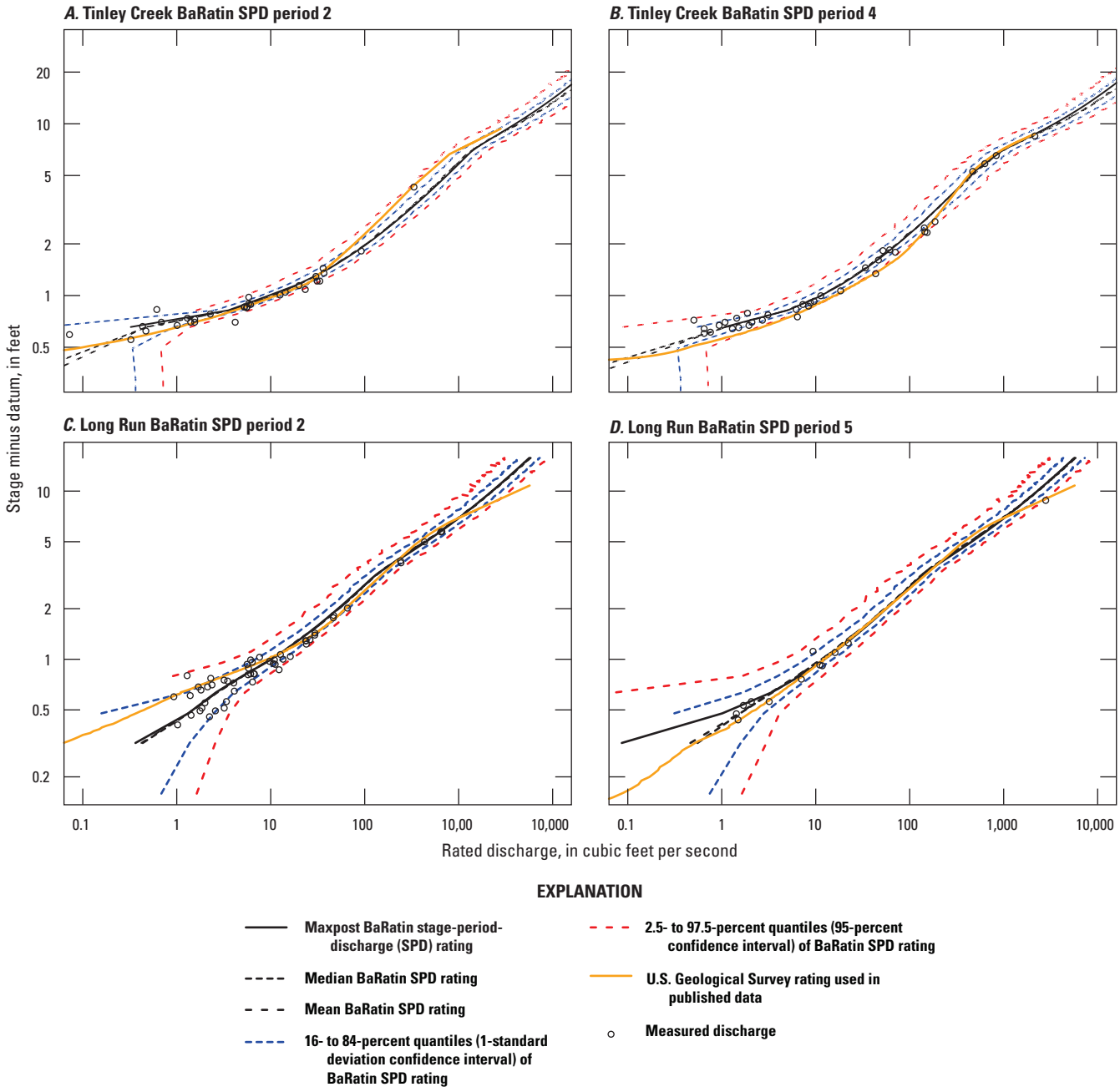

Figure 5. Selected rating curves obtained from application of the BaRatin stage-period-discharge (SPD) method in this study compared to the corresponding official U.S. Geological Survey (USGS) rating curves. *A*, BaRatin-SPD period 2 and *B*, BaRatin-SPD period 4 at Tinley Creek near Palos Park, Illinois, streamgage (USGS station 05536500). *C*, BaRatin-SPD period 2 and *D*, BaRatin-SPD period 4 at Long Run near Lemont, Illinois, streamgage (USGS station 05537500).

BaRatin SPD CIs contain the USGS rating curve, except at the upper tail at the Long Run streamgage, where, during SPD period 5 (table 2), the USGS rating curve was fitted through an observation that lies at the edge of the 95-percent BaRatin SPD interval. The USGS rating curve comes close to exceeding the 95-percent interval for the highest flow measurement during SPD period 2 (table 2) at the Tinley Creek streamgage, where again the USGS rating is drawn through a stage-discharge observation.

The highest measurements in period 4 at the Tinley Creek streamgage and in period 5 at the Long Run streamgage reflect a record rainstorm in this region of Illinois in July 1996, during which these peak-of-record discharges were measured at these streamgages. The Tinley Creek streamgage measurement was an indirect measurement based on culvert hydraulics and, therefore, was rated "poor" in accuracy, whereas the Long Run streamgage measurement was direct (instrument measured) and rated "fair" in accuracy.

The shapes of the rating curves depend on the activation stages where the control transitions from one to the next control. For the Tinley Creek streamgage, the weir (low-flow) to channel-control transition (activation stage $\kappa_2^{(k)}$) was estimated by this application of BaRatin SPD to be at 2.76 plus or minus (±)0.84 ft (mean±standard deviation) in period 2 and 6.72±1.47 ft in period 4. The change in mean and the substantial standard deviation indicate instability in the location of that transition, whereas the top-of-bank b_3 was assumed to be the same for all periods and was estimated to be at 8.52 ft with a smaller standard deviation (when compared to the weir to channel-control activation stage) of ±0.45 ft. In period 4 at the Tinley Creek streamgage, the four highest flow measurements agree well with the BaRatin SPD values in that the central tendency of the BaRatin SPD rating-curve distribution fits the measurements and closely fits the USGS rating curve. For the Long Run streamgage, during periods 2 and 5, these same quantities (periods 2 and 5 weir-to-channel flow activation stages $\kappa_2^{(m)}$ and top-of-bank stage b_3 throughout these periods) were estimated to be 0.92±0.08, 0.94±0.52, and 3.59±0.33 ft, respectively. Therefore, the low-flow to channel-flow transition was more stable at the Long Run streamgage than at the Tinley Creek streamgage. For both streamgages, the high-flow measurements from July 1996 are out-of-bank according to these parameters and according to the hydrographers' notes on the Long Run streamgage measurement (David Morgan, USGS, written commun., July 18, 1996).

The parameter distributions underlying the BaRatin SPD rating curves are illustrated for the Tinley Creek and Long Run streamgages in figs. 6 and 7, respectively. In most cases, the prior distributions are substantially wider than the posterior distributions, as expected. This result indicates the importance of the information gained from the measurements. By construction, the prior distributions of $b_1^{(k)}$ and $b_2^{(k)}$, the stage offsets for low-flow and channel-flow controls, widen with increasing period; however, their posterior distributions appear almost uniformly narrow at the scale of this plot, except for some widening of $b_2^{(k)}$ at the Tinley Creek streamgage. The mean value of at the Tinley Creek streamgage also increases from about 2.4 ft in periods 1 and 2 to about 3.5 ft for periods 3, 4, and 5 (tabulation not shown); the other mean values of $b_1^{(k)}$ and $b_2^{(k)}$ remain stable in all periods at both streamgages. Because only one shift in $b_1^{(k)}$ and $b_2^{(k)}$ is observed for either streamgage, despite substantial scatter in the low-flow measurements (fig. 3), additional analysis of the timing of rating-curve shifts would be beneficial.

The shifting of a_1 and a_2 of the posterior values relative to the centers of the prior value distributions for the Tinley Creek streamgage indicates that although available information for estimating the prior was uncertain, the measurements contained substantial and consistent information regarding these quantities. The structural error model parameters γ_1 and γ_2 also narrowed and shifted down between prior and posterior information and indicate relatively small errors in the posterior model as compared to the prior model.

Two exceptions to narrowing between prior and posterior distributions common to both streamgages are associated with the out-of-bank flows: b_3 (the top-of-bank stage) and a_3 ($K_s B_c \sqrt{S_0}$ for the out-of-bank flows), where measurements are few. Therefore, appreciable posterior uncertainty is expected to remain. The exponents c_1, c_2, and c_3 also failed to narrow in width because these exponents had narrow prior distributions, so additional narrowing is not expected.

Overall, the rating curves and associated parameters and their uncertainties obtained from the BaRatin SPD rating curves and their parameters appear to be meaningfully related to the prior information and measurements and to the USGS rating curves. As a result, using these parameters to obtain random realizations of predicted discharge at these streamgages is reasonable.

Uncertainty of Discharge Series from Uncertain Rating Curves

Following the methods described above, 500 daily discharge series realizations, 1 for each of the 500 posterior rating curves obtained from application of the BaRatin SPD method, were obtained for the Tinley Creek and Long Run streamgages. These rating curves were rescaled to agree in daily mean with the published USGS daily streamgage record (U.S. Geological Survey, 2020). The properties of these discharge series are briefly described in the following paragraphs.

The variability and mean of the daily discharge percentiles and mean over the simulation period (WYs 1996–2015) are presented and are compared to the published values in table 5. For a set of discharge percentiles and the mean and for the Tinley Creek and Long Run streamgages, the table gives the corresponding value from the published daily discharge and a set of statistics describing the distribution of the values of that percentile (or the mean) from the 500 simulated

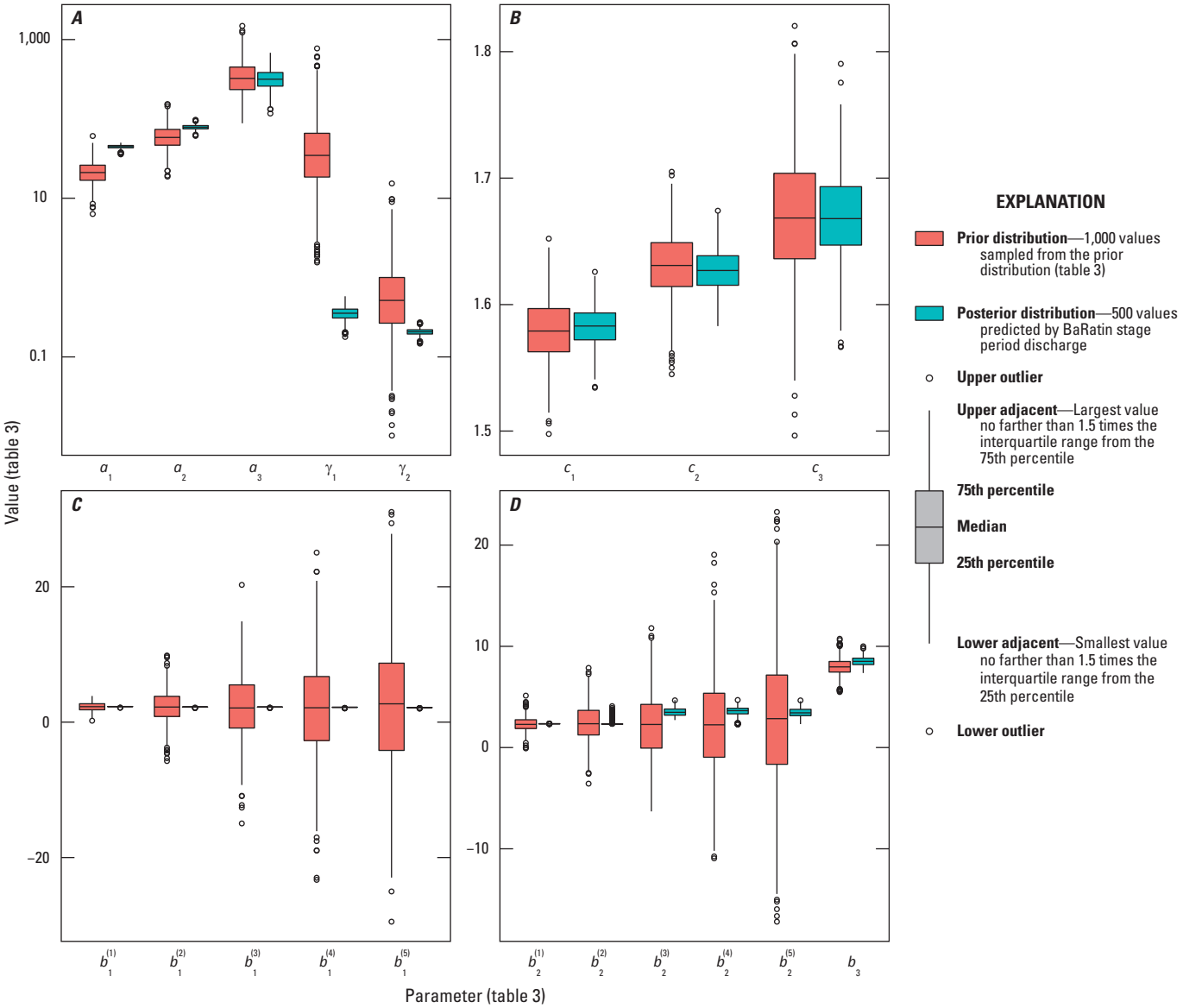

Figure 6. Distributions of the prior and posterior rating-curve parameters obtained with the BaRatin stage-period-discharge (SPD) method for this study at the Tinley Creek near Palos Park, Illinois, streamgage (U.S. Geological Survey station 05536500). *A*, stage factor (a_1, a_2, a_3) and structural error (γ_1,γ_2) parameters. *B*, stage exponent parameters. *C*, stage offset parameters, first control. *D*, stage offset parameters, second and third controls. See table 3 for definitions of parameters and analytical prior specifications.

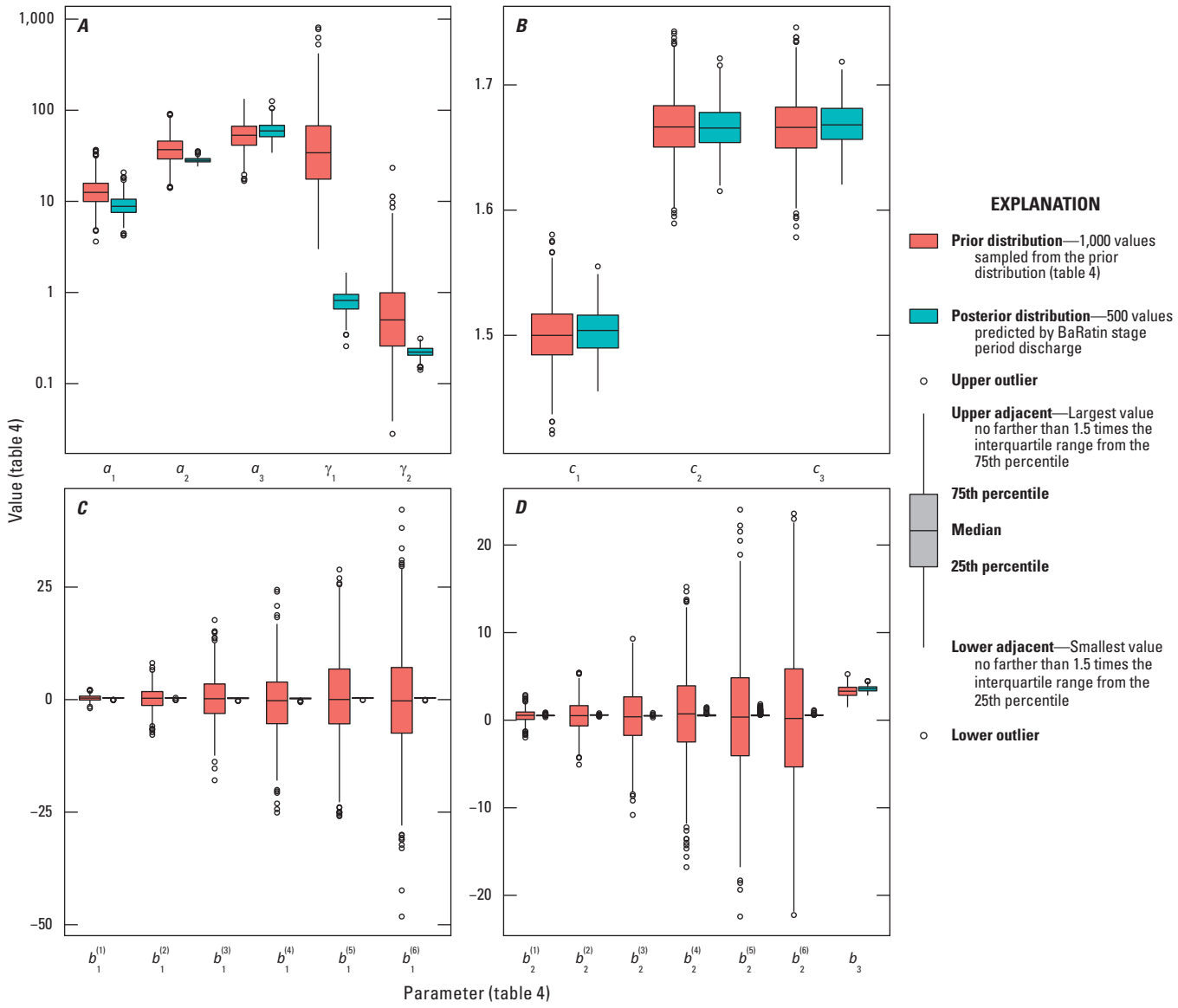

Figure 7. Prior and posterior rating-curve parameters obtained with the BaRatin stage-period-discharge (SPD) method for this study at the Long Run near Lemont, Illinois, streamgage (U.S. Geological Survey station 05537500). *A*, stage factor (a_1, a_2, a_3) and structural error (γ_1, γ_2) parameters. *B*, stage exponent parameters. *C*, stage offset parameters, first control. *D*, stage offset parameters, second and third controls. See table 4 for definitions of parameters and analytical prior specifications.

Table 5. Statistics describing variability of the mean and selected nonexceedance percentiles of the 500 BaRatin stage-period-discharge (SPD) daily discharge realizations from water years (WYs) 1996–2015 at the Tinley Creek and Long Run streamgages.

[Units are cubic feet per second, except coefficient of variation (CV), which is dimensionless; %, percent; Stdev, standard deviation]

Percentile	Tinley Creek						Long Run					
	Published	2.5% quantile	97.5% quantile	Mean	Stdev	CV	Published	2.5% quantile	97.5% quantile	Mean	Stdev	CV
1	0.18	0.13	0.20	0.168	0.019	0.111	0.78	0.47	0.93	0.70	0.12	0.165
5	0.47	0.39	0.55	0.461	0.044	0.095	1.35	1.02	1.54	1.29	0.13	0.100
10	0.87	0.72	0.99	0.850	0.069	0.081	1.90	1.57	2.13	1.85	0.14	0.076
25	1.90	1.70	2.06	1.88	0.089	0.047	3.60	3.24	4.02	3.62	0.20	0.054
50	4.57	4.26	4.80	4.55	0.14	0.031	8.60	7.99	9.27	8.60	0.32	0.037
75	12.0	11.2	12.6	12.0	0.36	0.030	20.9	19.6	22.2	20.9	0.66	0.032
90	31.8	29.0	34.0	31.8	1.26	0.040	53.1	49.3	58.0	53.4	2.13	0.040
95	60.0	55.7	63.8	59.8	2.00	0.033	92.7	84.9	100.7	92.1	4.07	0.044
99	187	174	206	190	8.06	0.043	274	244	312	277	17.6	0.063
Mean	15.0	14.2	15.8	15.0	0.42	0.028	24.6	23.0	26.2	24.6	0.80	0.032

discharge series. For example, the columns headed "2.5% quantile" contain the 2.5-percent quantiles of the 500 realizations of the quantities indicated in the leftmost column for example, the 2.5-percent quantile of the 1st percentile in the first row and the 2.5-percent quantile of the mean in the last row. Agreement between the means of the statistics and the published values are close because of rescaling. Variability among the 500 realizations as indicated by the coefficients of variation (CVs) is about 11 and 16.5 percent, respectively, for the 1-percent quantile at the Tinley Creek and Long Run streamgages, respectively, and decreases to about 3 percent for the 75-percent quantile before increasing again for higher quantiles. CVs are about 3 percent for the POR means. The 2.5–97.5-percent quantile intervals are approximately centered on the mean and about twice as wide as the 1σ interval, indicating approximately Gaussian variability.

Annual volumes are particularly relevant to the purpose of this report and are also of interest to show how the estimated uncertainty varies in time. Statistics of those volumes (presented as mean discharges) are shown in figure 8 and tables 6 and 7. First, prior to discussion of uncertainty, it may be observed that the annual mean discharges at the Tinley Creek and Long Run streamgages vary year to year by a factor of 2.4 at Tinley Creek and 3.1 at Long Run. These results presumably are because these are small watersheds and, thus, are particularly exposed to variations from year to year in the climatic forcing of runoff. Second, uncertainty varies by year, and not in a way that is strictly proportional to discharge (fig. 8; tables 6 and 7). Neither the dimensional (for example, standard deviation) and nondimensional (for example, CV) values are completely constant with time. The annual CVs at the Long Run streamgage scatter around 0.04 except for WYs 1997 and 1998, when the CVs are about 50 percent higher. The CV at the Tinley Creek streamgage is more complex than at the Long Run streamgage and appears to be almost uniformly distributed over its range from about 0.039 to 0.072. By construction, these variations result from the properties of the forcing stage time series and the rating-curve realizations, and because these averages are primarily controlled by the high flows. However, no substantial differences in rating-curve uncertainty by period were observed, as shown in figures 5, 6, and 7, nor in the rating curve periods not shown in figure 5. Investigation of the properties of the stage time series that may affect the discharge uncertainty variability is beyond the scope of this study; therefore, only the range and size of variation in the uncertainties of the annual volumes at the streamgages in the two calibration watersheds are noted.

Uncertainty of Discharge Series Realizations Selected for Recalibration

BaRatin SPD-generated discharge series realizations, where the total flow volumes correspond the closest to the 1-percent, 10-percent, 50-percent, 90-percent, and 99-percent quantiles of total volume during the HSPF simulation period, were selected for both streamgages at the calibration watersheds and paired (fig. 4). However, such ordering by magnitude in total volume does not necessarily carry over uniformly to other summary statistics, such as daily flow quantiles or annual volumes, or to the daily discharge time series for each day in the predicted record; instead, these statistics may cross one another. Because the HSPF calibration process used in this study depends on these and other observation groups, not simply on the total volume, these statistics for the selected series were examined.

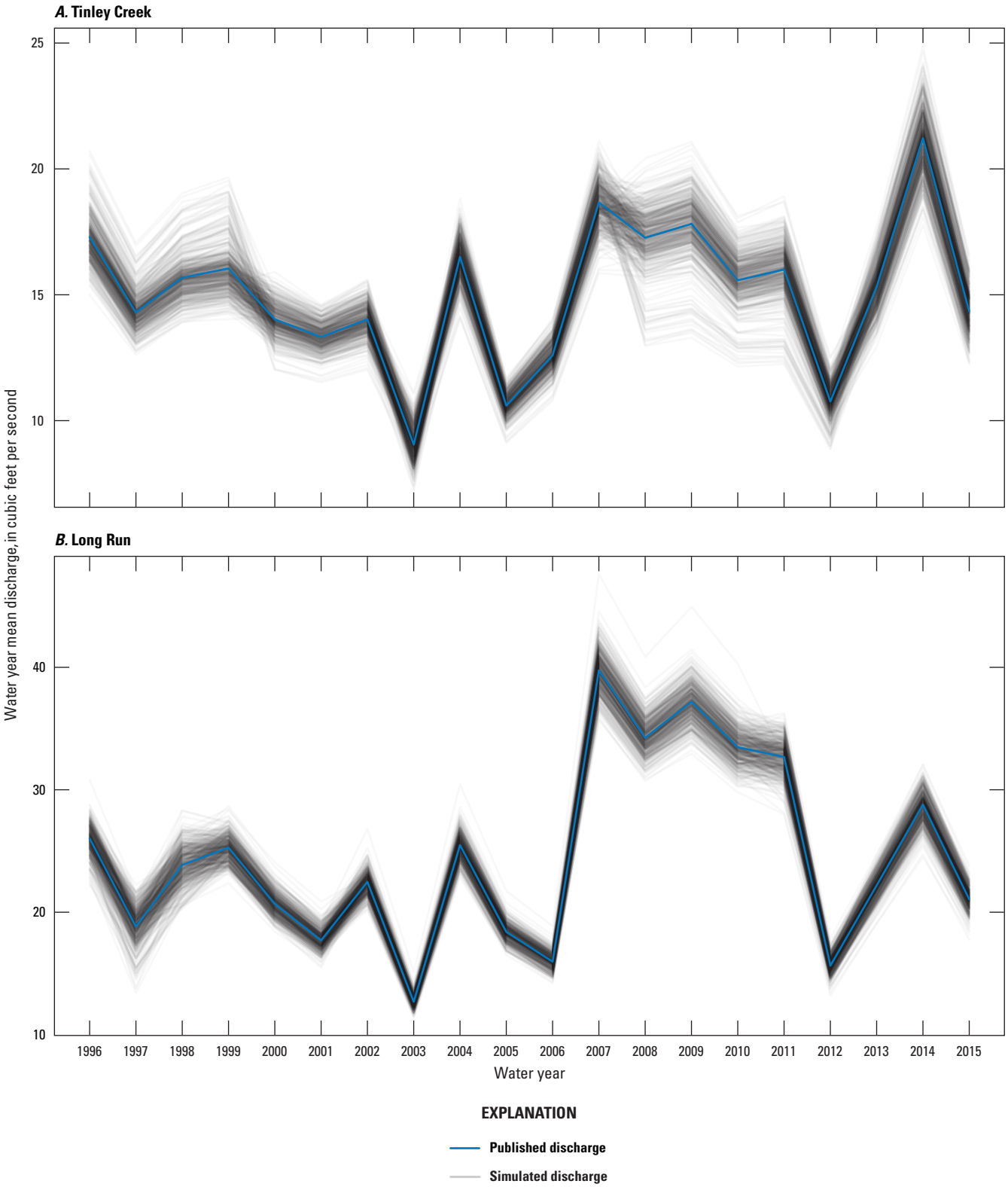

Figure 8. Water year mean discharge simulated with the BaRatin stage-period-discharge (SPD) method and published discharge series, water years 1996–2015. *A*, Tinley Creek near Palos Park, Illinois, streamgage (U.S. Geological Survey station 05536500). *B*, Long Run near Lemont, Illinois, streamgage (U.S. Geological Survey station 05537500).

Table 6. Water year mean statistics of discharge realizations generated by BaRatin stage-period-discharge (SPD) at Tinley Creek near Palos Park, Illinois, streamgage (U.S. Geological Survey station 05536500).

[Units are in cubic feet per second, except coefficient of variation (CV), which is dimensionless; SPD, stage-period-discharge; USGS, U.S. Geological Survey; Stdev, standard deviation; %, percent; NA, not applicable]

Water year	BaRatin SPD period	USGS rating-curve number(s)	Published	Mean	Stdev	CV	2.5% quantile	97.5% quantile	97.5%–2.5% quantile difference
1996	4–5	23–24	17.30	17.30	0.88	0.0507	15.77	19.56	3.79
1997	4	24	14.30	14.30	0.68	0.0477	13.08	15.99	2.90
1998	4	24	15.69	15.69	0.84	0.0536	14.31	18.01	3.70
1999	4	24	16.06	16.06	0.96	0.0595	14.58	18.78	4.19
2000	3–4	24–25	14.03	14.03	0.61	0.0435	12.78	15.26	2.48
2001	3	25	13.33	13.33	0.52	0.0389	12.30	14.36	2.06
2002	3	25	14.04	14.04	0.59	0.0420	12.85	15.18	2.32
2003	3	25	9.05	9.05	0.65	0.0717	7.86	10.35	2.49
2004	3	25	16.52	16.52	0.74	0.0449	15.07	17.97	2.90
2005	3	25	10.60	10.60	0.42	0.0398	9.73	11.41	1.69
2006	3	25	12.60	12.60	0.53	0.0418	11.49	13.61	2.12
2007	3	25	18.68	18.68	0.83	0.0444	17.09	20.32	3.23
2008	2	26	17.28	17.28	1.14	0.0662	13.91	19.09	5.18
2009	2	26	17.83	17.83	1.20	0.0675	14.34	19.82	5.47
2010	2	26	15.59	15.59	0.95	0.0612	12.93	17.27	4.34
2011	2	26	16.01	16.01	1.06	0.0659	12.97	17.88	4.91
2012	2	26	10.77	10.77	0.59	0.0547	9.38	11.93	2.54
2013	1	27	15.33	15.33	0.81	0.0526	13.77	17.02	3.25
2014	1	27	21.23	21.23	1.17	0.0551	18.98	23.59	4.62
2015	1	27	14.34	14.34	0.80	0.0556	12.76	15.93	3.17
Mean	NA	NA	15.03	15.03	0.80	0.0529	13.30	16.67	3.37

Table 7. Water year mean statistics of discharge realizations generated by BaRatin stage-period-discharge (SPD) at Long Run near Lemont, Illinois, streamgage (U.S. Geological Survey station 05537500).

[Units are in cubic feet per second, except coefficient of variation (CV), which is dimensionless; SPD, stage-period-discharge; USGS, U.S. Geological Survey; Stdev, standard deviation; %, percent; NA, not applicable]

Water year	BaRatin SPD period	USGS rating-curve number(s)	Published	Mean	Stdev	CV	2.5% quantile	97.5% quantile	97.5%–2.5% quantile difference
1996	5–6	27–28	26.03	26.03	1.16	0.0444	23.49	28.10	4.60
1997	5	28	18.83	18.83	1.42	0.0754	15.30	21.32	6.01
1998	4–5	28–29	23.87	23.87	1.45	0.0606	20.71	26.68	5.97
1999	3–4	29–30	25.29	25.29	0.98	0.0386	23.58	27.31	3.73
2000	3	30	20.69	20.69	0.86	0.0417	19.14	22.52	3.38
2001	3	30	17.71	17.71	0.76	0.0427	16.33	19.30	2.97
2002	2	31	22.49	22.49	0.86	0.0381	20.78	24.24	3.46
2003	2	31	12.69	12.69	0.48	0.0382	11.78	13.66	1.88
2004	2	31	25.49	25.49	0.98	0.0384	23.55	27.52	3.97
2005	2	31	18.40	18.40	0.70	0.0379	16.99	19.88	2.89
2006	2	31	15.97	15.97	0.64	0.0400	14.74	17.22	2.47
2007	2	31	39.78	39.78	1.54	0.0387	36.75	42.94	6.19
2008	2	31	34.26	34.26	1.31	0.0381	31.67	36.88	5.21
2009	2	31	37.21	37.21	1.49	0.0400	34.36	40.14	5.78
2010	2	31	33.50	33.50	1.33	0.0396	30.94	36.22	5.28
2011	1	32	32.70	32.70	1.31	0.0401	30.06	35.20	5.14
2012	1	32	15.64	15.64	0.67	0.0431	14.35	16.90	2.55
2013	1	32	22.20	22.20	0.91	0.0408	20.55	23.90	3.35
2014	1	32	28.80	28.80	1.20	0.0416	26.42	31.06	4.64
2015	1	32	21.06	21.06	0.87	0.0415	19.41	22.68	3.26
Mean	NA	NA	24.63	24.63	1.05	0.0430	22.55	26.68	4.14

Table 8. Comparison of nonexceedance probability percentiles and means of selected BaRatin stage-period-discharge (SPD) daily discharge realizations representing the 1-, 10-, 50-, 90-, and 99-percent nonexceedance probability quantiles of total volume at the Tinley Creek near Palos Park, Illinois, and Long Run near Lemont, Illinois, streamgages (U.S. Geological Survey stations 05536500 and 05537500, respectively) for the period from water year (WY) 1996 to WY 2015.

[Units are in cubic feet per second; %, percent]

Percentile	Tinley Creek						Long Run					
	Published	1% quantile	10% quantile	50% quantile	90% quantile	99% quantile	Published	1% quantile	10% quantile	50% quantile	90% quantile	99% quantile
1	0.18	0.16	0.16	0.20	0.14	0.19	0.78	0.76	0.59	0.68	0.66	0.53
5	0.47	0.45	0.43	0.50	0.40	0.53	1.35	1.32	1.13	1.31	1.25	1.06
10	0.87	0.87	0.80	0.90	0.80	0.91	1.90	1.86	1.69	1.87	1.81	1.66
25	1.90	1.87	1.82	1.93	1.84	1.96	3.60	3.56	3.33	3.62	3.69	3.45
50	4.57	4.54	4.43	4.69	4.60	4.79	8.60	8.42	8.30	8.73	8.23	8.60
75	12.0	11.6	12.0	12.1	12.1	12.7	20.9	20.0	20.4	20.8	21.3	21.8
90	31.8	29.5	32.7	31.9	33.4	34.6	53.1	50.3	51.2	52.1	56.9	57.1
95	60.0	56.0	61.7	59.8	62.3	64.0	92.7	84.9	88.6	89.2	97.9	99.2
99	187.0	173.4	190.6	188.9	193.1	196.9	274.0	244.9	263.2	279.4	297.6	308.9
Mean	15.0	14.1	15.3	15.0	15.6	16.0	24.6	22.9	23.6	24.6	25.6	26.5

The POR percentiles and means of the selected realizations are given in table 8. In this table, values of the 1st to 99th percentiles and the means of daily discharge are presented. The values for the published daily discharge are presented in the columns labeled "Published," and in the remaining columns the values for the selected realizations corresponding to the five quantiles of total volume considered are presented. For example, the columns headed "1% quantile" give the nonexceedance probability percentiles and mean of the realization closest to the 1 percent quantile of total volume for the Tinley Creek and Long run streamgages. These statistics show that the realizations are ordered as expected for higher flow percentiles (from 75th to 99th), but the orderings are unpredictable for the lower-flow percentiles (less than 75th). For example, the lowest value for the 1st, 5th, and 10th percentiles is the 90-percent realization for the Tinley Creek streamgage and the 99-percent realization for the Long Run streamgage, whereas the highest is the 99-percent realization for the Tinley Creek streamgage and the 1- or 50-percent realization for the Long Run streamgage.

Fewer such crossings are observed in the annual flow volumes presented in tables 9 and 10, but they are present. For example, at the Tinley Creek streamgage (table 9), the 90-percent quantile volume realization has the lowest total annual volume among all realizations from WY 1996 to 1999 (encompassing BaRatin SPD periods 4 and 5), and the highest from WY 2013 to 2015. That these two periods encompass BaRatin SPD periods 5, 4, and 1 indicates, not unexpectedly, that such crossings depend on the variation of rating-curve parameters among periods. The annual volumes for the Long Run streamgage (table 10) maintain their ordering more often than the Tinley Creek streamgage, but not always. For example, the 90-percent quantile volume realization has a relatively low volume at the Long Run streamgage when compared to the Tinley Creek streamgage from WY 1996 to 1998, which covers all or parts of BaRatin SPD periods 4, 5, and 6 at this streamgage.

Table 9. Water year annual mean discharges at selected BaRatin stage-period-discharge (SPD) realizations representing the 1-, 10-, 50-, 90-, and 99-percent nonexceedance probability quantiles of total volume at the Tinley Creek near Palos Park, Illinois, streamgage (U.S. Geological Survey station 05536500).

[Units are cubic feet per second; SPD, stage-period-discharge; USGS, U.S. Geological Survey; %, percent; NA, not applicable]

Water year	BaRatin SPD period	USGS rating curve(s)	Published	1% quantile	10% quantile	50% quantile	90% quantile	99% quantile
1996	4–5	23–24	17.30	16.62	17.18	17.48	16.39	18.45
1997	4	24	14.30	13.73	14.20	14.63	13.53	15.30
1998	4	24	15.69	15.03	15.64	16.02	14.99	16.43
1999	4	24	16.06	15.41	16.09	16.18	15.27	16.87
2000	3–4	24–25	14.03	13.05	14.42	13.97	14.19	14.45
2001	3	25	13.33	12.61	13.74	13.28	13.49	13.70
2002	3	25	14.04	13.05	14.51	13.91	14.21	14.42
2003	3	25	9.05	9.20	8.78	9.07	9.38	9.31
2004	3	25	16.52	15.42	17.14	16.50	16.73	16.87
2005	3	25	10.60	9.97	10.96	10.55	10.79	10.94
2006	3	25	12.60	11.90	13.05	12.55	12.77	12.97
2007	3	25	18.68	17.43	19.24	18.47	18.95	19.17
2008	2	26	17.28	16.20	17.74	17.21	18.74	19.33
2009	2	26	17.83	16.54	18.30	17.49	19.27	20.06
2010	2	26	15.59	14.74	15.77	15.42	16.62	17.34
2011	2	26	16.01	15.02	16.40	15.73	17.07	18.10
2012	2	26	10.77	10.35	10.79	10.64	11.26	12.08
2013	1	27	15.33	13.70	15.90	15.71	17.29	16.22
2014	1	27	21.23	18.89	22.07	21.32	24.08	22.62
2015	1	27	14.34	12.94	14.76	14.82	16.25	14.90
Mean	NA	NA	15.03	14.09	15.33	15.05	15.56	15.98

Table 10. Water year annual mean discharges at selected BaRatin stage-period-discharge (SPD) realizations representing the 1-, 10-, 50-, 90-, and 99-percent nonexceedance probability quantiles of total volume at the Long Run near Lemont, Illinois, streamgage (U.S. Geological Survey station 05537500).

[Units are cubic feet per second; SPD, stage-period-discharge; USGS, U.S. Geological Survey; %, percent; NA, not applicable]

Water year	BaRatin SPD period	USGS rating curve(s)	Published	1% quantile	10% quantile	50% quantile	90% quantile	99% quantile
1996	5–6	27–28	26.03	24.38	25.36	26.56	25.81	28.41
1997	5	28	18.83	17.93	18.73	18.83	16.87	21.58
1998	4–5	28–29	23.87	22.01	23.95	23.60	22.86	24.65
1999	3–4	29–30	25.29	23.58	24.61	25.17	26.56	26.27
2000	3	30	20.69	19.44	19.84	20.33	21.52	21.65
2001	3	30	17.71	16.78	17.35	17.56	18.54	18.22
2002	2	31	22.49	20.76	21.60	22.55	24.09	24.66
2003	2	31	12.69	11.76	12.07	12.81	13.42	13.89
2004	2	31	25.49	23.45	24.28	25.62	27.35	27.91
2005	2	31	18.40	17.01	17.64	18.34	19.60	19.92
2006	2	31	15.97	14.80	15.20	15.77	16.66	16.97
2007	2	31	39.78	36.33	37.70	39.46	42.68	43.23
2008	2	31	34.26	31.49	32.55	34.59	36.58	37.39
2009	2	31	37.21	33.89	35.47	37.45	40.15	40.89
2010	2	31	33.50	30.48	31.94	33.40	36.04	36.85
2011	1	32	32.70	30.54	31.10	32.72	34.03	34.98
2012	1	32	15.64	15.07	14.70	15.77	15.77	15.88
2013	1	32	22.20	20.91	21.09	22.68	23.04	23.41
2014	1	32	28.80	26.85	27.18	28.60	30.02	30.63
2015	1	32	21.06	20.12	19.96	21.08	21.37	21.81
Mean	NA	NA	24.63	22.88	23.62	24.65	25.65	26.46

Parameter Uncertainty

The two sets of HSPF simulations in this study are based on two corresponding HSPF parameter sets, one that characterizes the effects of the uncertainty of published discharge records and the other that characterizes the uncertainty of model parameters. These parameter sets are obtained as described in the sections "Model Recalibration with Uncertain Published Discharge Time Series" and "Estimating Uncertainty of Base-Model Parameters," respectively. The first set was obtained by recalibration to 17 discharge series pairs that characterize the uncertainty of the published discharge record. The second was obtained by Monte Carlo simulation of parameter sets based on the estimated base-model parameter covariance matrix. In this section, the uncertainty of these two sets of parameters are discussed in terms of statistics describing their distributions.

Uncertainty of Parameters from Recalibrations with Uncertain Published Discharges

The resulting parameters after recalibration with the 17 selected pairs of discharge realizations from the discharge uncertainty analysis showed changes in grassland and forest parameters, but not all adjustable parameters responded substantially to the differences in the calibration discharge series (fig. 9, table 11). Two types of changes are shown by the parameters. First, relative to the base-model parameters, the recalibrated values of LSUR, IRC, and KVARY (see table 1.1 for definitions) for grassland switched with those for forest as compared to the base model (fig. 9). Second, the parameter values from the 17 recalibrated parameter sets have a range of variations. Among those parameter sets, the FOREST, NSUR, LSUR, SLSUR, INTFW, IRC, KVARY, AGWRC, and DEEPFR parameters (see table 1.1 for definitions) in both land-cover types vary little. Quantifying the variation using the CV of the values shown in figure 9, the parameters INFILT, LZSN, UZSN, AGWETP, and LZETP (see table 1.1 for definitions) varied mostly only in forest, with CV greater than or equal to 0.01. Both PETMAX and PETMIN (see table 1.1 for definitions) had CV greater than 0.01 in grassland, and PETMIN also had CV greater than 0.01 in forest (table 11).

Having been recalibrated with the same climatic forcing data but different calibration discharges, the parameters that changed are primarily those pertinent to model evapotranspiration (ET) and losses to deep recharge (that is, processes to maintain water balance). PETMAX and PETMIN affect ET when snow processes are being simulated. With snow parameters not being calibrated, the large changes in PETMAX and PETMIN possibly reflect their taking surrogate roles for other parameters that also simulate ET but whose values are fixed throughout the year. FOREST is another parameter that affects ET when there is snowpack. In contrast to PETMAX and PETMIN, the FOREST values are site-specific data, and, therefore, there is not much information in the discharge data for calibrating its values. Snowpack has a larger effect on winter flows, but ET is a primary component of the water balance at all times. The PETMAX and PETMIN parameters showed large variations but their effects on predicted annual discharge may be small. Other parameters affecting ET volumes include LZETP, BASETP, AGWETP, and those parameters related to soil moisture storages (for example, UZSN, LZSN, and CEPSC). These parameters also changed after recalibration, except for BASETP, which was not among the adjustable parameters considered in this study.

Among the parameters that did not vary appreciably in the recalibration process, INTFW, IRC, and AGWRC are, however, parameters that are commonly selected for calibration. These parameters, along with NSUR, LSUR, SLSUR, and KVARY varied by only a small amount (see CV values in table 11) among the 17 recalibration sets. NSUR, LSUR, and SLSUR primarily affect the overland flows, INTFW and IRC simulate interflows pertinent to drainage mechanisms affecting when the flows return to the main channel, and KVARY and AGWRC mainly affect base flows that relate to groundwater-recession flow patterns. Both DEEPFR and KVARY parameters affect the simulation of groundwater processes; in particular, the fraction of the HSPF groundwater inflow that is lost to deep aquifer (DEEPFR) and the degree of nonlinearity of the base-flow recession (KVARY), respectively (table 1.1). DEEPFR and KVARY are usually the last parameters to be adjusted in manual calibrations because the processes they describe are abstract and not reflected in observed data (U.S. Environmental Protection Agency, 2000). The small variation of these seven parameters among the 17 recalibrated parameter sets may have resulted from their dependencies on the climate inputs, which are invariant in this study; from not being able to obtain information from observation groups for their calibration; from the fractions of individual observation groups in the final objective function not being approximately equal; or from some of these parameters being constrained not to vary seasonally.

Uncertainty of Base-Model Parameters

HSPF base-model parameter uncertainties, expressed by physical lower and upper bounds (determined based on expert knowledge) before the calibration, by their standard deviations obtained from the posterior covariance matrix estimated with the PEST PREDUNC7 utility, and by the standard deviations of the parameters sampled using the RANDPAR utility from the estimated posterior covariance matrix are presented in table 12. The CVs (that is, the ratio of the standard deviation to the mean) of the post-calibration parameters vary from about 0.1 to about 1, much larger than those in recalibrated parameters (table 11). Many of the smaller CVs are associated with the ratio parameters, such as the INFILTR, and with the PETMIN and PETMAX parameters, where units are degrees Fahrenheit (°F), such that a CV of 0.1 or 0.2 indicates

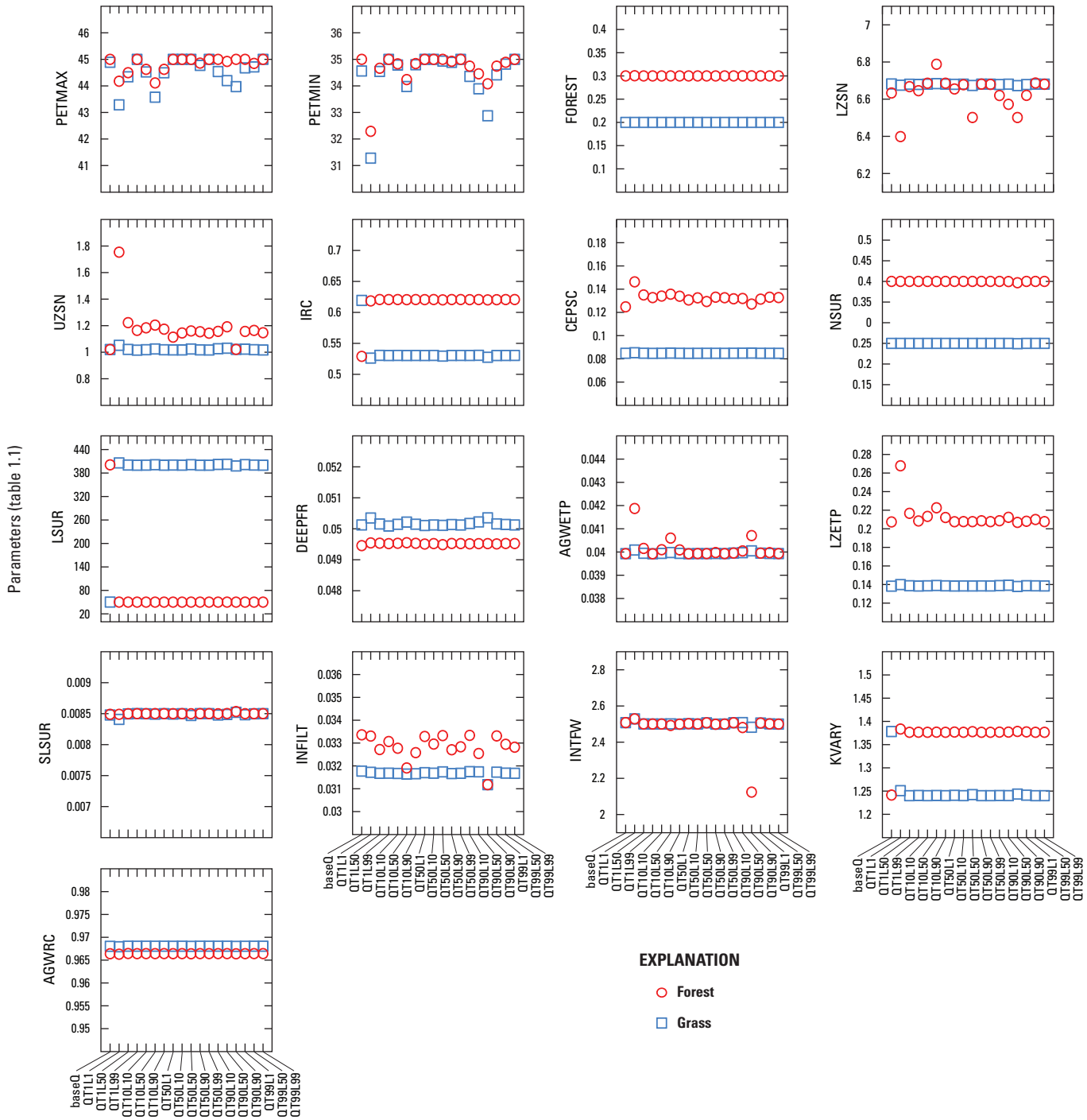

Figure 9. Hydrological Simulation Program–FORTRAN grassland and forest parameter values resulting from calibration using the published discharge records (baseQ) and from recalibration with the 17 pairs of discharge series realizations characterizing the uncertainty of the published discharge record. See table 1.1 for parameter definitions and figure 4 for definitions of recalibration pairs.

Table 11. Hydrological Simulation Program–FORTRAN parameter values resulting from calibration with published discharge records (base model) and statistics of 17 parameter sets from recalibration with discharge realizations characterizing the uncertainty of the published discharge records at the Tinley Creek near Palos Park, Illinois, and Long Run near Lemont, Illinois, streamgages (U.S. Geological Survey stations 05536500 and 05537500, respectively).

[Base model refers to calibration with published discharge records; CV, coefficient of variation]

Parameter name[1]	Grassland				Forest			
	Base model	Mean, recalibrations	CV, recalibrations	Range, recalibrations	Base model	Mean, recalibrations	CV, recalibrations	Range, recalibrations
FOREST	0.2	0.2	1.34E-06	1.99999–0.2	0.30003	0.30005	0.00034	0.299999–0.300419
LZSN	6.681	6.679	4.26E-04	6.673–6.684	6.604	6.632	0.014	6.398–6.788
INFILT	0.032	0.032	0.004	0.031–0.032	0.033	0.033	0.017	0.031–0.033
LSUR	50.2	400.8	0.004	398.1–405.9	400.8	50	0	50–50.1
SLSUR	0.0085	0.0085	0.003	0.00841–0.00852	0.0085	0.0085	0.001	0.00849–0.00853
KVARY	1.378	1.241	0.002	1.24–1.251	1.241	1.377	0.001	1.377–1.384
AGWRC	0.968	0.968	3.07E-05	0.9679–0.9681	0.966	0.966	3.36E-05	0.9663–0.9664
PETMAX	44.9	44.5	0.012	43.286–45	45	44.8	0.007	44.111–45
PETMIN	34.7	34.4	0.028	31.28–35	35	34.6	0.019	32.29–35
DEEPFR	0.05013	0.05017	0.00153	0.05009–0.05035	0.04947	0.04952	0.00029	0.04949–0.04955
AGWETP	0.0399	0.04	0.0011	0.0399–0.0401	0.0399	0.0402	0.0123	0.0399–0.0419
CEPSC	0.08479	0.08484	0.00151	0.0847–0.0853	0.12679	0.13318	0.02955	0.1272–0.1463
UZSN	1.019	1.022	0.008	1.015–1.052	1.019	1.191	0.127	1.021–1.754
NSUR	0.25	0.25	0.0007	0.2492–0.25	0.4	0.3998	0.0019	0.3968–0.4
INTFW	2.506	2.503	0.003	2.482–2.529	2.506	2.479	0.037	2.124–2.529
IRC	0.62	0.53	0.002	0.526–0.531	0.529	0.62	0.001	0.619–0.621
LZETP	0.138	0.139	0.003	0.1379–0.1399	0.208	0.214	0.068	0.2069–0.2678

[1]See table 1.1 for parameter definitions.

Table 12. Prior and posterior statistics of adjustable parameters of the base Hydrological Simulation Program–FORTRAN model and those of the 1,000 simulated parameter sets.

[stdev, standard deviation; CV, coefficient of variation]

Parameters[1]	Information for base-model parameters before calibration			Posterior parameter uncertainty estimation using PREDUNC7			Random parameter set: RANDPAR samples (1,000 simulations)		
	Initial value[2]	Lower bound[2]	Upper bound[2]	Input (prior) stdev[3,4]	Output (posterior) stdev[3]	Output CV[2]	Base model parameters (simulation input mean)[2]	Mean of simulated values[3]	Stdev of simulated values[3]
FOREST-grass	0.20	0.05	0.20	0.151	0.145	0.580	0.200	−0.758	0.084
FOREST-forest	0.30	0.20	0.50	0.099	0.098	0.465	0.300	−0.524	0.094
LZSN-grass	6.681	2	15	0.219	0.044	0.304	6.681	0.823	0.045
LZSNR-forest	1.00	0.7	1.3	0.067	0.057	0.347	0.988	−0.004	0.057
INFILT-grass	0.032	0.001	0.1	0.500	0.033	0.262	0.032	−1.499	0.033
INFILTR-forest	1.036	1	1.05	0.005	0.005	0.103	1.050	0.019	0.003
LSUR-forest	400	50	500	0.250	0.116	0.511	401	2.588	0.097
LSUR-grass	50	50	500	0.250	0.090	0.443	50	1.736	0.053
SLSUR-forest	0.0085	0.001	0.02	0.325	0.141	0.570	0.008	−2.070	0.133
SLSUR-grass	0.0085	0.0051	0.02	0.148	0.158	0.610	0.008	−2.079	0.151
KVARY-forest	1.240	0	3	1.369	0.266	0.839	1.241	0.076	0.248
KVARY-grass	1.377	0	3	1.369	0.187	0.674	1.378	0.135	0.184
AGWRCTR-grass	30.30	11	50	0.164	0.065	0.374	30.279	1.481	0.067
AGWRCTRR-forest	0.95	0.95	1.1	0.016	0.016	0.180	0.950	−0.016	0.009
PETMAX-grass	45	35	45	0.027	0.014	0.170	44.9	1.647	0.009
PETMAX-forest	45	35	45	0.027	0.022	0.213	45.0	1.645	0.013
PETMIN-grass	35	30	35	0.017	0.010	0.144	34.7	1.537	0.007
PETMIN-forest	35	30	35	0.017	0.013	0.162	35.0	1.539	0.008
DEEPFR-forest	0.0495	0.001	0.2	0.575	0.376	1.058	0.049	−1.299	0.357
DEEPFR-grass	0.0501	0.001	0.2	0.575	0.263	0.831	0.050	−1.295	0.251
AGWETP-grass	0.04	0.0001	0.1	0.750	0.044	0.303	0.040	−1.399	0.043
AGWETPR-forest	1.00	1	1.05	0.005	0.005	0.103	1.000	0.002	0.003
CEPSC-grass	0.085	0.03	0.2	0.206	0.052	0.330	0.085	−1.074	0.050
CEPSCR-forest	1.567	1	2.5	0.099	0.084	0.427	1.495	0.179	0.081
UZSN-grass	1.017	0.05	1.2	0.345	0.013	0.165	1.019	0.008	0.013
UZSNR-forest	1.130	1	2	0.075	0.033	0.260	1.000	0.013	0.019
NSUR-forest	0.25	0.15	0.25	0.055	0.049	0.322	0.400	−0.417	0.028
NSUR-grass	0.40	0.25	0.4	0.051	0.045	0.306	0.250	−0.621	0.027
INTFW-grass	2.50	1	5	0.175	0.026	0.231	2.506	0.400	0.026
INTFWR-forest	1.00	0.75	1	0.031	0.028	0.240	1.000	−0.010	0.015
IRCTR-forest	1.130	1	2.33	0.092	0.064	0.368	1.124	0.055	0.050
IRCTR-grass	1.637	1	2.33	0.092	0.052	0.332	1.629	0.216	0.052
LZETP-grass	0.138	0.1	0.9	0.239	0.027	0.238	0.138	−0.860	0.027
LZETPR-forest	1.5	1.5	3.6	0.095	0.050	0.326	1.500	0.196	0.030

[1]A suffix "R" to a parameter in forest land cover type means the parameter is a ratio of the grassland value. See table 1.1 for parameter definitions.

[2]Computed from untransformed values.

[3]Computed from $\log_{10}$-transformed values.

[4]The PREDUNC7 input stdev was computed as $\log_{10}$(Upper bound)/(Lower bound))/4.

variations of 4 to 8 °F. The parameters with the largest CVs are the DEEPFR and the KVARY parameters for the two types of land covers. In this study, focusing as it does on the long-term mean discharge, the effect of uncertainty in the loss to deep recharge, which is parameterized by DEEPFR, is especially important.

In comparing the parameter standard deviations before and after calibration as calculated by PREDUNC7 (table 12), it can be observed that the standard deviations are usually reduced, although to varying degrees. These reductions indicate the effects of the information provided by the observations in the calibration process. The KVARY parameters show the largest reductions but remain among the most uncertain model parameters. Large reductions are also evident for parameters LZSN, INFILT, and UZSN. These parameters govern surface runoff generation and infiltration in upper and lower soil zones (see table 1.1).

The standard deviations of the sampled random parameters provide a check on the accuracy of the sampling process. The values are about the same or up to about 30 percent smaller than the standard deviations of the posterior parameters. The reduction in the sampled standard deviations likely results because of trimming as the sampling is performed on the assumption of a multivariate Gaussian distribution applied to the $\log_{10}$-transformed parameter values. After back-transforming to real space, the values outside the parameter bounds are trimmed to the bounds.

As described previously, the underlying parameter uncertainties as estimated using PREDUNC7 are characterized by covariance matrices from which the standard deviations that are presented were abstracted. The off-diagonal terms of these matrices (not shown) include some significant correlations that are of interest for understanding the model-parameter interactions. A detailed analysis of these is beyond the scope of this study, but to summarize, the most common significant correlations are those on the order of −0.5, which are common but not universal among forest-grassland parameter pairs such as KVARY-forest and KVARY-grass. These correlations indicate that when the value for one land-use type goes up, the other goes down, indicating difficulty in identifying which land use to associate with some aspect of the calibration data, which is expected given that both calibration watersheds are mixtures of land-use types. A few parameters have significant correlations of at least 0.45 in absolute value beyond their land-use type pairing; most prominent among these is INFILT-grass, the primary infiltration parameter, which is negatively correlated with INTFW-grass and UZSN-grass, which parameterize interflow and the upper zone soil storage, which are processes closely connected with infiltration. INFILT-grass is also positively correlated with DEEPFR-grass and AGWETP-grass, which are the fraction of water lost to inactive groundwater and the fraction of land segment subject to direct evaporation from groundwater storage. All these processes are connected through the soil-water balance, so these correlations would be expected; however, they are not typically present for other parameters. For example, the parameter corresponding forest infiltration parameter INFILTR-forest does not have such correlations.

Normalized Variability Index for Uncertainty of Simulated Discharge Statistics

To investigate how the simulated discharge could vary because of possible variations in published discharge records through the use of LMDA HSPF modeling, two sources of uncertainty are considered: (1) the effect of the uncertainty of the published discharge that is characterized by 17 HSPF simulations each based on a parameter set obtained from recalibration to a random realization that characterizes the uncertainty of the published discharge, and (2) uncertainty of the base-model calibration parameter set characterized by 1,000 simulations from parameters sampled from its estimated covariance matrix. For a flow statistic X derived from either set of simulations, for example the period-of-study (POS) mean discharge, X has a central tendency (for example, the mean or median), and a range of variability. The range of variability indicates the uncertainty of X of the error source being modeled by the corresponding set of simulations, whereas the distance from its central tendency to the corresponding observed value indicates the simulation error.

The central tendency and the range of variability of statistics of the simulations can be expressed dimensionlessly by considering them as fractions of their corresponding observed values. Using the median as the measure of the central tendency, these dimensionless statistics are X_{med}/X_{obs}, where X_{med} is the median of the values of the statistic X from the simulations and X_{obs} is the observed value, and X_R/X_{obs}, where X_R is the range of the simulation values of the statistic X. With the choice of the median as the measure of the central tendency, the central tendency error is $X_{med} - X_{obs}$. The absolute magnitude of the central tendency error relative to the X_{obs} is called the relative median error in this report and has the following form for $X_{med} \geq 0$ and $X_{obs} > 0$:

$$|X_{med} - X_{obs}|/X_{obs} = |1 - X_{med}/X_{obs}|. \quad (9)$$

where

X_{med} is the median of the vector of simulation flow statistics; and

X_{obs} is the observed value of the statistic.

The conditions $X_{med} \geq 0$ and $X_{obs} > 0$ are satisfied for all flow statistic values considered in this report. Notice that when the median error is small, X_{med}/X_{obs} is close to 1 and the relative median error is close to zero.

To assess how large the simulation range, which characterizes magnitude of the uncertainty of associated error source, is relative to the error in the simulation, the two error indices, X_R/X_{obs} and $|X_{med} - X_{obs}|/X_{obs}$, can be combined by taking their ratio. Half of this ratio is defined for use in this study as the normalized variability index $V_N(X)$, which is given by

$$V_N(X) = (X_R / 2) / \left| (X_{med} - X_{obs}) \right|, \quad (10)$$

where

X_R is the range of simulated values of X;
X_{med} is the median of the vector of simulation flow statistics; and
X_{obs} is the observed value of the statistic.

The range X_R is calculated as the maximum less the minimum in the case of the recalibrations to the realizations of uncertain published discharge series at the streamgages in the calibration watersheds (because there are only 17 values) and as the 97.5-percentile less the 2.5-percentile (the 95-percent confidence interval) in the case of the simulations parameterized by sampling from the base model parameter covariance matrix. A graphical depiction of the quantities used to compute $V_N(X)$ is given in figure 10*A*,*B* for the case of the POS mean discharge.

Typically, simulation-to-observation errors have multiple sources, and if the range of variability of X is small relative to the simulation-to-observation error, that is, if $V_N(X)$ is much less than 1, it is clear that the source of uncertainty of the simulations can only explain a small fraction of the error; other sources must contribute the remainder of the error. If the range of variability of X is about the same as the simulation-to-observation error, that is, if index $V_N(X)$ is near 1, then the error possibly is caused by only the simulated source of uncertainty; however, the other, and usually more likely explanation, is that the other sources of uncertainty contributed offsetting errors. The likelihood of offsetting errors increases the larger the range of variability is compared to the simulation-to-observation error, that is, the larger $V_N(X)$ is compared to 1 because, in that case, the simulation results indicate that the error is likely to be larger than the observed error.

The relative variability of the two sets of simulations as measured by V_N as defined in equation 10 for discharge statistics X including the POS mean discharge, selected FDC quantiles, and water year (WY) mean discharge at the calibration and prediction watersheds are presented and discussed in sections of this report below. Abbreviations for discharge series of different sources are defined for ease of discussion. The published discharge is denoted as obsQ. The discharge simulated with the base-model parameters is denoted as baseQ. The discharge series simulated with the 17 recalibrated parameter sets are denoted as recalibratedQ, and the discharge series simulated with the 1,000 random parameter sets are termed as randomQ.

Uncertainty of Simulated Discharge at Calibration Watersheds

In this section, the HSPF discharge simulations based on the parameters recalibrated to the 17 pairs of uncertain published discharge and the 1,000 simulations based on sampling the uncertainty of the base-model parameters are compared to the published discharge, and to a lesser extent, the base-model discharge at the calibration watersheds. The discussion of the results considers the POS mean discharge, selected FDC quantiles, and WY mean discharge flow statistics, focusing on the comparison of the medians of the two sets of simulations with the observed values and on the resulting V_N values. Because the calibration watersheds are considered in this section, the results here indicate how much error can be explained by the uncertainty of the published discharge and by the base-model parameter uncertainty at watersheds where it was possible to tune the parameters to the observations.

Period-of-Study Mean Discharges

The POS mean discharges of the HSPF simulations using parameters recalibrated to the 17 pairs of discharge series characterizing the published discharge uncertainty (recalibratedQs), and the HSPF simulations using the 1,000 randomly generated parameter sets characterizing the uncertainty of the base-model parameters (randomQs) from the Tinley Creek and Long Run watersheds are plotted (figs. 10*A*, *B*) to present the variation of each and their covariation. The spread of data values along each axis indicates the distribution of the uncertainty for that streamgage, and deviation of individual pair(s) from the line of proportionality, which extends from the origin to the pair of base-model mean discharges and along which the ratio of mean discharges at the Long Run and Tinley Creek streamgages is constant, indicates the differential effects between the two watersheds. The scattering of the points around the line of proportionality indicates approximate direct proportionality between the simulated mean discharges at the two watersheds, implying that the changes in grassland and forest parameters have similar effects on total volume in both watersheds. It is important to note that for the Tinley Creek streamgage, the baseQ mean discharge (table 13) is beyond the upper end of the range of the 17 recalibrated mean discharges (fig. 10*A*) and at the upper end of the range of the randomQ mean discharges (fig. 10*B*), whereas the obsQ mean discharge is larger. The Long Run baseQ mean discharge is similarly positioned relative to the simulated discharges, but its obsQ discharge is less than that of any of the simulations.

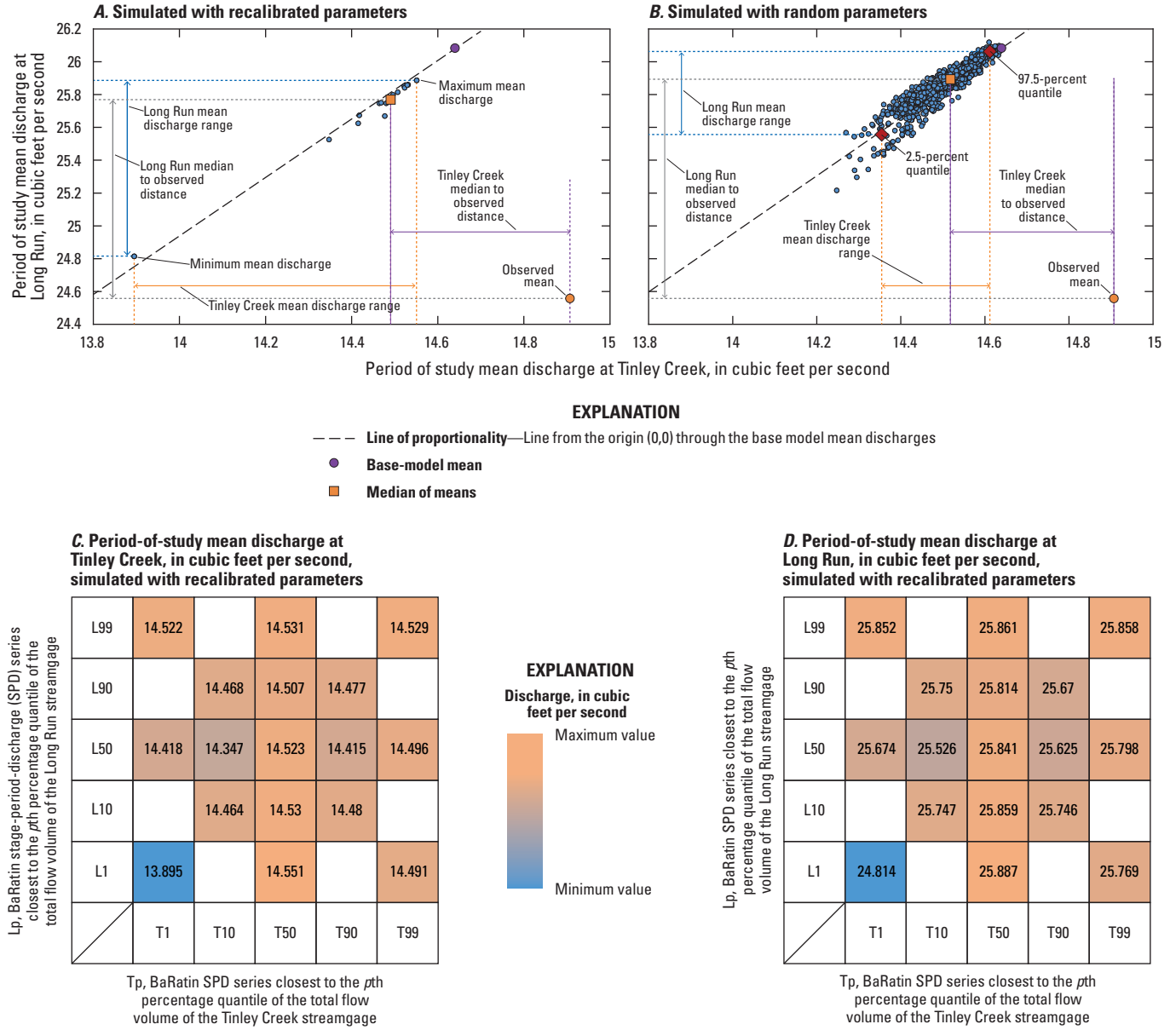

Figure 10. Relations between period of study (water years 1997–2015) daily discharges at Tinley Creek and Long Run watersheds simulated by Hydrological Simulation Program–FORTRAN. *A,* Simulated using 17 sets of recalibrated parameters that characterize uncertainty of published discharge (recalibratedQs). *B*, Simulated using 1,000 random parameter sets that characterize the uncertainty of the base-model parameters (randomQs), with definitions of quantities used to compute the normalized variability index values in table 13. The slope of the line of proportionality in panels *A* and *B*, the ratio of the simulated base-model mean discharge at Long Run to that at Tinley Creek, is 1.78. *C* and *D* display the simulated period of study mean discharges from the recalibratedQs at the Tinley Creek and the Long Run watersheds, respectively.

Table 13. Statistics of the means of daily discharge of the period of study (water years 1997–2015) obtained from published discharge records (obsQ), and Hydrological Simulation Program–FORTRAN simulated discharge series with base-model parameters (baseQ), with recalibrated parameters that characterize uncertainty of published discharge (recalibratedQs), and with random parameters that characterize the uncertainty of the base-model parameters (randomQs) for the Tinley Creek and Long Run watersheds.

[Units are in cubic feet per second except Median / obsQ, Relative median error, and V_N, which are unitless; Min, minimum; Max, maximum; %, percent; V_N, normalized variability index (eq. 10)]

Watershed	ObsQ mean	BaseQ mean	RecalibratedQs						RandomQs					
			Min	Max	Median	Median / obsQ	Relative median error	V_N	2.5%	97.5%	Median	Median / obsQ	Relative median error	V_N
Tinley Creek	14.9	14.6	13.9	14.6	14.5	0.97	0.028	0.787	14.4	14.6	14.5	0.97	0.026	0.335
Long Run	24.6	26.1	24.8	25.9	25.8	1.05	0.049	0.443	25.6	26.1	25.9	1.05	0.054	0.191

These opposite errors in mean discharge point to the effects of the two-watershed calibration scheme. The objective function, minimized during model calibration, consists of the sum of residuals in the selected observation groups from both calibration watersheds; therefore, the function is balancing the residuals between the two watersheds and does not return an optimal fit for either watershed.

For the recalibratedQs, one consideration is if the simulated POS mean discharges reflect the mean discharges of the pairs of BaRatin SPD realizations used for recalibration (fig. 4; table 8). The lowest pair (fig. 10*A*) consists of values from QT1L1 (which is the pair with the lowest mean discharge used for calibration for both watersheds), whereas the values from other pairs are clustered in a narrow range at some distance from the QT1L1 point. This separation of values results in a highly skewed distribution along the regression line. The highest discharge pair comes from QT50L1 and then the pair representing QT99L99 followed by the pair QT50L50 (figs. 10*A*, *C*, *D*). Therefore, the simulated POS mean discharge for the recalibratedQs is not directly proportional to the total volume of the calibration dataset. This lack of direct proportionality must result from the recalibrated parameters because they govern the simulations. It is important to note that the recalibration relies on various observation groups of which total volume is only one, and as discussed in the "Uncertainty of Discharge Series Realizations Selected for Recalibration" section, at least some of these statistics are not in direct proportion to total volume.

It is of interest that the range of observed discharges being used for the recalibrations (14.1–16.0 ft^3/s at Tinley Creek streamgage and 22.9–26.5 ft^3/s at Long Run streamgage; bottom row of table 8) is much wider than that of the POS mean discharge arising from the simulations results based on the recalibrated parameters (figs. 10*A*, *C*, *D*), although the POS mean discharge simulated with the QT1L1 parameters is less than the corresponding mean discharge used for calibration at the Tinley Creek streamgage (about 13.9 ft^3/s, figure 10*C*; compared to 14.1 ft^3/s, table 8). These differences also must be affected by constraints of the calibration process.

The POS mean discharges of the 1,000 randomQs from the Tinley Creek watershed plotted with the means from the Long Run watershed (fig. 10*B*) indicate the effects of random parameter variations sampled from the estimated covariance matrix. The data deviation around the line of proportionality is much wider than that shown in figure 10*A*, but because of the QTlL1 point in figure 10*A*, the data range is narrower. The set of points in figure 10*B* is also higher than those of recalibratedQs and closer to the POS-mean of baseQ simulation (notice that figure 10*A* and figure 10*B* have identical axes). The narrower range of the randomQ mean discharges is to be expected, considering that the points represent the parameter uncertainty for a fixed set of calibration data, whereas the recalibratedQ mean discharges represent recalibrations to a range of calibration discharges. It seems likely the narrower spread of the recalibratedQ mean discharges around the regression line compared to the randomQ mean discharges has a similar interpretation, and that it is the recalibration process that shifts the parameters resulting in a relatively tight line.

Considering the POS mean discharges of the two sets of simulations as collections of values characterizing the two sources of uncertainty, statistics of these collections and their relation to the observed mean discharge are presented in table 13. The relative median errors (eq. 9) are similar and small, within about 3 and 5 percent of the the obsQ mean at the Tinley Creek and Long Run streamgages, respectively; however, the range of randomQ POS mean discharges is narrower than that of the recalibratedQ mean discharges. As a result, the V_N values of the recalibratedQ mean discharges (0.787 and 0.443) are higher than those of the randomQs (0.335 and 0.191) in both watersheds, although less than one in all cases. Therefore, neither set of simulations appears to account for all the error or uncertainty in predicting the mean discharge, but according to these simulations, the published discharge uncertainty explains more than does the base-model

parameter uncertainty. The V_N values are also higher for the Tinley Creek results than for the Long Run results for each simulation type: the range of POS mean discharge is smaller for Tinley Creek than for Long Run but the smaller simulation median to obsQ differences for Tinley Creek affects the ratio more.

Flow-Duration Curves

Another flow characteristic commonly used to assess model fit is the FDCs of daily discharge, which can be viewed as the disaggregation of the POS mean discharge into daily magnitude bins and their associated frequencies. Summary statistics of selected FDC nonexceedance probability percentiles computed from obsQ, recalibratedQs, and randomQs daily discharges are presented in tables 14 and 15 for the Tinley Creek and Long Run watersheds, respectively. Judging from the median/obsQ ratios, generally underestimations in the medians of the FDC nonexceedance probability percentiles of the simulations resulted for nonexceedance probability percentiles up to the 20th at the Long Run streamgage and up to the 75th at the Tinley Creek streamgage. Underestimation is especially noticeable at the 1st nonexceedance probability percentile in both simulation sets for both watersheds. Overestimations resulted at the Long Run streamgage at the 75th percentile and above, whereas at the Tinley Creek streamgage overestimation begins at 90th nonexceedance probability percentile, but the simulated and observed values are close, except at the 99th percentile, where both sets of simulations substantially underestimate observed values. Recall that the POS mean discharge is overestimated at the Long Run streamgage but underestimated at the Tinley Creek streamgage by both simulation sets; the FDC results are consistent with the POS mean results because the mean is controlled by the high flows.

The relative median error is very high, nearly 100 percent, for the 1st percentile, but then drops quickly for both parameter sets and for both watersheds. At the Tinley Creek streamgage, the lowest relative median error is 0.4 percent for the recalibratedQs and 1.3 percent for the randomQs; the highest relative median error excluding the 1st percentile is 39.6 percent for the recalibratedQs and 41.5 percent for the randomQs (table 14). At the Long Run streamgage, the lowest relative median error is 1.2 percent for the recalibratedQs and 0.6 percent for the randomQs; the highest relative median error excluding the 1st percentile is 38.1 percent for the recalibratedQs and 41.5 percent for the randomQ (table 15).

The normalized variability index (V_N) values of both simulation sets are low at the low discharge percentiles, increase as discharge magnitudes increase, and then turn lower again at high discharges. The very low V_N values at the 1st percentile discharge by the recalibratedQs in both watersheds indicate that a negligible fraction of the error of this quantity can be explained by the uncertainty of the recalibrated discharges. According to the V_N values, a middle to large fraction of the error is explained by recalibratedQs as discharge magnitudes increase, with values at two percentiles in both watersheds exceeding 1. The V_N values for the randomQ FDC nonexceedance probability percentiles are roughly similar to those from the recalibratedQs, except that for Long Run, the V_N values are generally larger for the recalibratedQs as a result of wider ranges, because the relative median errors are similar. As discussed previously, one way to interpret V_N values exceeding 1 is to consider errors from multiple sources that offset each other; in this particular case, it is also useful to consider that an FDC is a curve and a set of quantiles that approximate a distribution. Therefore, if the FDC of the simulated distribution crosses that of the observed distribution, there will be 1 or 2 percentiles per crossing where they closely agree. Because of this phenomenon, it may be most useful to consider the medians of the V_N values, which range of 0.308 to 0.603 and are thus similar to those of the POS mean discharges.

Water Year Mean Discharges

The WY mean discharges can be viewed as the disaggregation of the POS mean discharge to annual time steps. These discharges for the obsQs, baseQs, recalibratedQs, and randomQs for the Tinley Creek and Long Run watersheds are presented in figure 11. WY mean discharge for the published records (obsQ) and selected statistics consisting of median/obsQ, relative median error, and V_N for the recalibratedQ and randomQ simulations are presented in tables 16 and 17 for the Tinley Creek and Long Run watersheds, respectively. The recalibratedQs and randomQs have similar median/obsQ ratios in each watershed. At the Tinley Creek streamgage, overestimation occurred for 7 WYs with the largest median/obsQ being about 1.34 and a corresponding relative median error of 34 percent; underestimation occurred for 11 WYs with the lowest median/obsQ being about 0.76 and a corresponding relative median error of 24 percent (table 16). At the Long Run streamgage, overestimation occurred for 13 WYs with the largest median/obsQ being about 1.61 and a corresponding relative median error of 61 percent; underestimation occurred for 7 WYs with the lowest median/obsQ being about 0.82 and a corresponding relative median error of 18 percent (table 17).

At Tinley Creek, the range of the recalibratedQ WY mean discharges encloses the obsQ WY mean for 3 years, overestimates the obsQ mean for 5 years, and underestimates the obsQ mean for 11 years (fig. 11*A*); the corresponding values for randomQs are 1, 7, and 11 years, respectively (fig. 11*B*). At Long Run, the range of the recalibratedQ WY mean discharges encloses the obsQ WY mean for 4 years, overestimates the obsQ mean for 9 years, and underestimates the obsQ mean for 6 years (fig. 11*C*); the corresponding values for randomQs are 0, 13, and 6 years, respectively (fig. 11*D*). This mixture of enclosure, overestimation, and underestimation of WY means

Table 14. Statistics of selected nonexceedance probability percentiles of flow-duration curves derived from published discharge records and Hydrological Simulation Program–FORTRAN simulated discharge at Tinley Creek watershed.

[ObsQ, published discharge records; ft³/s, cubic feet per second; V_N, normalized variability index (eq. 10); NA, not applicable]

Percentile	ObsQ (ft³/s)	RecalibratedQs			RandomQs		
		Median / obsQ	Relative median error	V_N	Median / obsQ	Relative median error	V_N
1	0.20	0.002	0.998	0.00059	0.003	0.997	0.0401
5	0.50	0.604	0.396	0.121	0.585	0.415	0.266
10	0.80	0.803	0.197	0.227	0.771	0.229	0.303
20	1.50	0.871	0.129	0.229	0.838	0.162	0.328
25	1.90	0.895	0.105	0.341	0.864	0.136	0.347
50	4.50	0.932	0.068	0.431	0.907	0.093	0.322
75	12	0.956	0.044	0.508	0.939	0.061	0.308
80	15	1.009	0.009	2.193	0.995	0.005	4.784
90	32	1.028	0.028	0.599	1.034	0.034	0.238
95	60	1.004	0.004	4.618	1.013	0.013	0.682
99	186	0.890	0.110	0.308	0.898	0.102	0.129
Median	NA	NA	NA	0.341	NA	NA	0.308

Table 15. Statistics of selected nonexceedance probability percentiles of flow-duration curves derived from published discharge records and Hydrological Simulation Program–FORTRAN simulated discharge at Long Run watershed.

[ObsQ, published discharge records; ft³/s, cubic feet per second; V_N, normalized variability index (eq. 10); NA, not applicable]

Percentile	ObsQ (ft³/s)	RecalibratedQs			RandomQs		
		Median / obsQ	Relative median error	V_N	Median / obsQ	Relative median error	V_N
1	0.80	0.036	0.964	0.00161	0.033	0.967	0.0385
5	1.30	0.619	0.381	0.077	0.585	0.415	0.197
10	1.90	0.846	0.154	0.215	0.810	0.190	0.286
20	3.00	0.973	0.027	0.753	0.938	0.062	0.660
25	3.60	1.012	0.012	2.461	0.966	0.034	1.376
50	8.80	0.962	0.038	0.688	0.925	0.075	0.431
75	21	1.028	0.028	0.738	1.006	0.006	4.259
80	27	1.045	0.045	0.603	1.021	0.021	1.061
90	54	1.066	0.066	0.329	1.061	0.061	0.203
95	93	1.115	0.115	0.180	1.121	0.121	0.099
99	275	1.023	0.023	1.277	1.059	0.059	0.333
Median	NA	NA	NA	0.603	NA	NA	0.333

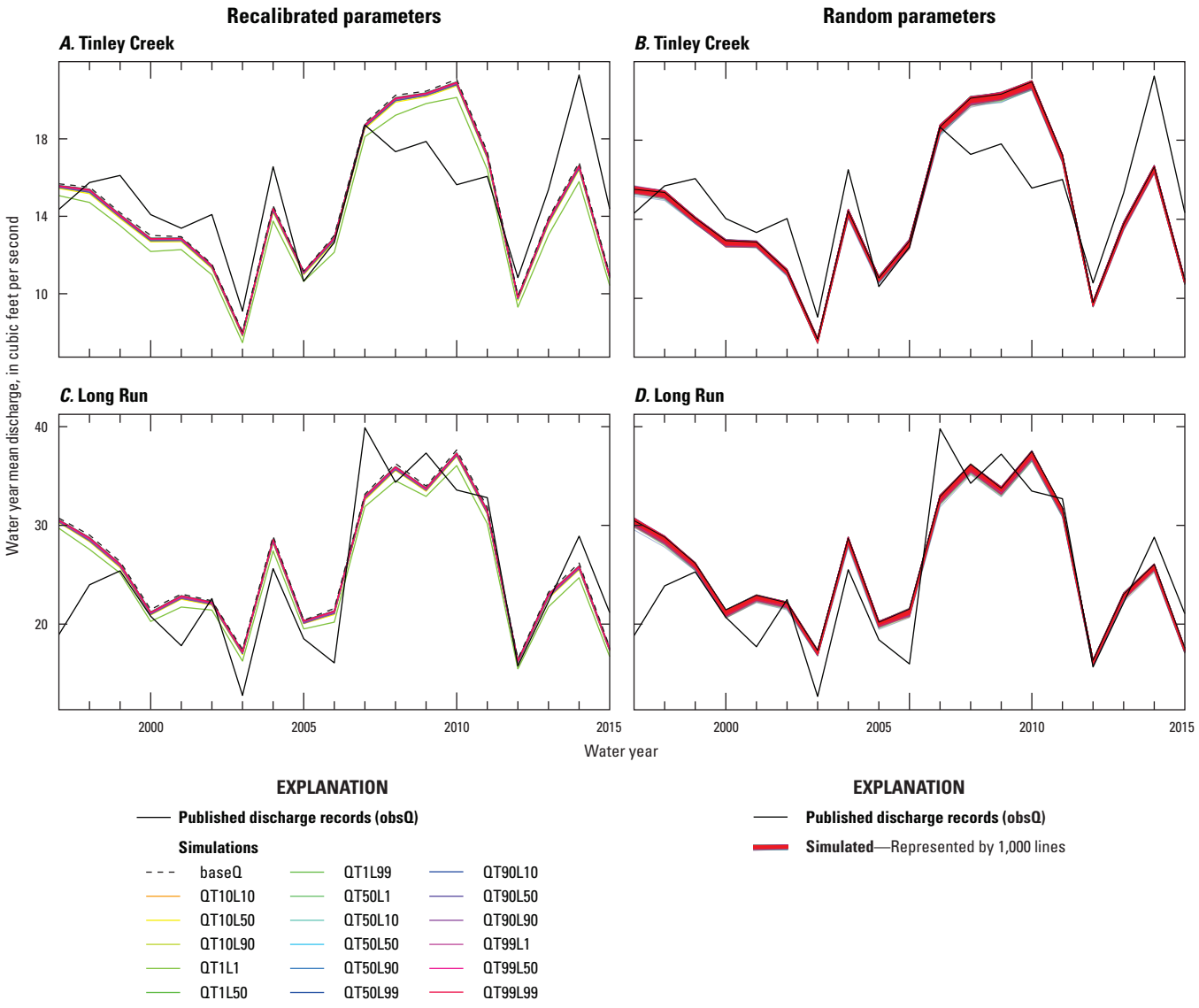

Figure 11. Water year (WY) mean discharges from Hydrological Simulation Program–FORTRAN simulations and published daily discharges for WYs 1997–2015. *A*, Discharge computed with the base and 17 recalibrated parameters that characterize uncertainty of published discharge at Tinley Creek watershed. *B*, Discharge computed with 1,000 randomly sampled parameter sets that characterize the uncertainty of the base-model parameters at Tinley Creek watershed. *C*, Discharge computed with the base and 17 recalibrated parameters that characterize uncertainty of published discharge at Long Run watershed. *D*, Discharge computed with 1,000 randomly sampled parameter sets that characterize the uncertainty of the base-model parameters at Long Run watershed.

of obsQs indicates limited overall bias, although the proportions of years that are overestimated and underestimated for each of the streamgages are consistent with the POS mean discharge differences, for which the Tinley Creek watershed values underestimate the obsQ and the Long Run watershed values overestimate the obsQ (fig. 10*A*, *B*).

The V_N values of the recalibratedQ WY means are higher than those of the randomQs, as a result of wider ranges (the medians are similar) (tables 16 and 17), but are generally lower than the V_N values of the POS mean discharges and the FDCs. The V_N values and the rating-curve periods (table 2) do not appear to be statistically correlated.

Table 16. Water year (WY) mean discharge of published discharge records and statistics of WY mean discharge of simulated discharge based on recalibrated and random parameters at the Tinley Creek watershed.

[WY, water year; USGS, U.S. Geological Survey; no., number; ObsQ, published daily discharge records; ft^3/s, cubic feet per second; RecalibratedQs, discharge series simulated with the recalibrated parameters; RandomQs, discharge series simulated with random parameters; V_N: normalized variability index (eq. 10); NA, not applicable]

WY	USGS rating curve no.	ObsQ (ft^3/s)	RecalibratedQs			RandomQs		
			Median / obsQ	Relative median error	V_N	Median / obsQ	Relative median error	V_N
1997	24	14.31	1.084	0.08	0.221	1.085	0.08	0.117
1998	24	15.69	0.974	0.03	0.847	0.976	0.02	0.404
1999	24	16.06	0.870	0.13	0.137	0.872	0.13	0.060
2000	24, 25	14.03	0.910	0.09	0.276	0.915	0.09	0.115
2001	25	13.33	0.958	0.04	0.538	0.959	0.04	0.222
2002	25	14.04	0.810	0.19	0.093	0.810	0.19	0.043
2003	25	9.05	0.869	0.13	0.206	0.871	0.13	0.083
2004	25	16.51	0.868	0.13	0.158	0.870	0.13	0.068
2005	25	10.60	1.043	0.04	0.559	1.039	0.04	0.278
2006	25	12.59	1.016	0.02	1.883	1.018	0.02	0.698
2007	25	18.66	0.998	0.00	8.384	0.998	0.00	5.815
2008	25, 26	17.28	1.158	0.16	0.168	1.160	0.16	0.063
2009	26	17.81	1.138	0.14	0.114	1.138	0.14	0.067
2010	26	15.58	1.337	0.34	0.077	1.341	0.34	0.032
2011	26	16.01	1.067	0.07	0.367	1.071	0.07	0.127
2012	26	10.78	0.904	0.10	0.260	0.904	0.10	0.105
2013	26, 27	15.33	0.895	0.11	0.246	0.895	0.10	0.084
2014	27	21.24	0.778	0.22	0.091	0.781	0.22	0.031
2015	27	14.33	0.756	0.24	0.072	0.758	0.24	0.031
Median	NA	NA	NA	NA	0.233	NA	NA	0.094

Table 17. Water year (WY) mean discharge of published discharge records and statistics of WY mean discharge of simulated discharge based on recalibrated and random parameters at Long Run watershed.

[WY, water year; USGS, U.S. Geological Survey; no., number; ObsQ, published discharge records; ft³/s, cubic feet per second; RecalibratedQs, discharge series simulated with the recalibrated parameters; RandomQs, discharge series simulated with random parameters; V_N: normalized variability index (eq. 10); NA, not applicable]

WY	USGS rating curve no.	ObsQ (ft³/s)	RecalibratedQs			RandomQs		
			Median / obsQ	Relative median error	V_N	Median / obsQ	Relative median error	V_N
1997	28	18.83	1.613	0.61	0.037	1.617	0.62	0.026
1998	28, 29	23.88	1.195	0.19	0.130	1.201	0.20	0.063
1999	29	25.29	1.023	0.02	0.799	1.029	0.03	0.343
2000	29	20.69	1.016	0.02	1.421	1.024	0.02	0.548
2001	29, 30	17.71	1.279	0.28	0.117	1.284	0.28	0.048
2002	30, 31	22.48	0.980	0.02	0.936	0.982	0.02	0.576
2003	31	12.68	1.344	0.34	0.115	1.354	0.35	0.046
2004	31	25.51	1.113	0.11	0.212	1.120	0.12	0.095
2005	31	18.42	1.095	0.09	0.240	1.092	0.09	0.140
2006	31	15.97	1.323	0.32	0.112	1.331	0.33	0.055
2007	31	39.79	0.823	0.18	0.076	0.825	0.18	0.047
2008	31	34.27	1.043	0.04	0.497	1.048	0.05	0.202
2009	31	37.22	0.902	0.10	0.117	0.905	0.10	0.087
2010	31	33.48	1.109	0.11	0.177	1.114	0.11	0.088
2011	31, 32	32.70	0.957	0.04	0.489	0.964	0.04	0.249
2012	32	15.66	1.030	0.03	0.933	1.034	0.03	0.383
2013	32	22.18	1.029	0.03	1.023	1.033	0.03	0.342
2014	32	28.81	0.891	0.11	0.190	0.898	0.10	0.087
2015	32	21.06	0.823	0.18	0.107	0.828	0.17	0.054
Median	NA	NA	NA	NA	0.201	NA	NA	0.092

Uncertainty of Simulated Discharge at Prediction Watersheds

Applying the recalibrated and random parameter sets to the seven prediction watersheds provides an opportunity to evaluate the transferability of regional parameters at noncalibration locations where the characteristics of climatic forcing, watershed properties, and measured discharge are different from those used in calibration. Thus, this application gives an indication of the fraction of the prediction error that the processes considered in this report, calibration discharge uncertainty and parameter uncertainty, can explain at ungaged sites. In addition to the inherent differences between calibration and prediction sites (that is, that the parameters of the model had been adjusted to account for the various model and input errors at the calibration watersheds but not the prediction watersheds), there are a few special considerations regarding why the prediction errors may increase at the prediction sites used in this study:

(1). The sites are located on the border of or outside the CCPN coverage; therefore, precipitation inputs were supplemented with the use of disaggregated data from NWS–COOP daily stations (see "Watersheds and Model-Input Data" section).

(2). Although the prediction watersheds have published discharge computed using the same basic approach as the calibration watersheds (that is, using a stage-discharge rating), and therefore may be expected to have a similar errors, the Hart Ditch at Munster streamgage (fig. 1, table 1), which is subject to backwater effects and therefore has a more complex rating, is expected to have reduced data accuracy compared to the other streamgages.

(3). The Flag Creek, Hickory Creek, Hart Ditch at Dyer, and Hart Ditch at Munster streamgages have effluent discharges added to their simulated values. The hourly

effluent discharges were approximated from their reported monthly data using a uniform distribution and therefore are less accurate (Soong and Over, 2015).

(4). The Midlothian Creek, Hickory Creek, and Butterfield Creek watersheds have reservoirs that were not explicitly simulated (as does Long Run) (table 1).

As discussed in the "Uncertainty of Simulated Discharge at Calibration Watersheds" section, the discussion of the results at the prediction watersheds will focus on the V_N values computed using the POS mean discharge, FDCs, and WY mean discharge flow statistics.

Period of Study Mean Discharge

The presentation of the results on uncertainty of simulated discharge begins with the POS mean discharge. The published POS mean discharge (obsQ), median/obsQ ratio, relative median error, and V_N statistics based on the recalibratedQ and randomQ simulations at the seven prediction watersheds are given in table 18. For comparison, the information for the two calibration watersheds is listed at the bottom of the table. The differences consist of overestimation (median/obsQ > 1) in five watersheds and underestimation (median/obsQ < 1) in two watersheds that have less CCPN coverages. The relative median errors of POS mean discharge simulated with the two parameter groups vary from approximately 6 percent at the Midlothian Creek streamgage to approximately 36 percent at the Butterfield Creek streamgage. These relative median errors are substantially larger than the relative median errors of 3 and 5 percent at the calibration streamgages (table 13), except at the Midlothian Creek streamgage. The corresponding V_N values vary from 0.029 at Flag Creek to 0.248 at Midlothian Creek for the recalibratedQ simulations and from 0.026 at Flag Creek to 0.131 at Midlothian Creek for the randomQ simulations. These values are low relative to those obtained in Tinley Creek and Long Run, which range from 0.443 to 0.787 for the recalibratedQs and from 0.191 to 0.335 for the randomQs (table 13), at least in part because of the larger relative median errors. Low V_N values indicate a smaller fraction of the prediction errors in simulated POS mean discharge can be explained in the predictive watersheds than in the calibration watersheds. Between the two sets of simulations, the recalibratedQs produce larger V_N values, largely because their ranges are wider (their medians, and therefore their median-to-observation errors, are similar).

Flow-Duration Curves

Regarding the FDC results at the prediction watersheds, to focus the discussion on explaining the predictive errors, only the V_N values for selected FDC quantiles from the recalibratedQs and randomQs are presented here (table 19). Among the seven prediction watersheds, the V_N values from both sets of simulations in the Butterfield and Flag Creek watersheds (fig. 1) are low for almost all discharge percentiles and the lowest overall, measured by median V_N. The Hart Ditch at Munster watershed has the highest V_N overall, with a median V_N value of 0.388 for the recalibratedQs and 0.920 for the randomQs, despite concerns regarding discharge measurement accuracy discussed in the "Uncertainty of Simulated Discharge at Prediction Watersheds" section. Furthermore, these V_N values are as high as or higher than the V_N values of the FDCs of the calibration watershed simulations (tables 14 and 15); nevertheless, the V_N values for the FDCs at all the other prediction watersheds are lower than those of the calibration watersheds. As in the case of the calibration watersheds, the V_N values associated with the FDCs in the prediction watersheds follow similar patterns with respect to variation with percentile (that is, V_N values vary from low to high flows so that there is occasionally a percentile where the V_N values are large, such as for Hart Ditch at Dyer and Hart Ditch at Munster in the recalibratedQ results and for Skokie River in the randomQ results). However, contrary to the calibration watershed results, the randomQ simulations mostly resulted in higher V_N values than did the recalibratedQs. This difference may have resulted because other sources of errors, such as approximations of the magnitudes of effluents and (or) the effects of flood-control reservoirs (see "Watersheds and Model-Input Data" section) that are unrelated to discharge uncertainty, are better explained by the parameter uncertainty of randomQs than by recalibratedQs. Another consequence of the higher FDC V_N values for the randomQs is that they are substantially larger than those for the POS mean values for the randomQs (table 18), whereas the values for the recalibratedQ simulations are similar between POS mean (table 18) and FDCs.

Water Year Mean Discharges

An indication of the temporal structure of the variability and accuracy of the two simulation sets for the prediction watersheds can be obtained from examining the annual mean discharges. WY mean discharges from WY 1997 to 2015 based on the published discharge records and simulated discharge from recalibratedQs and randomQs are presented in figure 12. As it is for the calibration watersheds (fig. 11), the two sets of simulations are similar for each watershed. Further, the overall directions of the errors are consistent

Table 18. Published period of study (POS, water years 1997 to 2015) mean discharge and statistics of POS mean discharge simulated with Hydrological Simulation Program–FORTRAN using the recalibrated and random parameters at the seven prediction watersheds. For comparison, the POS mean discharge statistics from the two calibration watersheds are listed in the last two rows of the table.

[Units are in cubic feet per second, except Median / ObsQ, Relative median error, and V_N, which are unitless. ObsQ, published discharge records; RecalibratedQs, discharge series simulated with the recalibrated parameters; RandomQs, discharge series simulated with random parameters; Min, minimum; Max, maximum; V_N, normalized variability index (eq. 10); %, percent]

Watershed	ObsQ (ft^3/s)	RecalibratedQs						RandomQs					
		Min	Max	Median	Median / obsQ	Relative median error	V_N	2.5%	97.5%	Median	Median / obsQ	Relative median error	V_N
Butterfield Creek	22.6	30.0	30.8	30.7	1.358	0.36	0.052	30.5	31.0	30.9	1.365	0.36	0.035
Hart Ditch at Dyer	41.4	31.4	33.5	33.3	0.805	0.20	0.133	32.9	33.9	33.5	0.810	0.19	0.060
Flag Creek	32.3	38.8	39.2	39.1	1.209	0.21	0.029	39.0	39.3	39.2	1.212	0.21	0.026
Hickory Creek	125.4	137.7	142.5	141.9	1.132	0.13	0.144	140.9	143.5	142.6	1.138	0.14	0.075
Midlothian Creek	16.9	17.5	17.9	17.9	1.056	0.06	0.248	17.8	18.0	17.9	1.060	0.06	0.131
Hart Ditch at Munster	97.8	85.8	89.3	88.9	0.908	0.09	0.191	88.2	89.8	89.3	0.913	0.09	0.094
Skokie River	24.8	29.2	30.1	30.0	1.211	0.21	0.089	29.8	30.3	30.1	1.216	0.22	0.046
Tinley Creek[1]	14.9	13.9	14.6	14.5	0.972	0.03	0.787	14.4	14.6	14.5	0.974	0.03	0.335
Long Run[1]	24.6	24.8	25.9	25.8	1.049	0.05	0.443	25.6	26.1	25.9	1.054	0.05	0.191

[1]Calibration watershed.

Table 19. Normalized variability index (V_N, eq. 10) of discharge magnitudes at selected nonexceedance probability percentiles of the flow-duration curves at the seven prediction watersheds for water years 1997 to 2015 based on Hydrological Simulation Program–FORTRAN simulations.

[Watershed locations shown in figure 1. RecalibratedQs, discharge series simulated with the recalibrated parameters; RandomQs, discharge series simulated with random parameters]

Percentile	RecalibratedQs							RandomQs						
	Butterfield Creek	Hart Ditch at Dyer	Flag Creek	Hickory Creek	Midlothian Creek	Hart Ditch at Munster	Skokie River	Butterfield Creek	Hart Ditch at Dyer	Flag Creek	Hickory Creek	Midlothian Creek	Hart Ditch at Munster	Skokie River
1	0.197	0.003	0.000	0.124	3.73E−05	0.016	0.010	15.436	0.015	0.011	0.433	0.007	0.108	0.173
5	0.023	0.008	0.011	0.104	0.028	0.297	0.235	0.270	0.038	0.034	0.274	0.165	1.646	2.001
10	0.045	0.038	0.010	0.053	0.106	0.388	0.570	0.179	0.082	0.046	0.140	0.203	1.744	4.702
20	0.045	0.172	0.016	0.056	0.232	4.009	0.201	0.142	0.229	0.055	0.127	0.411	0.925	0.404
25	0.041	0.230	0.013	0.076	0.331	17.640	0.244	0.121	0.260	0.046	0.136	0.454	1.155	0.357
50	0.065	1.280	0.017	0.054	0.149	0.606	0.082	0.164	2.316	0.044	0.108	0.198	0.441	0.135
75	0.057	0.325	0.028	0.153	0.216	0.596	0.175	0.120	0.226	0.066	0.201	0.249	1.076	0.242
80	0.042	0.275	0.039	0.168	0.164	0.541	0.126	0.061	0.178	0.076	0.183	0.121	0.920	0.119
90	0.041	0.095	0.019	0.152	0.096	0.175	0.056	0.033	0.048	0.032	0.095	0.078	0.096	0.034
95	0.042	0.061	0.019	0.194	0.088	0.091	0.034	0.028	0.033	0.040	0.117	0.050	0.070	0.022
99	0.168	0.221	0.058	0.675	0.061	0.065	0.230	0.110	0.125	0.056	0.365	0.039	0.053	0.067
Median	0.045	0.172	0.017	0.124	0.106	0.388	0.175	0.121	0.125	0.046	0.140	0.165	0.920	0.173

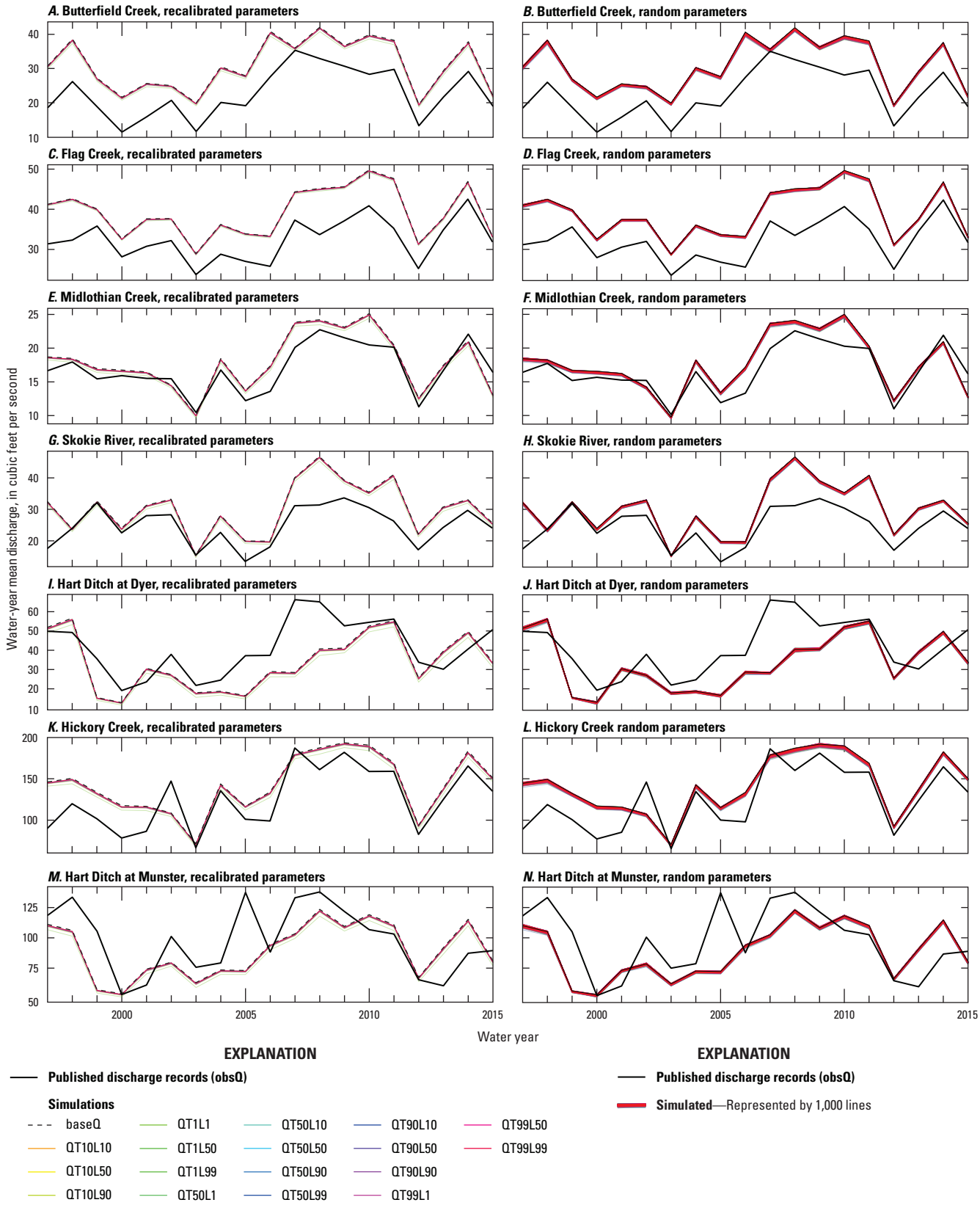

Figure 12. Water year (WY) mean discharges from Hydrological Simulation Program–FORTRAN simulations and from published daily discharges for WYs 1997-2015, computed with the base model and 17 recalibrated parameters that characterize uncertainty of published discharge (left column) and with 1,000 randomly sampled parameter sets that characterize the uncertainty of the base-model parameters (right column).

with the mean POS discharge errors (table 18); for example, the Flag Creek and Butterfield Creek simulations are usually higher than the published, whereas the Hart Ditch at Dyer simulations are usually lower. Also, similar to the calibration watersheds (fig. 11), the temporal patterns of higher and lower years among the prediction watersheds generally agree, as would be expected because they are in the same region.

The V_N values computed from the WY mean discharges from the recalibratedQs are higher than those from randomQs (table 20), but the values, having medians ranging from 0.025 to 0.172, generally are lower than the V_N values in the two calibration watersheds (tables 16 and 17), largely because of larger simulation to observational errors. As with the FDC nonexceedance probability percentiles and the POS mean discharges, the V_N values in Butterfield Creek and Flag Creek watersheds (fig. 1) are the lowest in both simulation sets. These relatively low V_N values for both sets of simulations indicate that neither source of uncertainty explains a large fraction of the simulation-to-observation error in the WY mean discharges in the prediction watersheds.

The V_N values from the WY mean discharges are similar to those from the POS mean values, both in overall magnitude and in variation by watershed. For the recalibratedQs, they are also similar to the FDC V_N values in overall magnitude but are substantially smaller for the randomQs. By contrast, as noted in the "Uncertainty of Simulated Discharge at Calibration Watersheds" section, the WY mean V_N values decreased relative to the POS mean V_N values for the calibration watersheds and decreased for both recalibratedQs and randomQs as compared to the V_N values for the FDCs. Overall, although FDCs represent increased detail with respect to magnitude compared to POS mean discharge and WY mean discharges increased detail with respect to timing, the V_N values have a general trend with these increases in detail when both the recalibratedQs and randomQs and the calibration and prediction watersheds are considered together. One reason for this lack of trend with detail may be the normalized nature of the V_N statistic itself; trends may be present in its components.

Table 20. Normalized variability index (V_N, eq. 10) values of water year (WY) mean discharges from WY 1997 to 2015 at the seven prediction watersheds based on Hydrological Simulation Program–FORTRAN simulations.

[Watershed locations shown in figure 1. WY, water year; RecalibratedQs, discharge series simulated with the recalibrated parameters; RandomQs, discharge series simulated with random parameters]

WY	RecalibratedQs							RandomQs						
	Butterfield Creek	Hart Ditch at Dyer	Flag Creek	Hickory Creek	Midlothian Creek	Hart Ditch at Munster	Skokie River	Butterfield Creek	Hart Ditch at Dyer	Flag Creek	Hickory Creek	Midlothian Creek	Hart Ditch at Munster	Skokie River
1997	0.029	0.937	0.017	0.038	0.096	0.207	0.027	0.026	0.435	0.022	0.026	0.076	0.138	0.019
1998	0.043	0.211	0.022	0.095	0.654	0.070	0.962	0.034	0.096	0.021	0.058	0.367	0.036	0.703
1999	0.045	0.030	0.044	0.075	0.170	0.023	3.437	0.032	0.013	0.040	0.042	0.091	0.011	0.785
2000	0.032	0.088	0.032	0.055	0.371	7.721	0.380	0.026	0.072	0.038	0.035	0.196	6.509	0.188
2001	0.044	0.171	0.030	0.080	0.275	0.142	0.168	0.028	0.080	0.024	0.042	0.148	0.062	0.087
2002	0.080	0.079	0.030	0.048	0.172	0.063	0.098	0.065	0.047	0.032	0.030	0.118	0.041	0.053
2003	0.039	0.252	0.031	0.379	0.346	0.129	1.706	0.028	0.104	0.023	0.214	0.208	0.056	1.365
2004	0.045	0.136	0.029	0.417	0.170	0.229	0.094	0.028	0.074	0.025	0.201	0.089	0.121	0.051
2005	0.046	0.035	0.026	0.141	0.138	0.021	0.060	0.033	0.023	0.025	0.086	0.092	0.012	0.036
2006	0.043	0.126	0.025	0.081	0.076	0.368	0.263	0.031	0.062	0.024	0.045	0.044	0.166	0.157
2007	0.980	0.025	0.029	0.284	0.067	0.053	0.060	0.706	0.009	0.033	0.215	0.051	0.028	0.040
2008	0.060	0.057	0.020	0.128	0.241	0.148	0.039	0.039	0.028	0.020	0.065	0.126	0.071	0.020
2009	0.069	0.078	0.022	0.236	0.152	0.113	0.081	0.057	0.043	0.028	0.158	0.113	0.070	0.053
2010	0.041	0.484	0.024	0.089	0.061	0.178	0.104	0.032	0.261	0.028	0.055	0.040	0.085	0.057
2011	0.062	0.838	0.019	0.363	2.532	0.313	0.044	0.042	0.464	0.018	0.163	0.682	0.135	0.023
2012	0.054	0.114	0.031	0.196	0.174	1.244	0.085	0.039	0.053	0.025	0.109	0.095	0.511	0.043
2013	0.071	0.174	0.088	0.253	0.496	0.075	0.096	0.037	0.062	0.057	0.098	0.195	0.029	0.042
2014	0.057	0.155	0.056	0.177	0.244	0.079	0.145	0.036	0.065	0.044	0.081	0.130	0.036	0.071
2015	0.117	0.062	0.127	0.125	0.048	0.168	0.351	0.075	0.031	0.098	0.066	0.031	0.092	0.135
Median	0.046	0.126	0.029	0.128	0.172	0.142	0.098	0.034	0.062	0.025	0.066	0.113	0.070	0.053

Summary

A Lake Michigan Diversion Accounting (LMDA) system for estimating the volume of water diverted annually from Lake Michigan by the State of Illinois has been developed by U.S. Army Corps of Engineers-Chicago District. Discharge from the diverted Lake Michigan watershed in northeastern Illinois and northwestern Indiana, approximately 673 square miles (mi^2), has been estimated with the Hydrological Simulation Program–FORTRAN (HSPF) because most catchments in the diverted watershed are ungaged. This report presents a study of two components of the overall uncertainty of prediction in the ungaged watersheds. One component is the effect of the uncertainty of the published discharge records used for calibrating the HSPF parameters. The other component is the uncertainty of simulated discharge resulting from the uncertainty of the parameters. A two-watershed calibration scheme developed previously is used to determine parameters that minimize an objective function that is computed using discharge records from two calibration watersheds. The Parameter ESTimation and uncertainty analysis (PEST) package is applied for optimizing the parameters. The uncertainty components are investigated at nine gaged watersheds in or near the diverted watershed. Tinley Creek and Long Run in Illinois are the calibration watersheds, and Flag Creek, Skokie River, Butterfield Creek, Midlothian Creek, and Hickory Creek, all in Illinois, and Hart Ditch at Dyer and Hart Ditch at Munster in Indiana are used as prediction watersheds.

In the HSPF calibration, only the parameters used for simulating runoff from the pervious areas of the watersheds were adjusted in this study. The pervious areas were assigned two land-cover types, forest and grassland. Each is modeled with a set of 17 adjustable parameters. The resulting set of 34 parameters were calibrated jointly at the two calibration watersheds.

The uncertainty of published daily discharge originates from the rating curves used for converting measured stages to discharges at a streamgage. At the calibration watershed streamgages, the uncertainty is computed by adapting the Bayesian rating curve estimation (BaRatin) method with a Stage-Period-Discharge (SPD) extension, which provides rating curves, their uncertainties, and the uncertainties of computed discharges given a sequence of known rating curve periods. The BaRatin SPD analysis provided a set of 500 equally likely discharge realizations, from which uncertainty statistics were computed. As a ratio of standard deviation to the mean, that is, as the coefficient of variation (CV), the published daily discharge uncertainty of the mean discharge of the period used in applying the technique, water years (WYs) 1995–2016, was estimated to be about 3 percent at both streamgages.

To characterize the effects of the uncertainty of published daily discharge records on the pervious land parameters, 17 pairs of BaRatin SPD discharge series realizations, which span the 1-percent to 99-percent quantiles of mean discharge for the two calibration watersheds during the period used in HSPF simulations, WYs 1996–2015, were selected to use for recalibration. These few realizations were selected because the calibration technique is computationally intensive and dependent on expert judgement.

Uncertainty of the HSPF model parameters was also estimated by a Bayesian technique, applying a utility program provided with PEST to the pre- and post-optimization parameters of the base model, which is the model whose calibration discharges are the published daily discharge records at the two calibration streamgages. The result was an estimated parameter covariance matrix, from which 1,000 parameter sets were sampled and used for simulation, to estimate the effects of the parameter uncertainty on simulated discharge.

As a result of the different Bayesian techniques applied in the uncertainty analyses, in particular the nonrandom sampling of the uncertain discharge realizations, the uncertainties predicted by the two sets of simulations could not be added but instead were analyzed separately using a common normalized variability index (V_N). V_N is defined as the ratio of one-half the range of simulated values of a given flow statistic to the absolute value of the simulation error (that is, the difference between the simulation median and the observed value). Using this definition, large values of V_N relative to 1 indicate the source of uncertainty modeled by the simulations can explain a large fraction of the error, whereas small values indicate that the source of uncertainty modeled by the simulations can only explain a small fraction of the error, and the components of uncertainty that were not considered must explain most of the error. The flow statistics analyzed consist of mean discharge during the period of study (POS, WYs 1997–2015), nonexceedance probability percentiles of the flow-duration curves (FDCs), and WY mean discharges during the POS.

At the calibration watersheds, for both sets of simulations, the median simulation POS mean discharges had opposite errors because of the two-watershed calibration scheme balancing errors between the watersheds: the Long Run watershed had a higher simulated mean than observed and the Tinley Creek watershed had a lower simulated mean than observed. The results for the other flow statistics considered, FDCs and WY mean discharges, were consistent with the overall bias, but the bias was not uniform. At the Tinley Creek streamgage, the FDC quantiles of the simulations were biased low at all but the lowest quantiles, although slightly high at the highest. At the Long Run streamgage the results were similar but the quantiles where these biases occurred were shifted upward. For the WY mean discharges at each watershed, high bias was noted for a group of years, and low bias was noted for another group of years, but these groups of years were not the same between the two watersheds. The Tinley Creek watershed, consistent with its long-term low discharge bias, had more undersimulated years and the Long Run watershed had more oversimulated years.

At the calibration watersheds for both sets of simulations, the V_N values were determined to range from about 0.2 to about 0.8 for the POS mean discharges, from about 0.3 to about 0.6 as a median of the V_N values over each set of FDC quantiles, and from about 0.1 to 0.2 as a median for the WY

mean discharge series. In agreement with its smaller POS mean discharge simulation error, the Tinley Creek streamgage POS mean V_N values were larger than those at the Long Run streamgage. The pattern that V_N values were appreciably larger at the Tinley Creek streamgage was not observed in the FDC and WY mean discharge V_N values. Although not strictly comparable because of the sampling differences, the V_N values associated with the 17 recalibrations that were measuring the effect of uncertain published discharge were larger than those from the random parameter sampling because of having a wider range of values. The resulting V_N values, being primarily smaller than 1, indicate a substantial amount of unexplained error at the calibration watersheds.

At the seven prediction watersheds for both sets of simulations, it was observed that the relative median errors in POS mean discharges at these watersheds, ranging from 6 to 36 percent, were larger than those at the calibration watersheds, being 3 and 5 percent at Tinley Creek and Long Run watersheds, respectively. In agreement with their larger median errors, the corresponding V_N values overall were smaller at the prediction watersheds, ranging from 0.026 to 0.248, compared to 0.191 to 0.787 for the calibration watersheds. The V_N values for the FDC and WY mean discharges at the prediction watersheds were also almost always substantially smaller than those at the calibration watersheds, probably also as a result of larger median error. These larger median errors and the correspondingly smaller V_N values at the prediction watersheds likely result from parameter-transfer error, which is reduced at the calibration watersheds to the balancing of errors between the two watersheds. Also, considering the percent quantiles of FDC to be more detailed statistics than the POS mean from a magnitude perspective, and WY mean discharges to be more detailed statistics than the POS mean from a time-scale perspective, it is interesting to note that the corresponding V_N values are similar among the POS mean, FDCs, and WY mean discharge statistics. The V_N values do not trend consistently up or down with increasing detail in magnitude or time scale in the flow statistics analyzed. The reason for this lack of trend with detail may be the normalized nature of the V_N statistic itself; trends may be present in its components.

The ranking of the V_N values among the seven prediction watersheds was largely consistent across the different flow statistics. The Flag Creek watershed consistently produced the lowest V_N values for the three statistics and the two sets of simulated discharges, and the Butterfield Creek watershed produced the second lowest. The Midlothian Creek produced the highest V_N values for the POS mean and WY mean discharges, whereas it ranked near the middle for the FDCs, and the Hart Ditch at Munster watershed produced the highest.

The V_N values determined in this study, being generally less than 1 and often substantially less than 1, particularly at the prediction watersheds, indicate that the two sources of uncertainty considered here do not explain all the uncertainty of the HSPF streamflow simulations considered. This result indicates the need to consider other sources of uncertainty in the HSPF simulations for the LMDA system. These potential sources that were not simulated include those pertaining the forcing data (particularly, precipitation, which may have measurement and spatial distribution errors); the estimates of the forest and grassland fractions, the uncertainty of the impervious land parameters; and, for the prediction watersheds, parameter-transfer error. In addition, to allow the uncertainty from model recalibrations that simulated the effect of published discharge uncertainty to be added to the other components, a more efficient recalibration approach would be beneficial, so that a sample that is larger than the one used in this study and has known statistical properties can be used.

References Cited

Bauer, E., and Westcott, N., 2017, Continued operation of a 25-raingage network for collection, reduction, and analysis of precipitation data for Lake Michigan Diversion Accounting—Water year 2016: Illinois State Water Survey, Prairie Research Institute, University of Illinois at Urbana-Champaign, Contract Report 2017-03, 88 p., accessed September 2020 at https://www.isws.illinois.edu/pubdoc/CR/ISWSCR2017-03.pdf.

Beven, K., and Freer, J., 2001, Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology: Journal of Hydrology (Amsterdam), v. 249, no. 1–4, p. 11–29, accessed September 2020 at https://doi.org/10.1016/S0022-1694(01)00421-8.

Bicknell, B.R., Imhoff, J.C., Kittle, J.L., Jr., Jobes, T.H., and Donigian, A.S., Jr., 2005, HSPF version 12.2 user's manual: Prepared by AQUA TERRA Consultants for Office of Surface Water, U.S. Geological Survey, Reston, Va., and U.S. Environmental Protection Agency, Athens, Ga., EPA/600/R-97/080, 755 p.

Bloschl, G., Sivapalan, M., Wagener, T., and Viglione, A., eds., 2013, Runoff prediction in ungauged basins—Synthesis across processes, places, and scales: Cambridge University Press, 465 p. [Also available at https://doi.org/10.1017/CBO9781139235761.]

Chow, V.T., 1964, Handbook of applied hydrology: McGraw Hill, New York, [variously paged].

Diaz-Ramirez, J.N., Johnson, B.E., McAnally, W.H., Martin, J.L., Alarcon, V.J., and Camacho, R.A., 2013, Estimation and propagation of parameter uncertainty in lumped hydrological models—A case study of HSPF model applied to Luxapallila Creek watershed in Southeast USA: Journal of Hydrogeology and Hydrologic Engineering, v. 2, no. 1, 9 p. [Also available at https://doi.org/10.4172/2325-9647.1000105.]

Doherty, J., 2010, Methodologies and software for PEST-based model predictive uncertainty analysis: Watermark Numerical Computing, 157 p., accessed September 2015 at https://www.pesthomepage.org/About_Us.php.

Doherty, J., 2013, Getting the most out of PEST: Watermark Numerical Computing, 12 p., accessed September 2020 at https://pesthomepage.org/documentation.

Doherty, J., 2015, Calibration and uncertainty analysis for complex environmental models, PEST—Complete theory and what it means for modelling the real world: Brisbane, Australia, Watermark Numerical Computing, 237 p.

Doherty, J., and Christensen, S., 2011, Use of paired simple and complex models to reduce predictive bias and quantify uncertainty: Water Resources Research, v. 47, no. 12, W12534, 21 p. [Also available at https://doi.org/10.1029/2011WR010763.]

Doherty, J., and Welter, D., 2010, A short exploration of structural noise: Water Resources Research, v. 46, no. 5, W05525, 14 p. [Also available at https://doi.org/10.1029/2009WR008377.]

Doherty, J., and Johnston, J.M., 2003, Methodologies for calibration and predictive analysis of a watershed model: Journal of the American Water Resources Association, v. 39, no. 2, p. 251–265. [Also available at https://doi.org/10.1111/j.1752-1688.2003.tb04381.x.]

Dunn, S.M., and Lilly, A., 2001, Investigating the relationship between a soils classification and the spatial parameters of a conceptual catchment-scale hydrological model: Journal of Hydrology (Amsterdam), v. 252, no. 1-4, p. 157–173. [Also available at https://doi.org/10.1016/S0022-1694(01)00462-0.]

Espey, W.H., Melching, C.S., and Muste, M., 2019, Lake Michigan Diversion Committee—Findings of the eighth technical committee for review of diversion flow measurements and accounting procedures: U.S. Army Corps of Engineers, Chicago District, 175 p., accessed June 2019 at https://www.lrc.usace.army.mil/Portals/36/docs/divacct/technical/Eighth_Technical_Complete.pdf.

Espey, W.H., Melching, C.S., and Muste, M., 2009, Lake Michigan Diversion Committee—Findings of the sixth technical committee for review of diversion flow measurements and accounting procedures: U.S. Army Corps of Engineers, Chicago District, 212 p., accessed June 2019 at https://www.lrc.usace.army.mil/Portals/36/docs/divacct/technical/Sixth_Technical_Complete.pdf.

Espey, W.H., Melching, C.S., and Mades, D., 2004, Lake Michigan Diversion Committee—Findings of the fifth technical committee for review of diversion flow measurements and accounting procedures: U.S. Army Corps of Engineers, Chicago District, 194 p. , accessed June 2019 at https://www.lrc.usace.army.mil/Portals/36/docs/divacct/technical/Fifth_Technical_Complete.pdf.

Espey, W.H., Lara, O.G., and Barkau, R.L., 1993, Lake Michigan Diversion Committee—Findings of the third technical committee for review of diversion flow measurements and accounting procedures: U.S. Army Corps of Engineers, Chicago District, 219 p., accessed July 2019 at https://www.lrc.usace.army.mil/Portals/36/docs/divacct/technical/third_Technical_Complete.pdf.

Everitt, B.S., and Skrondal, A., 2010, Cambridge Dictionary of Statistics (4th ed.): Cambridge, Cambridge University Press, 468 p.

Farmer, W., and Vogel, R.M., 2016, On the deterministic and stochastic use of hydrologic models: Water Resources Research, v. 52, no. 7, p. 5619–5633. [Also available at https://doi.org/10.1002/2016WR019129.]

Gábor, A., and Banga, J.R., 2015, Robust and efficient parameter estimation in dynamic models of biological systems: BMC Systems Biology, v. 9, no. 74, 25 p. [Also available at https://doi.org/10.1186/s12918-015-0219-2.

Greenspan, H.P., and Benney, D.J., 1973, Calculus—An introduction to applied mathematics: New York, McGraw-Hill, 784 p.

Gupta, H.V., Beven, K.J., and Wagener, T., 2006, Model calibration and uncertainty estimation, *in* Anderson, M.G., and others, eds., Encyclopedia of Hydrological Sciences: Chichester, United Kingdom, John Wiley & Sons Ltd, 17 p. [Also available at https://doi.org/10.1002/0470848944.hsa138.]

Hill, L., 2007, The Chicago River—A natural and unnatural history: Chicago, Lake Claremont Press, 302 p.

Homer, C., Dewitz, J., Yang, L., Jin, S., Danielson, P., Xian, G., Coulston, J., Herold, N., Wickham, J., and Megown, K., 2015, Completion of the 2011 National Land Cover Database for the conterminous United States—Representing a decade of land cover change information: Photogrammetric Engineering and Remote Sensing, v. 81, no. 5, p. 345–354. [Also available at https://www.ingentaconnect.com/content/asprs/pers/2015/00000081/00000005/art00002.]

Hrachowitz, M., Savenije, H.H.G., Blöschl, G., McDonnell, J.J., Sivapalan, M., Pomeroy, J.W., Arheimer, B., Blume, T., Clark, M.P., Ehret, U., Fenicia, F., Freer, J.E., Gelfan, A., Gupta, H.V., Hughes, D.A., Hut, R.W., Montanari, A., Pande, S., Tetzlaff, D., Troch, P.A., Uhlenbrook, S., Wagener, T., Winsemius, H.C., Woods, R.A., Zehe, E., and Cudennec, C., 2013, A decade of Predictions in Ungauged Basins (PUB)—A review: Hydrological Sciences Journal, v. 58, no. 6, p. 1198–1255. [Also available at https://doi.org/10.1080/02626667.2013.803183.]

Hunt, R.J., Doherty, J., and Tonkin, M.J., 2007, Are models too simple? Arguments for increased parameterization: Ground Water, v. 45, no. 3, p. 254–262. [Also available at https://doi.org/10.1111/j.1745-6584.2007.00316.x.]

Kennedy, E.J., 1983, Computation of continuous records of streamflow: U.S. Geological Survey Techniques of Water-Resources Investigations, book 3, chap. A13, 53 p., accessed September 2020 at https://pubs.usgs.gov/twri/twri3-a13/.

Kennedy, E.J., 1984, Discharge ratings at gaging stations: U.S. Geological Survey Techniques of Water-Resources Investigations, book 3, chap. A10, 59 p., accessed September 2020 at https://pubs.usgs.gov/twri/twri3-a10/.

Le Coz, J., Renard, B., Bonnifait, L., Branger, F., and Le Boursicaud, R., 2014, Combining hydraulic knowledge and uncertain gaugings in the estimation of hydrometric rating curves—A Bayesian approach: Journal of Hydrology (Amsterdam), v. 509, p. 573–587. [Also available at https://doi.org/10.1016/j.jhydrol.2013.11.016.]

Lumb, A.M., McCammon, R.B., and Kittle, J.L., Jr., 1994, Users manual for an expert system (HSPEXP) for calibration of the Hydrological Simulation Program-FORTRAN: U.S. Geological Survey Water-Resources Investigations Report 94–4168, 102 p. [Also available at https://water.usgs.gov/software/code/water_quality/hspexp/doc/hspexp.pdf.]

Mansanarez, V., Renard, B., Le Coz, J., Lang, M., and Darienzo, M., 2019, Shift happens! Adjusting stage-discharge rating curves to morphological changes at known times: Water Resources Research, v. 55, no. 4, p. 2876–2899. [Also available at https://doi.org/10.1029/2018WR023389.]

Moore, C., and Doherty, J., 2005, Role of the calibration process in reducing model predictive error: Water Resources Research, v. 41, no. 5. [Also available at https://doi.org/10.1029/2004WR003501.]

National Centers for Environmental Information, 2018, Climate data online: accessed April 6, 2018, at https://www.ncdc.noaa.gov/cdo-web/.

National Research Council, 2012, Assessing the reliability of complex models—Mathematical and statistical foundations of verification, validation, and uncertainty quantification: Washington D.C., The National Academic Press. [Also available at https://www.nap.edu/catalog/13395/assessing-the-reliability-of-complex-models-mathematical-and-statistical-foundations.]

National Weather Service, 2020, Cooperative Observer Program webpage: accessed November 27, 2020, at https://www.weather.gov/coop/.

Over, T.M., Soong, D.T., and Sortor, R.N., 2022, Models, inputs, and outputs for estimating the uncertainty of discharge simulations for the Lake Michigan Diversion using the Hydrological Simulation Program – FORTRAN model: U.S. Geological Survey data release, https://doi.org/10.5066/P9UC21B0.

Over, T.M., Soong, D.T., and Su, T.Y., 2010, HSPF snow modeling parameters in an urban setting—Evaluation using NWS Coop and SNODAS data and re-calibration using PEST: Proceedings of EWRI Watershed Management Conference 2010, Madison, Wisconsin, August 2010. [Also available at https://doi.org/10.1061/41143(394)110.]

Peppier, R.A., 1991, Installation and operation of a dense raingage network to improve precipitation measurements for Lake Michigan diversion accounting—Water year 1990: Atmospheric Science Section, Illinois State Water Survey, SWS Contract Report 517, 87 p., accessed September 2020 at https://core.ac.uk/download/pdf/158299482.pdf.

R Core Team, 2019, R—A language and environment for statistical computing: Vienna, Austria, R Foundation for Statistical Computing. [Also available at https://www.R-project.org/.]

Rantz, S.E., 1982, Measurement and computation of streamflow: U.S. Geological Survey Water-Supply Paper 2175, [variously paged], accessed September 2020 at https://doi.org/10.3133/wsp2175.

Refsgaard, J.C., van der Sluijs, J.P., Højberg, A.L., and Vanrolleghem, P.A., 2007, Uncertainty in the environmental modelling process—A framework and guidance: Environmental Modelling & Software, v. 22, no. 11, p. 1543–1556. [Also available at https://doi.org/10.1016/j.envsoft.2007.02.004.]

Resource Coordination Policy Committee, 1998, Our community and flooding—A report of the status of floodwater management in the Chicago Metropolitan Area: Resource Coordination Policy Committee, 72 p. [Also available at https://www2.illinois.gov/dnr/WaterResources/Documents/OurCommunityAndFlooding_Oct1998.pdf.]

RUST Environment and Infrastructure, 1993, Technical memorandum on HSPF parameter assignments used in the Lake Michigan Diversion Accounting Program: Submitted to the U.S. Army Corps of Engineers, Chicago District, 7 p. with 6 appendices.

Sauer, V.B., and Turnipseed, D.P., 2010, Stage measurement at gaging stations: U.S. Geological Survey Techniques and Methods, book 3, chap. A7, 45 p. [Also available at https://doi.org/10.3133/tm3a7.]

Sharpe, J.B., and Soong, D.T., 2015, Lake Michigan Diversion Accounting land cover change estimation by use of the National Land Cover Dataset and raingage network partitioning analysis: U.S. Geological Survey Open-File Report 2014–1258, 12 p., accessed September 2020 at https://doi.org/10.3133/ofr20141258.

Sitterson, J., Knightes, C., Parmar, R., Wolfe, K., Muche, M., and Avant, B., 2017, An overview of rainfall-runoff model types: U.S. Environmental Protection Agency, Office of Research and Development, Oak Ridge Institute for Science and Education., EPA/600/R-14/152, 30 p. [Also available at https://cfpub.epa.gov/si/si_public_record_report.cfm?dirEntryId=339328&Lab=NERL.]

Sivapalan, M., 2003, Prediction in ungauged basins—A grand challenge for theoretical hydrology: Hydrological Processes, v. 17, no. 15, p. 3163–3170. [Also available at https://doi.org/10.1002/hyp.5155.]

Sivapalan, M., Takeuchi, K., Franks, S.W., Gupta, V.K., Karambiri, H., Lakshmi, V., Liang, X., McDonnell, J.J., Mendiondo, E.M., O'Connell, P.E., Oki, T., Pomeroy, J.W., Schertzer, D., Uhlenbrook, S., and Zehe, E., 2003, IAHS Decade on Predictions in Ungauged Basins (PUB), 2003–2012—Shaping an exciting future for the hydrological sciences: Hydrological Sciences Journal, v. 48, no. 6, p. 857–880. [Also available at https://doi.org/10.1623/hysj.48.6.857.51421.]

Soong, D.T., and Over, T.M., 2015, Analysis of regional rainfall-runoff parameters for the Lake Michigan Diversion hydrological modeling: U.S. Geological Survey Scientific Investigations Report 2015-5053, 55 p., accessed September 2020 at https://doi.org/10.3133/sir20155053.

Tolson, B.A., and Shoemaker, C.A., 2007, Dynamically dimensioned search algorithm for computationally efficient watershed model calibration: Water Resources Research, v. 43, no. 1, W01413, 16 p. [Also available at https://doi.org/10.1029/2005WR004723.]

Turnipseed, D.P., and Sauer, V.B., 2010, Discharge measurements at gaging stations: U.S. Geological Survey Techniques and Methods, book 3, chap. A8, 87 p. [Also available at https://doi.org/10.3133/tm3a8.]

U.S. Army Corps of Engineers, Chicago District, [undated] a, Lake Michigan Diversion Accounting Program: accessed June 26, 2020, at https://www.lrc.usace.army.mil/Missions/Lake-Michigan-Diversion-Accounting/.

U.S. Army Corps of Engineers, Chicago District, [undated] b, Lake Michigan Diversion Accounting: accessed June 26, 2020, at https://www.lrc.usace.army.mil/Missions/Civil-Works-Projects/Lake-Michigan-Diversion-Accounting/.

U.S. Army Corps of Engineers, Chicago District, 2009, Lake Michigan Diversion Accounting—Water year 2009 report: U.S. Army Corps of Engineers, 85 p. [Also available at https://www.lrc.usace.army.mil/Portals/36/docs/divacct/annual/2009-Annual.pdf.]

U.S. Environmental Protection Agency, 2000, BASINS Technical Note 6—Estimating hydrology and hydraulic parameters for HSPF: U.S. Environmental Protection Agency, Office of Water, EPA-823-R00-012, 34 p., accessed April 2020 at https://www.epa.gov/sites/production/files/2015-08/documents/2000_08_14_basins_tecnote6.pdf.

U.S. Geological Survey, 2014, National Land Cover Database (NLCD) 2011 Land Cover Conterminous United States: U.S. Geological Survey data release, accessed March 2017 at https://doi.org/10.5066/P97S2IID.

U.S. Geological Survey, 2020, USGS water data for the Nation: U.S. Geological Survey National Water Information System database, accessed July 2020 at https://doi.org/10.5066/F7P55KJN.

Wagener, T., Wheater, H.S., and Gupta, H.V., 2004, Rainfall-runoff modeling in gauged and ungauged catchments: London, Imperial College Press, 332 p. [Also available at https://doi.org/10.1142/p335.]

Watermark Numerical Computing, 2020a, PEST, model-independent parameter estimation, User manual part 1—PEST, SENSAN and Global Optimisers (7th ed.) with February 2020 additions: 393 p., accessed November 2020 at https://www.pesthomepage.org/Downloads.php.

Watermark Numerical Computing, 2020b, PEST, model-independent parameter estimation, User manual part 2—PEST utility support software (7th ed.) with February 2020 additions: 267 p., accessed November 2020 at https://www.pesthomepage.org/Downloads.php.

Westenbroek, S.M., Doherty, J., Walker, J.F., Kelson, V.A., Hunt, R.J., and Cera, T.B., 2012, Approaches in highly parameterized inversion—TSPROC, a general time-series processor to assist in model calibration and result summarization: U.S. Geological Survey Techniques and Methods, book 7, chap. C7, 79 p., 3 apps. [Also available at https://pubs.usgs.gov/tm/tm7c7.]

Appendix 1. Initial and Ranges of Parameter Values for Calibrating the Grassland and Forest Land Segments of the Hydrological Simulation Program–FORTRAN Model

Table 1.1. A list of parameter names for pervious lands with their abbreviated explanations, unit in English system, and range and initial values to start the calibration of grassland and forest land parameters of the base model.

[Base model refers to calibration with published discharge records. 1/in, inverse of an inch; ET, evapotranspiration; PET, potential evapotranspiration]

Parameter[1]	Explanation	Unit	Range (lower bound–upper bound) of parameter values		Initial parameter values for calibrating the base model	
			Grassland	Forest	Grassland	Forest
FOREST	Fraction of land that covered in (for example, conifer) forest that can transpire when there is snowpack	None	0.05–0.2	0.2–0.5	0.2	0.3
LZSN	Lower zone nominal soils moisture storage	Inches	2.0–15.0	1.4–19.5	6.6808	6.6808
INFILT	Index of infiltration capacity	Inches per hour	0.001–0.10	0.001–0.105	0.032	0.033
LSUR	Length of overland flow plane	Feet	50–500	50–500	50.0	400.0
SLSUR	Average slope of assumed overland flow plane	Feet per feet	0.0051–0.02	0.001–0.02	0.0085	0.0085
KVARY	A constant describes the nonlinear groundwater recession	1/in	0.00001–3.0	0.00001–3.0	1.377	1.240
AGWRC	A constant describes the base groundwater recession rate	none	0.917–0.98	0.913–0.982	0.968	0.966
PETMAX	Temperature below which ET is reduced to 50 percent of that in the input time series. Active only when snow processes are being simulated	Degrees Fahrenheit	35–45	35–45	45.0	45.0
PETMIN	Temperature below which ET is zero. Active only when snow processes are being simulated	Degrees Fahrenheit	30–35	30–35	35.0	35.0
INFEXP	Exponent in infiltration equation that determines how much a deviation from nominal lower zone storage affects the infiltration rate	None	2.0	2.0	2.0	2.0
INFILD	Ratio of maximum/mean soil infiltration capacities	None	2.0	2.0	2.0	2.0
DEEPFR	Fraction of infiltrating water lost to inactive groundwater (deep percolation)	None	0.001–0.20	0.001–0.20	0.0501	0.0495
AGWETP	Fraction of PERLND subject to direct evaporation from groundwater storage	None	0.0001–0.1	0.0001–0.105	0.0399	0.0399
CEPSC	Interception storage capacity by vegetation	Inches	0.03–0.20	0.03–0.50	0.0848	0.1329
UZSN	Nominal upper zone soil moisture storage	Inches	0.05–1.2	0.05–2.4	1.017	1.130
NSUR	Manning's roughness for overland flows	None	0.15–0.25	0.25–0.40	0.25	0.40

Table 1.1. A list of parameter names for pervious lands with their abbreviated explanations, unit in English system, and range and initial values to start the calibration of grassland and forest land parameters of the base model.—Continued

[Base model refers to calibration with published discharge records. 1/in, inverse of an inch; ET, evapotranspiration; PET, potential evapotranspiration]

Parameter[1]	Explanation	Unit	Range (lower bound–upper bound) of parameter values		Initial parameter values for calibrating the base model	
			Grassland	Forest	Grassland	Forest
INTFW	Determines interflow	None	1.0–5.0	0.75–1.0	2.500	2.500
IRC	Interflow recession parameter	None	0.5–0.7	0.5–0.85	0.649	0.531
LZETP	Parameter determines lower zone ET	None	0.1–0.9	0.15–3.24	0.138	0.208

[1]Parameters INFEXP, INFILD, and BASETP are not varied in the Lake Michigan Diversion Accounting Hydrological Simulation Program—FORTRAN analysis.

For more information about this publication, contact:

Director, USGS Central Midwest Water Science Center
405 North Goodwin
Urbana, IL 61801
217–328–8747

For additional information, visit: https://www.usgs.gov/centers/cmwater

Publishing support provided by the
Rolla Publishing Service Center

ISSN 2328-0328 (online)
https://doi.org/10.3133/sir20225102